A globally exponentially stable filter for bearing-only simultaneous localization and mapping with monocular vision

Pedro Lourenço, Pedro Batista, Paulo Oliveira, and Carlos Silvestre

Robotics and Autonomous Systems, vol. 100, pp. 61-77, February 2018

https://doi.org/10.1016/j.robot.2017.11.001

A Globally Exponentially Stable filter for Bearing-Only Simultaneous Localization and Mapping with Monocular vision

Pedro Lourenço[a,*], Pedro Batista[a,b], Paulo Oliveira[a,c], Carlos Silvestre[a,d]

[a] *Institute for Systems and Robotics, Laboratory for Robotics and Engineering Systems, Portugal.*
[b] *Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
[c] *Department of Mechanical Engineering, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1049-001 Lisboa, Portugal*
[d] *Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Macau*

Abstract

This paper proposes a novel filter for sensor-based bearing-only simultaneous localization and mapping in three dimensions with globally exponentially stable (GES) error dynamics. A nonlinear system is designed, its output transformed, and its dynamics augmented so that the proposed formulation can be considered as linear time-varying for the purpose of observability analysis. This allows the establishment of observability results related to the original nonlinear system that naturally lead to the design of a Kalman filter with GES error dynamics. The performance of the proposed algorithm is assessed resorting to real experiments based on the *Rawseeds* dataset as well as further realistic simulations.

Keywords: Simultaneous localization and mapping, 3-D mapping, Sensor fusion, Monocular vision, Global exponential stability

1. Introduction

Navigation using directions to known sources has been in use for centuries. Initially, in marine applications, several tools to measure the elevation of stars such as sextants and mariner's astrolabes were employed to derive the position of ships, and lighthouses were used in triangulation techniques. In the last century, aviation brought into use more advanced technologies supported on bearings (azimuth and/or elevation) readings: the automatic direction finder

*Corresponding author
Email addresses: plourenco@isr.ist.utl.pt (Pedro Lourenço), pbatista@isr.ist.utl.pt (Pedro Batista), pjcro@isr.ist.utl.pt (Paulo Oliveira), csilvestre@umac.mo (Carlos Silvestre)

(ADF), the VHF omnidirectional range (VOR), and the instrument landing system (ILS) are the most common still in use today. The advent of global positioning systems has gradually replaced the use of these techniques, but in GPS-denied environments, aided relative algorithms are still called for to navigate unmanned vehicles. A family of algorithms that addresses this issue is simultaneous localization and mapping (SLAM), a concept introduced in the scientific community in the 1980's [1] and first coined in [2].

The most studied version of the SLAM problem is what is called range-bearing SLAM, where the coordinates of measured landmarks are readily available (see [3] and [4] for a survey on the algorithms proposed in the first decades of SLAM research, [5] for a more up to date review focused in the recent theoretical achievements and [6] for an overarching survey on the history and remaining present and future challenges of SLAM, e.g., robustness and scalability, within others). This is known in the scientific community as fully observable SLAM, as a single measurement is sufficient to estimate landmark positions. However, there are versions of the problem that omit one of the two informations available, either range-only SLAM (RO-SLAM) or bearing-only SLAM (BO-SLAM). These approaches are named partially-observable, as a single noise-free observation provides only a line or surface as an estimate for the position of a landmark. The bearing-only case is even more difficult to treat than the range-only one, because an observation corresponds to an unbounded region. This raises serious issues on the initialization of a landmark, which has been the main topic of research in partially-observable SLAM, yielding initially only delayed solutions, i.e., algorithms that try to obtain a preliminary landmark estimate from readings at different viewpoints before introducing the initial estimate into the filter. This can be done through triangulation or more advanced probabilistic approaches, using, for example, a sum of Gaussians [7] or deferring the initialization until an approximately Gaussian estimate is achieved [8]. There were some notable exceptions in [9] and [10] that recur to multiple hypothesis directly in the filter. More recently, algorithms with undelayed initialization, at least to some extent, have been proposed. The concept of inverse depth parametrization [11] was inspired by computer vision and it brought to bearing-only EKF-SLAM several advantages, as the inverse depth has better linearity and allows for very low parallax features. However, it requires a six-dimensional (6-D) representation of the feature state, instead of the traditional tridimensional (3-D) representation. In this case, feature depth, or rather, inverse depth, is initialized with a generic prior that accounts for statistically feasible depths. Another technique was proposed in [12], where the authors transform the measurement in image coordinates to a 3-D vector in camera coordinates by assuming an arbitrary depth while defining a very large covariance for that measurement. Even though these two approaches brought a fresh perspective to the field and seem to have good results in practice, there are no guarantees of convergence.

Although research in BO-SLAM is not as prolific as it is in range-bearing SLAM, the community has provided several approaches depending on the underlying filtering technique and the sensors used. Most algorithms are based on extended kalman filters (EKF). However, some methods are inspired on

expectation-maximization or particle filters (see [13] for a comparison of these approaches). Another source of diversity in BO-SLAM algorithms is the type of sensor used. Even though bearing-only localization is historically related to the computation of the angle-of-arrival of signals from beacons through the time difference of arrival at different elements of a receiving array, BO-SLAM is mostly associated with monocular vision [14] or even catadioptric omnidirectional systems [15], even though there are still BO-SLAM systems using acoustic sensors [16]. Monocular SLAM and other vision-based algorithms have been the focus of a proficuous research effort, yielding several relevant algorithms, of which one of the most important is [17], as it was one of the first real-time SLAM algorithm with a single camera as the only data source (even though it was preceded in computer vision by [18]). This algorithm initializes features as semi-infinite lines lying in the direction of the detected features. Other, more recent, approaches vary in application, the way features are handled, and the underlying filtering engine. For example, [19] represents a departure from the usual exploitation of the ubiquity of cameras, as it presents possible new applications. In this case, underwater hull inspection. Other interesting approach is ORB-SLAM [20] that uses ORB features [21] which are rotation invariant, as SURF features [22], and have faster extraction. The main algorithm performs feature tracking and on top has external loop closing and re-localization procedures. Demonstrating the possibilities of graph-based solutions, the authors of [23] introduce a closed-form pose-chain optimization algorithm that uses sparse graphs as well as appearance-based loop detection. Another very important approach to monocular SLAM was imported from computer vision, where structure-from-motion (SFM) has been a subject of research for many years [24]. In that field, there have been some notable advances that can be related to BO-SLAM, such as [18], where the authors propose an EKF to address SFM in real-time and perform a characterization of observability and minimal realization. Also in SFM, bundle adjustment (BA) techniques [25] have become the standard. In bearing-only SLAM this concept is much more recent, but there have been notable efforts as detailed in [26]. There the authors perform a comparison between established bundle-adjustment and filtering techniques, concluding that in general BA has a better accuracy/computational cost ratio than EKF filtering. However, it is also mentioned that BA procedures have no general robustness to initialization parameters and that filters may be more capable of dealing with the high uncertainty present in that phase.

One of the greatest problems in any SLAM framework is data association. In range-bearing SLAM this issue is mostly solved and there are several different algorithms that tackle it. In BO-SLAM, however, there are extra difficulties in the initialization process. Some approaches try to deal with all hypothesis when initializing [27], but vision-based algorithms may use image information if the frame-rate is high enough to disambiguate measurements. This was the approach followed successfully by the authors in a previous work in range-bearing SLAM [28], and recovered for the algorithm presented in this paper. Feature extraction provides tools with which to distinguish different features, that can be combined with well tested association methods, such as the sequen-

tial compatibility nearest neighbour or joint compatibility branch and bound [29]. Connected with the association problem is the issue of loop closing, i.e., the ability to recognize previously visited terrain. Before the widespread use of vision in SLAM, loop closing was based on landmark position or map-to-map comparison, or scan matching when dealing with laser range finders. With the information present in images, other possibilities arose, such as image-to-image or image-to-map correspondence. These are discussed at length in [30].

This work proposes a solution to the initialization problem by introducing a BO-SLAM algorithm with exponentially fast global convergence, which allows for undelayed initialization at any arbitrary depth. With its tridimensional (3-D) sensor-based approach, the pose of the vehicle is eliminated from the filter state and the inclusion of odometry-like measurements and relative bearings comes naturally. This aspect, coupled with a state augmentation and output transformation, leads to the design of an LTV system whose observability is analysed in this paper, resulting in constructive conditions with clear physical insight that are important for motion planning. The underlying idea of this paper is influenced by the source-localization algorithm presented in [31], as the proposed filter results from similar state and output transformations. This work was first presented in a conference version in [32], and is herein presented with new observability results, complete proofs and experimental validation using a monocular camera, and data from a widely available dataset, the *Rawseeds* dataset [33, 34].

1.1. Paper organization and notation:

The paper is organized as follows. A short description of the problem, with the definition of the system dynamics, is presented in Section 2. Section 3 details the proposed solution, including the observability analysis and filter design. The implementation of the algorithm is described in Section 4 including landmark detection, data association, and loop closing procedures. Experimental results based on the *Rawseeds* dataset are described and discussed in Section 5, accompanied by simulation results, whereas in Section 6 the concluding remarks and direction for further research are provided.

The following symbol convention is used in this paper: vectors are represented in bold small letters, matrices in bold capital letters and scalar symbols are expressed in italic, constants by capital letters, and variables in small letters. The superscript ^{E} indicates a vector or matrix expressed in the Earth-fixed frame $\{E\}$. For the sake of clarity, when no superscript is present, the vector is expressed in the body-fixed frame $\{B\}$. $\mathbf{I}_n$ is the identity matrix of dimension n, and $\mathbf{0}_{n\times m}$ is a n by m matrix filled with zeros. If m is omitted, the matrix is square. $\mathbf{S}[\mathbf{a}]$ is a special skew-symmetric matrix, henceforth called the cross-product matrix, as $\mathbf{S}[\mathbf{a}]\,\mathbf{b} = \mathbf{a}\times\mathbf{b}$ with $\mathbf{a},\mathbf{b}\in\mathbb{R}^3$.

2. The bearing-only SLAM problem

Consider a vehicle operating in a static environment, capable of measuring the relative azimuth and elevation of landmarks installed in unknown locations,

as well as its linear and angular velocities in its own reference frame. Possible examples are ground robots equipped with wheel encoders, or aerial vehicles equipped with optical flow sensors and an altimeter. Both of these assisted by an inertial measurement unit (IMU) to measure the angular velocities and a camera or an acoustic/electromagnetic receiver to detect landmarks. The landmarks can be artificial or natural, i.e., previously installed or extracted from the scenery. This situation falls under the scope of BO-SLAM, which is the problem of navigating a vehicle in an unknown environment, building a map of metric landmarks by measuring bearings and using this map to deduce its location, without the need for *a priori* information about landmark inertial location.

2.1. The sensor-based approach

The sensor-based approach, or, as it is commonly known in the SLAM community, the robocentric approach to SLAM has been proven more consistent than its inertial, world-centric, counterpart [35]. Furthermore, previous observability studies using piece-wise linearizations showed that this approach becomes fully observable in two time steps, in opposition to what happens in the world-centric case [36]. In addition, in this family of problems where the measurements are all expressed in local coordinates it makes sense to operate in a sensor-based framework, as it is a way of avoiding the inclusion of the pose of the vehicle in the filter state, one of the main sources of nonlinearity. That is the idea behind the nonlinear system that underlies the filter to be detailed. Paramount to the sensor-based idea is the fact that only the relative coordinates of the map with respect to the vehicle are estimated. In fact, it is possible to achieve similar results in a world-centric strategy, as long as this relative property is kept and external estimates for the attitude are available. An example of this, inspired in the authors sensor-based approach to SLAM [37, 38, 28] and the theoretical results associated, is given in [39]. Earth-fixed global estimates can still be obtained from relative information using an algorithm such as the one proposed in [40, 41], using at least two landmarks as anchors with known coordinates, or acquiring an external estimate of the attitude of the vehicle (e.g. with an Attitude Heading Reference System [42]). Recall the situation described above, and consider two different reference frames. One fixed to the vehicle, denoted as body-fixed frame $\{B\}$, and the other fixed in the environment, denoted as the Earth-fixed frame $\{E\}$. The two frames are related through the rotation matrix $\mathbf{R}(t) \in \mathrm{SO}(3)$ and the translation ${}^{E}\mathbf{p}(t) \in \mathbb{R}^3$. The former represents the attitude of the vehicle and satisfies $\dot{\mathbf{R}}(t) = \mathbf{R}(t)\mathbf{S}\left[\boldsymbol{\omega}(t)\right]$, where $\boldsymbol{\omega}(t) \in \mathbb{R}^3$ is the angular velocity of the vehicle expressed in $\{B\}$. Similarly, the translation represents the position of the vehicle in the Earth-fixed frame, coincident with the origin of the body-fixed frame expressed in $\{E\}$.

The environment, i.e., the map, consists of N static landmarks ${}^{E}\mathbf{p}_i \in \mathbb{R}^3$ that compose the landmark set $\mathcal{M} = \{1, \ldots, N\}$. Depending on the pose of the vehicle, some of these landmarks may be visible or not, which motivates the definition of two subsets of landmarks, $\mathcal{M}_o = \{1, \ldots, N_O\}$ and $\mathcal{M}_u = \{N_O + 1, \ldots, N\}$. The first contains the N_O observed or visible landmarks while

the latter contains the unobserved, or non-visible, ones. Note that, without loss of generality, the landmarks are ordered for simplicity of analysis. In the body-fixed frame, the i-th landmark is denoted by $\mathbf{p}_i(t) = \mathbf{R}^T(t)\left({}^E\mathbf{p}_i - {}^E\mathbf{p}(t)\right)$ and its derivative satisfies

$$\dot{\mathbf{p}}_i(t) = -\mathbf{S}\left[\boldsymbol{\omega}(t)\right]\mathbf{p}_i(t) - \mathbf{v}(t),$$

where $\mathbf{v}(t) \in \mathbb{R}^3$ is the linear velocity of the vehicle expressed in its own frame.

From the problem definition, it is known that both the linear and the angular velocities are measured, as well as relative bearings to the landmarks. This last quantity is described by the unit vector $\mathbf{b}_i(t)$ that defines the line between the position of the vehicle and landmark i, and is given by

$$\mathbf{b}_i(t) = \frac{\mathbf{p}_i(t)}{\|\mathbf{p}_i(t)\|} \in \mathcal{S}(2),$$

with $i \in \mathcal{M}_o$. As the information this measurement carries is limited, not only several measurements from the same landmark are needed to unambiguously determine its position but also some measure of scale is required. This is provided by the linear velocity measurements and it is the reason why they must be available.

This section culminates naturally with the nonlinear system that puts all this information together. The positions of the landmarks in the body-fixed frame are its states, the linear and angular velocities are the inputs, and the measured quantities are its outputs ($\mathbf{b}_i(t)$). The resulting system model is

$$\begin{cases} \dot{\mathbf{p}}_i(t) = -\mathbf{S}\left[\boldsymbol{\omega}(t)\right]\mathbf{p}_i(t) - \mathbf{v}(t) \\ \mathbf{b}_j(t) = \dfrac{\mathbf{p}_j(t)}{\|\mathbf{p}_j(t)\|} \end{cases} \tag{1}$$

where $i \in \mathcal{M}$ and $j \in \mathcal{M}_o$. The output $\mathbf{b}_j(t)$ can be stacked in a column vector to obtain $\mathbf{b}(t) = \begin{bmatrix}\mathbf{b}_1^T(t) & \cdots & \mathbf{b}_{N_O}^T(t)\end{bmatrix}^T$.

2.2. Problem statement

The problem addressed in this paper is that of designing a navigation system for a vehicle operating in the environment here described, by means of a filter for the dynamics in (1), assuming noisy measurements. The algorithm consists of a BO-SLAM filter in the space of sensors, and, therefore, the pose of the vehicle is deterministic as, by construction, it corresponds to the position and attitude of the body-fixed frame expressed in that same frame.

3. Proposed solution: GES BO-SLAM

The system presented in the last section is still nonlinear, even though the sensor-based approach allowed one to avoid including the pose of the vehicle in the dynamics. In some problems, where the nonlinearity occurs in the output

equation, a state augmentation can help to remove the nonlinearity, as was done successfully in [38], where the idea was applied to RO-SLAM. In this paper, the proposed solution relies on an output transformation that leads to a state augmentation, inspired by the results presented in [31].

3.1. State augmentation and output transformation

The objective of this subsection is to obtain a linear-like system that mimics the dynamics of the original nonlinear system while avoiding the nonlinearity on the bearing output. Consider then the following physically sensible mild assumption.

Assumption 1. *The position of the vehicle cannot coincide with a landmark, i.e., a visible bearing vector is always defined.*

With this in mind, the manipulation of the output of (1) yields

$$\mathbf{p}_i(t) - \mathbf{b}_i(t)\|\mathbf{p}_i(t)\| = \mathbf{0}, \quad i \in \mathcal{M}_o. \tag{2}$$

If the norm of the i-th landmark is added as a state, this expression becomes in fact linear. That is the idea behind the augmented state

$$\mathbf{x}_F(t) := \begin{bmatrix} \mathbf{x}_L^T(t) & \mathbf{x}_R^T(t) \end{bmatrix}^T,$$

where $\mathbf{x}_L(t) \in \mathbb{R}^{n_L}$ is the stacking of all landmarks, both visible and non-visible, and $\mathbf{x}_R(t) \in \mathbb{R}^{n_R}$ agglomerates all the norms of the landmarks, i.e., the distance from each landmark to the vehicle. These correspondences are summarized by the state constraints

$$\begin{cases} \mathbf{x}_{L_i}(t) := \mathbf{p}_i(t) \\ x_{R_i}(t) := \|\mathbf{x}_{L_i}(t)\| \end{cases}, \tag{3}$$

for all $i \in \mathcal{M}$, where $\mathbf{x}_{L_i}(t) \in \mathbb{R}^3$ and $x_{R_i}(t) \in \mathbb{R}$ are i-th components of the landmark subscript L) and range (subscript R) state vectors, respectively. Note that both the landmark and range states are composed by visible and non-visible parts, denoted by subscripts O and U respectively.

Consider the derivative of the range state, given by

$$\dot{x}_{R_i}(t) = -\frac{\mathbf{x}_{L_i}^T(t)}{x_{R_i}(t)}\mathbf{v}(t)$$

which is needed to write the full state dynamics. When a landmark is observed and its bearing is available, the quotient $\frac{\mathbf{x}_{L_i}(t)}{x_{R_i}(t)}$ can be replaced by the bearing $\mathbf{b}_i(t)$ for all $i \in \mathcal{M}_o$. Knowing this, the resulting system reads

$$\begin{cases} \dot{\mathbf{x}}_F(t) = \mathbf{A}_F(t)\,\mathbf{x}_F(t) + \mathbf{B}_F(t, \mathbf{b}(t), \mathbf{x}_{L_U}(t), \mathbf{x}_{R_U}(t))\mathbf{v}(t) \\ \mathbf{y}(t) = \mathbf{C}_F(t, \mathbf{b}(t))\,\mathbf{x}_F(t) \end{cases}, \tag{4}$$

where the dynamics matrix is

$$\mathbf{A}_F(t) = \begin{bmatrix} \mathbf{A}_L(t) & \mathbf{0}_{n_L \times n_R} \\ \mathbf{0}_{n_R \times n_L} & \mathbf{0}_{n_R \times n_R} \end{bmatrix}$$

with component $\mathbf{A}_L(t) = -\operatorname{diag}\left(\mathbf{S}\left[\boldsymbol{\omega}(t)\right], \ldots, \mathbf{S}\left[\boldsymbol{\omega}(t)\right]\right)$. The input matrix is given by

$$\mathbf{B}_F(t, \mathbf{b}(t), \mathbf{x}_{L_U}(t), \mathbf{x}_{R_U}(t)) = \begin{bmatrix} \mathbf{B}_L \\ \mathbf{B}_R(t, \mathbf{b}(t), \mathbf{x}_{L_U}(t), \mathbf{x}_{R_U}(t)) \end{bmatrix}$$

where $\mathbf{B}_L = -\begin{bmatrix} \mathbf{I} & \cdots & \mathbf{I} \end{bmatrix}^T \in \mathbb{R}^{n_L \times 3}$ and

$$\mathbf{B}_R(t, \mathbf{b}(t), \mathbf{x}_{L_U}(t), \mathbf{x}_{R_U}(t)) = \\ -\begin{bmatrix} \mathbf{b}_1(t) & \cdots & \mathbf{b}_{N_O}(t) & \frac{\mathbf{x}_{L_{N_O+1}}(t)}{x_{R_{N_O+1}}(t)} & \cdots & \frac{\mathbf{x}_{L_N}(t)}{x_{R_N}(t)} \end{bmatrix}^T.$$

The output matrix is

$$\mathbf{C}_F(t, \mathbf{b}(t)) = \begin{bmatrix} \mathbf{I}_{n_O} & \mathbf{0}_{n_O \times n_U} & \mathbf{C}_b(t, \mathbf{b}(t)) & \mathbf{0}_{n_O \times N_U} \end{bmatrix}, \tag{5}$$

with $\mathbf{C}_b(t, \mathbf{b}(t)) = -\operatorname{diag}\left(\mathbf{b}_1(t), \ldots, \mathbf{b}_{N_O}(t)\right)$. Finally, the output is $\mathbf{y}(t) = \begin{bmatrix} \mathbf{0}_{n_O \times 1} \end{bmatrix}$.

Remark 1. *The measured quantities $\mathbf{b}_i(t)$ are signals external to this augmented system that tie it to the original nonlinear one. They can be seen as inputs, available for observer design purposes, and they provide the information necessary in the output equation, which is identically zero. The observability analysis in the next subsection addresses these issues.*

Even though the output nonlinearity as first brought up disappeared with the state augmentation and output transformation proposed in this section, the process introduced two new non-linearities. The first is on the input matrix, as it depends both on a measured quantity, the bearing, and on the state, when the measurement is not available. The second is on the output matrix that also depends on a measured quantity. However, the presence of the measurement in the input and output matrices is not really a problem, as, for observability purposes, a system whose dynamics matrix depends on known quantities can be seen as a linear time-varying (LTV) system. The presence of the state in the input matrix only affects the non-visible landmarks ($\mathbf{x}_{L_U}(t)$ and $\mathbf{x}_{R_U}(t)$). These are not observable, and therefore will be propagated in open loop.

Another important aspect that must be stressed is the fact that the there is nothing in the augmented system (4) that imposes the constraints (3), particularly the nonlinear relation $x_{R_i}(t) = \|\mathbf{x}_{L_i}(t)\|$, and as such the relation between the nonlinear and the augmented systems must be carefully analysed.

The current state-of-the-art approach in BO-SLAM filters, inverse depth-based algorithms [11], also employs a state augmentation to try to cope with the

nonlinearity in the measurement model. In that case, each feature is represented by a 6-D vector containing the camera coordinates from the first sighting, the inverse depth and the image coordinates. The approach presented in this paper is also not minimal, but has lower dimensionality and ensures that the system is truly linear time-varying which is important for observability, stability, and convergence purposes as explained in the subsequent subsections.

3.2. Observability analysis

The subject of this subsection is the observability analysis of the nonlinear and augmented systems presented previously. The augmented system (4) contains non-visible landmarks and associated ranges that are clearly not observable as the corresponding bearing is not available. Hence, in the observability analysis, the non-visible landmarks and the associated ranges are discarded, following the successful approach first used by the authors in [37] and [28]. Furthermore, since each landmark-range-bearing group is independent of the others, it is possible to simplify the analysis greatly by assuming that only one landmark is visible, i.e., $\mathcal{M}_o := \{1\}$.

The new reduced system is given by

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{v}(t) \\ \mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) \end{cases} \tag{6}$$

where the dynamics matrix is

$$\mathbf{A}(t) = \begin{bmatrix} -\mathbf{S}\left[\boldsymbol{\omega}(t)\right] & \mathbf{0}_{3\times 1} \\ \mathbf{0}_{1\times 3} & 0 \end{bmatrix},$$

the input matrix is

$$\mathbf{B}(t) = \begin{bmatrix} -\mathbf{I}_3 \\ -\mathbf{b}_1^T(t) \end{bmatrix},$$

and the output matrix is

$$\mathbf{C}(t) = \begin{bmatrix} \mathbf{I}_3 & -\mathbf{b}_1(t) \end{bmatrix}.$$

The dependence of the input matrix on the non-visible landmarks and ranges has disappeared with the reduction of the state of this system. However, both the input and the output matrices still depend on the visible bearing. Given that this quantity is a known function of time, the reduced system can be considered as linear time-varying for observability analysis and observer design. In fact, as shown in [43, Lemma 1], if the observability Gramian associated with a system whose dynamics matrix depends on the system input and output is invertible, then the system is observable. This result will be exploited throughout this subsection.

The forthcoming analysis requires the definition of ${}^E\mathbf{b}_1(t) = \mathbf{R}(t)\mathbf{b}_1(t)$ as the Earth-fixed or absolute bearing. This theorem addresses the observability analysis of system (6) regarded as LTV.

Theorem 1. *Take system* (6)*, regarded as LTV, and let* $\mathcal{T} := [t_0, t_f]$*. The system is observable in* $\mathcal{T}$ *if and only if the absolute bearing associated with the visible landmark is not constant in* $\mathcal{T}$*, i.e., there exists a* $t_1 \in \mathcal{T}$ *such that* ${}^{E}\dot{\mathbf{b}}_1(t_1) \neq 0$*.*

Proof. Consider the Lyapunov transformation (see [44] for details)

$$\mathbf{z}(t) = \mathbf{T}(t)\mathbf{x}(t)$$

where $\mathbf{T}(t) = \operatorname{diag}(\mathbf{R}(t), 1)$ preserves the observability properties of the original system. It is a matter of computation to obtain the transformed system, given by

$$\begin{cases} \dot{\mathbf{z}}(t) = \boldsymbol{\mathcal{B}}(t)\mathbf{v}(t) \\ \mathbf{y}(t) = \boldsymbol{\mathcal{C}}(t)\mathbf{z}(t) \end{cases} \tag{7}$$

where the input matrix is

$$\boldsymbol{\mathcal{B}}(t) = \begin{bmatrix} -\mathbf{R}(t) \\ -\mathbf{b}_1^T(t) \end{bmatrix},$$

and the output matrix is

$$\boldsymbol{\mathcal{C}}(t) = \begin{bmatrix} \mathbf{R}^T(t) & -\mathbf{b}_1(t) \end{bmatrix}.$$

The proof, made by contraposition, will follow with the transformed system for simplicity of analysis. This system is assumed not observable, which, using [43, Lemma 1] implies that the observability Gramian is singular. Then it is shown that the conditions of the theorem cannot hold.

Given the nature of (7), it is clear that its transition matrix is $\boldsymbol{\phi}(t, t_0) = \mathbf{I}$. With the transition matrix at hand, the observability Gramian given by

$$\boldsymbol{\mathcal{W}}(t_0, t_f) = \int_{t_0}^{t_f} \left(\boldsymbol{\mathcal{C}}(\tau)\boldsymbol{\phi}(\tau, t_0)\right)^T \boldsymbol{\mathcal{C}}(\tau)\boldsymbol{\phi}(\tau, t_0)\mathrm{d}\tau \tag{8}$$

can now be analysed.

Suppose that the LTV system (6) is not observable, meaning that the transformed system is also not observable, which implies that the observability Gramian (8) is singular. Then, there exists a unit vector $\mathbf{c} := \begin{bmatrix} \mathbf{c}_p^T & c_r \end{bmatrix} \in \mathbb{R}^{nz}$ such that

$$\mathbf{c}^T \boldsymbol{\mathcal{W}}(t_0, t_f)\mathbf{c} = \int_{t_0}^{t_f} \|\mathbf{f}(\tau, t_0)\|^2 \mathrm{d}\tau \tag{9}$$

is zero. In that case, the function $\mathbf{f}(\tau, t_0)$, given by

$$\mathbf{f}(\tau, t_0) = \mathbf{R}^T(\tau)\mathbf{c}_p - \mathbf{b}_1(\tau)c_r, \tag{10}$$

must be zero too for all $\tau \in \mathcal{T}$. The following condition arises after left multi-

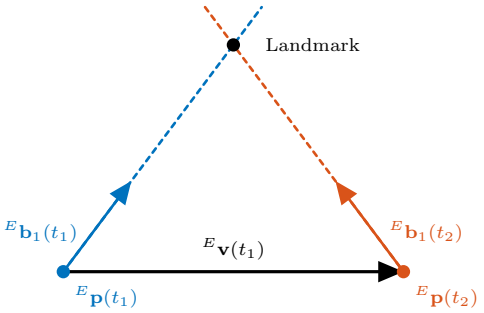


Figure 1: Triangulation for positioning a landmark.

plication of (10) by $\mathbf{R}(\tau)$

$$^E\mathbf{b}_1(\tau)c_r = \mathbf{c}_p, \quad \forall \tau \in \mathcal{T}.$$

From this it follows that $^E\mathbf{b}_1(\tau)$ must be constant for all $\tau \in \mathcal{T}$, which contradicts the conditions of the theorem. Hence, if the system is not observable, the conditions of the theorem cannot hold. Therefore, by contraposition, if the conditions of the theorem do hold, the transformed system (7) is observable, and, as it is related to the LTV system (6) through a Lyapunov transformation, the latter is also observable, thus concluding the first part of the proof.

For the necessity part, assume that the conditions of the theorem do not hold, i.e., $^E\mathbf{b}_1(t) = {}^E\mathbf{b}_1(t_0)$. In that case, if it is possible to find a unit vector $\mathbf{c}$ such that (9) is zero, then the system is not observable, thus proving the necessity of the conditions. Under this assumption, (10) becomes

$$\mathbf{f}(\tau, t_0) = \mathbf{R}^T(\tau)(\mathbf{c}_p - {}^E\mathbf{b}_1(t_0)c_r).$$

Choosing $\mathbf{c}_p = {}^E\mathbf{b}_1(t_0)c_r$ and $c_r = \frac{\sqrt{2}}{2}$, it is clear that $\mathbf{f}(\tau, t_0) = \mathbf{0}$ for all $\tau \in \mathcal{T}$. Then, it is possible to conclude that (9) is zero, which means that the system is not observable. Hence, by contraposition, if the system is observable, the conditions of the theorem must hold, thus concluding the proof of the necessity part of the theorem. □

Remark 2. *The sufficient and necessary condition introduced by this theorem is in fact a requirement on the motion of the vehicle. For the system to be observable, i.e., in order to be possible to obtain the initial condition of a landmark, the trajectory of the vehicle must not be restricted to the line described by the absolute bearing as exemplified in Figure 1.*

This theorem established sufficient and necessary conditions for the observability of the system (6) that is a reduced version of the augmented nonlinear system (4). Given that the discarded states are not observable and do not influence the others, the two systems are equivalent in what concerns observability, when discarding the non-visible landmarks. Hence, this result also applies to

the augmented system with a note of caution: when dealing with multiple landmarks, the observability condition requires that, in the considered time interval, all landmarks must be visible in instants where the vehicle did not travel in their direction. This does not mean that all landmarks should be visible at the same time, but that in the time intervals in which each landmark is visible the conditions of Theorem 1 must hold. As to the original nonlinear system, this observability result cannot be extrapolated without special attention. Recall that although the augmented system (4) mimics the dynamics of the nonlinear one, there is nothing imposing the state relations (3). The sequel addresses this aspect, following the approach in previous works such as [31] and [38].

Theorem 2. *If the conditions of Theorem 1 and Assumption 1 hold, then:*

(i) the state of the original nonlinear system (1) *and that of the LTV system* (6) *are the same and uniquely determined, and the constraints* (3) *are imposed by the dynamics;*

(ii) an observer for the LTV system with globally exponentially stable error dynamics is also a state observer for the underlying nonlinear system with error dynamics that converge exponentially.

Proof. Once more, the proof will focus on the transformed system (7) for simplicity of analysis. The output of this system is $\mathbf{y}(t) = \mathbf{0}$.

Consider the time evolution of the transformed LTV system

$$\mathbf{z}(t) = \boldsymbol{\phi}(t, t_0)\mathbf{z}(t_0) + \int_{t_0}^{t} \boldsymbol{\phi}(t, \tau)\boldsymbol{\mathcal{B}}(\tau)\mathbf{v}(\tau)\mathrm{d}\tau.$$

It is a matter of computation to obtain

$$\mathbf{z}_{L_1}(t) = \mathbf{z}_{L_1}(t_0) - \boldsymbol{\Delta}\mathbf{p}(t, t_0) \tag{11}$$

where $\boldsymbol{\Delta}\mathbf{p}(t, t_0) = \int_{t_0}^{t} \mathbf{R}(\tau)\mathbf{v}(\tau)\mathrm{d}\tau$, and

$$z_{R_1}(t) = z_{R_1}(t_0) - \int_{t_0}^{t} \mathbf{b}_1^T(\tau)\mathbf{v}(\tau)\mathrm{d}\tau. \tag{12}$$

Noting that $\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{p}_1(t)\| = -\mathbf{b}_1^T(t)\mathbf{v}(t)$, it is also possible to write

$$\mathbf{R}(t)\mathbf{p}_1(t) = \mathbf{R}(t_0)\mathbf{p}_1(t_0) - \boldsymbol{\Delta}\mathbf{p}(t, t_0) \tag{13}$$

and

$$\|\mathbf{p}_1(t)\| = \|\mathbf{p}_1(t_0)\| - \int_{t_0}^{t} \mathbf{b}_1^T(\tau)\mathbf{v}(\tau)\mathrm{d}\tau. \tag{14}$$

From the output of the transformed system (7) and using (11) and (12), it follows that

$$\mathbf{R}^T(t)\mathbf{z}_{L_1}(t_0) - \mathbf{R}^T(t)\boldsymbol{\Delta}\mathbf{p}(t,t_0) - \mathbf{b}_1(t)z_{R_1}(t_0) - \mathbf{b}_1(t)\int_{t_0}^{t} \mathbf{b}_1^T(\tau)\mathbf{v}(\tau) = \mathbf{0},$$

which, left-multiplying by $\mathbf{R}(t)$ and using the inverse transformation $\mathbf{x}(t) = \mathbf{T}^{-1}(t)\mathbf{z}(t)$, can be manipulated to yield

$$\mathbf{R}(t_0)\mathbf{x}_{L_1}(t_0) - \boldsymbol{\Delta}\mathbf{p}(t,t_0) - {}^{E}\mathbf{b}_1(t)x_{R_1}(t_0) - {}^{E}\mathbf{b}_1(t)\int_{t_0}^{t} \mathbf{b}_1^T(\tau)\mathbf{v}(\tau) = \mathbf{0}.$$

Finally, recalling that the output of the original nonlinear system (1) can be expressed as (2) if Assumption 1 holds, and using the explicit relations of the nonlinear state with the initial conditions in (13) and (14), it is possible to write

$$\mathbf{R}(t_0)\mathbf{p}_1(t_0) - \boldsymbol{\Delta}\mathbf{p}(t,t_0) - {}^{E}\mathbf{b}_1(t)\|\mathbf{p}_1(t_0)\| - {}^{E}\mathbf{b}_1(t)\int_{t_0}^{t} \mathbf{b}_1^T(\tau)\mathbf{v}(\tau) = \mathbf{0}.$$

The comparison of these two expressions yields

$$\mathbf{R}(t_0)\left(\mathbf{x}_{L_1}(t_0) - \mathbf{p}_1(t_0)\right) = {}^{E}\mathbf{b}_1(t)\left(x_{R_1}(t_0) - \|\mathbf{p}_1(t_0)\|\right),$$

for all $t \in \mathcal{T}$. If the conditions of the theorem hold, then ${}^{E}\mathbf{b}_1(t)$ is not the same for all time, and thus both the left-hand and right-hand sides must be zero, which implies that

$$\mathbf{x}_{L_1}(t_0) = \mathbf{p}_1(t_0)$$

and

$$x_{R_1}(t_0) = \|\mathbf{p}_1(t_0)\|.$$

The initial conditions of both systems are proven to be the same, and, as their dynamics are the same by construction, the states do indeed correspond to each other and the state constraints (3) are imposed by the dynamics when the conditions of the theorem hold, hence concluding the first part of the proof.

The proof of the second part of the theorem follows directly from the first. As the state of the LTV system (6) corresponds to that of the underlying non-

linear system when discarding the non-visible landmarks if the conditions of the theorem hold, the estimates of an observer with globally exponentially stable error dynamics will converge exponentially fast not only to the true LTV state but also to the state of the nonlinear system (1). □

The two previous results address the observability of the LTV system and the correspondence between the states of that and of the original nonlinear system. Combining the two results, it is possible to determine in which conditions the nonlinear system is observable, when the non-visible landmarks are discarded. The following theorem addresses this issue.

Theorem 3. *The nonlinear system* (1) *is observable, when discarding the non-visible landmarks, if and only if the conditions of Theorem 1 and Assumption 1 hold.*

Proof. The sufficiency part of the proof is readily provided by Theorems 1 and 2. The former establishes conditions for the observability of the LTV system (6) and the latter relates the state of that system to that of the nonlinear system in analysis. For the necessity part of the proof, note that if, for the same input, there exist two initial conditions that lead to the same output at all times, then the system is not observable. Assume then that the conditions of the theorem do not hold, and recall the output of the nonlinear system (1). Given that ${}^{E}\mathbf{b}_1(t)$ is constant, the vehicle can only move in the line defined by its initial position and the landmark, which means that any initial state corresponding to a landmark defined in that line will yield the same output, thus implying that the system is not observable when the conditions do not hold, or, conversely, if the system is observable, the conditions must hold. □

3.3. Filter design

The results of the previous subsection show that, in certain conditions with physical insight, the augmented system is equivalent to the nonlinear system, and that if a filter with GES error dynamics can be constructed to the LTV system, it will also be applicable to the original nonlinear system. It can be shown that the error dynamics of the Kalman filter for linear time-varying systems are globally exponentially stable if the pair $(\mathbf{A}(t), \mathbf{C}(t))$ is uniformly completely observable, a form of observability stronger than the ones addressed previously. This result can be achieved following the steps in [45, Example 8.11] and [46]. This last theorem addresses the uniform complete observability of the LTV system.

Theorem 4. *Let* $\mathcal{T}_\delta := [t, t+\delta]$. *The pair* $(\mathbf{A}(t), \mathbf{C}(t))$ *associated with the system* (6)*, regarded as LTV, is uniformly completely observable if there exist positive constants* δ *and* α_b *such that, for all* $t \geq t_0$*, it is possible to find a* $t_1 \in \mathcal{T}_\delta$ *for which the absolute bearing to the visible landmark respects*

$$\left\| \int_t^{t_1} {}^{E}\dot{\mathbf{b}}_1(\tau) \mathrm{d}\tau \right\| \geq \alpha_b. \tag{15}$$

Proof. The proof, provided in Appendix AppendixA, follows similar steps to the proofs of Theorem 1, but considering uniform bounds for all $t \geq t_0$ and intervals $[t, t+\delta]$. □

A Kalman filter for LTV systems can now be implemented for (4), and it is done so in its discrete version. Considering additive disturbances, the discretized system for time-steps of length T_s is given by

$$\begin{cases} \mathbf{x}_{F_{k+1}} = \mathbf{F}_{F_k}\mathbf{x}_{F_k} + T_s\mathbf{B}_{F_k}\mathbf{v}_k + \boldsymbol{\xi}_k \\ \mathbf{y}_{k+1} = \mathbf{C}_{F_{k+1}}\mathbf{x}_{F_{k+1}} + \boldsymbol{\theta}_{k+1} \end{cases}, \tag{16}$$

where the dynamics matrix is

$$\hat{\mathbf{F}}_{F_k} = \begin{bmatrix} \mathbf{F}_{L_k} & \mathbf{0}_{n_L \times n_R} \\ \mathbf{0}_{n_R \times n_L} & \mathbf{I}_{n_R} \end{bmatrix}, \tag{17}$$

with $\mathbf{F}_{L_k} = \text{diag}\left(\mathbf{R}_{k+1}^T\mathbf{R}_k, \ldots, \mathbf{R}_{k+1}^T\mathbf{R}_k\right)$ and $\mathbf{R}_{k+1}^T\mathbf{R}_k = \mathrm{e}^{-\mathbf{S}[\boldsymbol{\omega}_k]T_s}$. The vectors $\boldsymbol{\xi}_k$ and $\boldsymbol{\theta}_k$ represent the model disturbance and measurement noise, respectively. They are assumed to be zero-mean discrete white Gaussian noises with covariances $\boldsymbol{\Xi}_k$ and $\boldsymbol{\Theta}_k$. This is an approximation of the real noise descriptions, since the presence of the angular velocity in $\mathbf{F}_{L_k}$ and of the measured bearings in $\mathbf{B}_{F_k}$ introduces multiplicative terms. The prediction and update equations are the standard LTV Kalman filter equations [47], with the detail that the non-visible landmarks must be propagated in open loop and that the optimality of the filter is lost due to the noise approximation above. However, the convergence and stability properties are maintained.

4. Algorithm implementation

This section addresses the practical implementation of the filter introduced in the previous section. Aside from the Kalman filter that serves as the estimation engine of the algorithm, there are several other relevant components that need to be discussed. Given that the primary exteroceptor in the proposed implementation is a camera, and only natural landmarks are being dealt with, there needs to exist a feature detection process and a consequent landmark association stage. Therefore, this section starts by explaining how the algorithm extracts bearings from the images provided by the camera, bearings that are then used in the update step of the Kalman filter.

4.1. Obtaining bearing measurements from a camera

The first step in obtaining bearing measurements to landmarks from an image is *feature detection*. This step takes place every time a new image is available from the installed camera. This image is fed to a SURF implementation [22], which detects points-of-interest (features) on 2-D pictures of the environment. Aside from the measured feature locations in image coordinates $\mathbf{f}_{m_i}$, SURF also

provides a metric to determine which features have higher *quality* (the hessian), a characteristic that distinguishes types of features (the laplacian), and a 64-dimensional descriptor for association. After feature detection, some processing is necessary before the insertion of these measurements in the filter. Note, however, that this processing serves only the purpose of transforming the features from image coordinates to bearings in body-fixed coordinates, as the landmarks are initialized undelayed.

Transforming image coordinates to bearings. This step depends greatly on the used camera model. This work uses a model [48] that relates the undistorted normalized coordinates $\mathbf{f}_i \in \mathbb{R}^2$ with the measured coordinates, given by

$$\begin{cases} \mathbf{f}_{d_i} = \left(1 + k_c^1 \|\mathbf{f}_i\|^2 + k_c^2 \|\mathbf{f}_i\|^4 + k_c^5 \|\mathbf{f}_i\|^6\right) \mathbf{f}_i + \mathbf{d}_i \\ \mathbf{d}_i = \begin{bmatrix} 2k_c^3 x_i y_i + k_c^4 \left(\|\mathbf{f}_i\|^2 + 2x_i^2\right) \\ 2k_c^4 x_i y_i + k_c^3 \left(\|\mathbf{f}_i\|^2 + 2y_i^2\right) \end{bmatrix} \\ \mathbf{f}_{m_i} = \begin{bmatrix} f_c^1 & \alpha_c f_c^1 & c_c^1 \\ 0 & f_c^2 & c_c^2 \end{bmatrix} \begin{bmatrix} \mathbf{f}_{d_i} \\ 1 \end{bmatrix} \end{cases} . \tag{18}$$

where $\mathbf{f}_{d_i} \in \mathbb{R}^2$ are the distorted feature coordinates, $\mathbf{d}_i$ is the tangential distortion, and the remaining quantities are the intrinsic model parameters: f_c^j is the camera focal length on each axis $j = 1, 2$; the c_c^j, with $j = 1, 2$, are the principal point coordinates; k_c^j, for $j = 1, \ldots, 5$, are the distortion coefficients; and α_c is the skew coefficient.

This model was chosen because it is the one used by the authors of the *Rawseeds* dataset [33, 34] from which the data for the experiments detailed in this paper was obtained. To solve (18) for the normalized feature coordinates, consequently removing distortion and projection, the calibration toolbox [49] is employed.

From the normalized feature locations, given by

$$\mathbf{f}_i = \begin{bmatrix} \frac{x_i}{z_i} & \frac{y_i}{z_i} \end{bmatrix}^T ,$$

it is possible to compute a normalized range,

$$\bar{r}_i = \sqrt{\|\mathbf{f}_i\|^2 + 1} = |z_i|^{-1} \sqrt{x_i^2 + y_i^2 + z_i^2}.$$

Noting that z_i is always positive due to the limited field of view of the camera, it can be seen that this is also equal to the norm of the three-dimensional point normalized by the z coordinate. Careful observation of the information available, $(\mathbf{f}_i, \bar{r}_i)$, it is clear that it is possible to compute directions to the

landmarks, i.e., $\frac{\mathbf{p}_i}{\|\mathbf{p}_i\|}$, which are then given by

$$\mathbf{b}_i = \frac{\mathbf{p}_i}{\|\mathbf{p}_i\|} = \begin{bmatrix} \frac{x_i}{z_i} \\ \frac{y_i}{z_i} \\ \frac{z_i}{z_i} \end{bmatrix} \frac{z_i}{\|\mathbf{p}_i\|} = \begin{bmatrix} \mathbf{f}_i \\ 1 \end{bmatrix} \frac{1}{\bar{r}_i}.$$

Following this process, the remaining step before the actual update step of the filter is landmark association, the subject of the sequel.

4.2. Landmark association

In typical bearing-only algorithms, data association is sometimes not an issue, as the measurements originate from identified acoustic or RF signals. However, when monocular vision is utilized to obtain the bearing measurements, as is the case here, association becomes extremely important. Due to this importance, several algorithms stemming from a fertile research effort in the field have appeared. This is not the focus of this paper and any method could be employed. A brief description of the chosen methodology is included for the sake of completeness.

The first step in this process is to define the sets to associate with. A maximum number N_{max} of observed features is defined and the N_{max} features with highest metric are chosen. Then, the set of landmarks in the current map to associate is defined as the ones within the field-of-view (FOV) of the camera (only azimuth and elevation are checked). Both sets are divided in two, depending on the sign of the Laplacian of each feature/landmark.

The next step avails itself of the possibilities permitted by the image processing procedure used, namely the description of the landmarks by labels invariant and robust to several transformations. This follows a previous successful implementation by the authors in [28], where the landmarks are first associated with respect to their descriptor, with individual compatibility gating based on the Mahalanobis distance of the measurement (2). The process is summarized in the following steps: (i) the filtered landmarks inside the predicted FOV are selected for possible association; (ii) the observed features and selected landmarks are divided according to their Laplacian; (iii) the differences between descriptors for all combinations of pairs of observed-filtered landmarks are computed; (iv) the pair with the lowest difference is selected and its Mahalanobis distance [29], given by

$$m_{ij} = (\mathbf{y}_{i_k} - \mathbf{C}_{i_k}\hat{\mathbf{x}}_k)^T \left(\mathbf{C}_{i_k}\mathbf{\Sigma}_k\mathbf{C}_{i_k}^T + \mathbf{\Theta}_{i_k}\right)^{-1} (\mathbf{y}_{i_k} - \mathbf{C}_{i_k}\hat{\mathbf{x}}_k) \sim \chi_3^2,$$

is computed; (v) the Mahalanobis distance is compared with the 95% percentile of the χ^2 distribution with three degrees of freedom. An association is considered valid if the distance is below the threshold, and the feature is considered a new landmark if it is not. (vi) the pair is removed from the possible associations

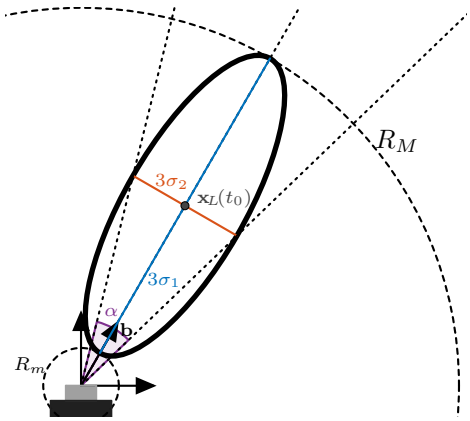


Figure 2: A schematic of the construction of the initial covariance. R_m and R_M are the minimum and maximum range of the camera, α is the noise error, and σ_i^2 are the eigenvalues of the new covariance.

and steps (iv) to (vi) are repeated until there are no features/landmarks left to associate.

Given that the number of features detected by SURF can be very high, and to avoid the size of the filter state getting prohibitively high, after the association procedure only the associated bearings are directly used as observations in the filter. All non-associated measurements are discarded as long as the total number of *valid* bearings is greater than a predefined threshold. If it is not, then the best new landmarks are added until the threshold is passed. The next step is landmark initialization. The guaranteed convergence of the filter frees the user from careful considerations on how to initialize each landmark in the filter. A possible method is depicted in Figure 2, and includes the following steps: (i) define a minimum (R_m) and maximum range (R_M) for the camera; (ii) define a circle of error in the image plane that is equivalent to a cone in 3-D space with an aperture angle of α centered on the measured bearing; and (iii) approximate a Gaussian distribution to this volume using the ellipsoid corresponding to the 3σ uncertainty bound, i.e., $\hat{\mathbf{x}}_{L_i}(k_0) := \frac{R_M+R_m}{2}\mathbf{b}_i(k_0)$ and $\mathbf{\Sigma}_{L_i}(k_0) := \mathbf{U}(\mathbf{b}_i(k_0))\,\mathrm{diag}((\frac{1}{6}(R_M - R_m))^2, (\frac{1}{6}\|\hat{\mathbf{x}}_{L_i}(k_0)\|\sin\alpha)^2, (\frac{1}{6}\|\hat{\mathbf{x}}_{L_i}(k_0)\|\sin\alpha)^2)\mathbf{U}(\mathbf{b}_i(k_0)^T$, where $\mathbf{U}(\mathbf{b}_i(k_0))$ is the rotation matrix that aligns the ellipsoid with the bearing measurement. In the particular implementation of the filter used in the experimental results of the next section, the above method was used with the detail that the center of the ellipsoid was placed in a random point in the line defined by $\mathbf{b}_i(k_0)$, to better showcase the convergence of the algorithm.

4.3. Map maintenance

Even though the association process is designed to try to reduce the number of landmarks in the state, the process employed is rather naïve. Furthermore,

it favours the inclusion of spurious measurements, when the non-associated features are inserted in the filter. This can lead to a very large state with a great number of spurious features. To deal with this issue, a simple test was implemented to check whether a landmark is spurious or not, depending on two counters: (i) a visibility counter that is increased each time a landmark is associated with a new measurement; and (ii) a counter that is increased each iteration that a landmark is not associated and reset every time it is, thus counting the time without observations. When the second counter reaches $T_{\max}$ and the first counter is lower than $N_{\min}$, the landmark is considered spurious and discarded.

4.4. Complexity reduction

As mentioned previously, the size of the filter state can get very high very fast which will eventually slow down the algorithm. As the filter operates in the body-fixed frame, all the non-visible landmarks must be propagated in each step according to the motion of the vehicle, This can be very time consuming for too large a state. This is in opposition to the traditional inertial SLAM approach, where the prediction step is trivial (the landmarks are static) but the update step is very computationally expensive as all landmarks (and their covariances) need to be updated.

Since the dynamics of each landmark (and its associated range state) are independent of each other for observability purposes, it can be thought that that independence will be apparent in the stochastic filtering setting. The following result shows in which conditions that insight is correct.

Lemma 1. *Consider the Kalman filter for the linear time-varying discrete system* (16)*. If the initial covariance* $\mathbf{\Sigma}_0$ *and the process covariance* $\mathbf{\Xi}_k$ *for all* $k > 0$ *have the following structure*

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{L_1} & \cdots & \mathbf{0} & \mathbf{M}_{L_1 R_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{M}_{L_N} & \mathbf{0} & \cdots & \mathbf{M}_{L_N R_N} \\ \mathbf{M}_{R_1 L_1} & \cdots & \mathbf{0} & M_{R_1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{M}_{R_N L_N} & 0 & \cdots & M_{R_N} \end{bmatrix}. \tag{19}$$

and the measurement covariance $\mathbf{\Theta}_k$ *is block-diagonal, then both the predicted covariance* $\mathbf{\Sigma}_{k+1|k}$ *and the updated covariance* $\mathbf{\Sigma}_{k|k}$ *maintain the structure in* (19) *for all* $k > 0$*.*

Proof. The proof is focused in showing that at a given instant $k > 0$ the update and predict steps of the Kalman filter do not change the structure of the covariance if the covariance at $k-1$ already had the structure of (19). It can be found in AppendixB. □

Using this result, if the model parameters $\mathbf{\Sigma}_0$, $\mathbf{\Theta}_k$, and $\mathbf{\Xi}_k$ are chosen accordingly it is possible to simplify the computations by exploiting the sparsity of the covariance matrix, as only a few entries need to be computed, yielding, for all $i \in \mathcal{M}$,

$$\begin{cases} \mathbf{\Sigma}_{L_{i\,k+1|k}} = \mathbf{R}_{k+1}^T \mathbf{R}_k \mathbf{\Sigma}_{L_{i\,k|k}} \mathbf{R}_k^T \mathbf{R}_{k+1} + \mathbf{\Xi}_{L_{i\,k+1|k}} \\ \mathbf{\Sigma}_{R_{i\,k+1|k}} = \mathbf{\Sigma}_{R_{i\,k|k}} + \mathbf{\Xi}_{R_{i\,k+1|k}} \\ \mathbf{\Sigma}_{R_i L_{j\,k+1|k}} = \mathbf{\Sigma}_{R_i L_{j\,k|k}} \mathbf{R}_k^T \mathbf{R}_{k+1} + \mathbf{\Xi}_{R_i L_{j\,k+1|k}} \end{cases} .$$

If it is necessary to include more cross-covariance terms, for example, the structure

$$\hat{\mathbf{\Sigma}}_{k+1|k} = \begin{bmatrix} \mathbf{\Sigma}_{L_{1\,k+1|k}} & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{\Sigma}_{LR_{k+1|k}} \\ \mathbf{0} & \cdots & \mathbf{\Sigma}_{L_{N\,k+1|k}} & \\ & \mathbf{\Sigma}_{RL_{k+1|k}} & & \mathbf{\Sigma}_{R_{k+1|k}} \end{bmatrix}$$

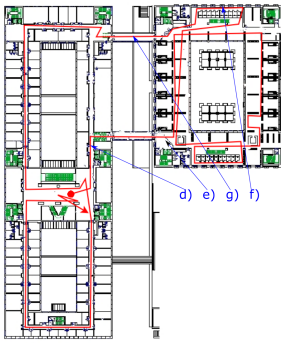
can still be assumed for the propagated covariance. Due to the presence of the noisy linear velocity in the dynamics for all the landmarks, there should exist cross-covariance terms in $\mathbf{\Xi}_k$. However, it is argued that sacrificing the exactness of the model allows the computation time to decrease greatly, given that the prediction step is the greatest contributor for the computational cost. This behaviour becomes more apparent especially for large maps, such as the one presented in the following section. In this particular unoptimized implementation, designed to demonstrate experimentally the performance and convergence properties of the sensor-based BO-SLAM filter, the algorithm scales with N^2.

5. Experimental results

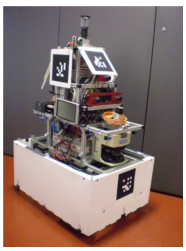
The theoretical results presented in Section 3 lead to the design of a BO-SLAM filter implemented as described in the previous section. This algorithm was tested with real data from datasets acquired by the *Rawseeds* Project [33, 34]. This section covers the relevant details of the dataset used and presents the results of the experiments, while also providing a discussion of the latter.

5.1. Rawseeds dataset

From the available datasets in the *Rawseeds* project, for the purpose of validating the proposed algorithm, an indoors dataset with natural lighting and no purposed dynamic objects, named `Bicocca_2009-02-25b`, was chosen. In this dataset, the traversed path of 774 m includes several small loops, aside from the main loop which is covered by the robot in 29 minutes. The top view of the schematic of the area and the sketched trajectory can be found in Figure 3 along with pictures of several regions in the area illustrating some of the difficulties in visual navigation, namely corridors without many distinguishable features, very dark places, very bright places, within others.



(a) A schematic of the top view of the area travelled by the robot in the dataset.



(b) The robot used in experiments.

(c) The interior of the university library (the building on the right).

(d) A typical corridor.

(e) A very bright elevator lobby.

(f) Another typical corridor with clear repetition of features.

(g) One of the two very dark corridors between the two buildings.

Figure 3: Information on the dataset. Schematic of the area and the trajectory, the robot used in the experiments and examples of places visited by the robot.

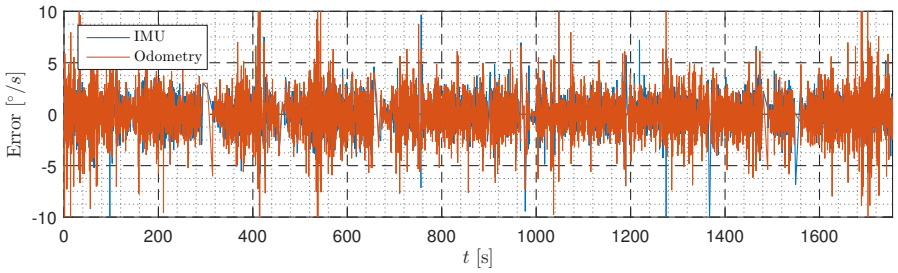


Figure 4: Inertial measurement unit and odometry-based angular velocities error evolution.

The robot used as a data acquisition platform for the dataset, depicted in Figure 3(b), is equipped with a very complete sensor suite, comprising 3 *Videre Design* black and white cameras in a forward-looking trinocular configuration, a *Prosilica* camera with an hyperbolic mirror to achieve omni-directional vision, a SONAR belt, an *Unibrain* low-cost camera, a *Maxbotix* SONAR belt, an *Xsense* inertial measurement unit (IMU), two double laser range finder configurations to cover the front and rear of the vehicle with both a cheaper solution by *Hokuyo* and a long range one by *Sick*, and wheel encoders for odometry. Aside from the various streams of data provided by the sensors on board the robot, the dataset also provides ground truth from a system based on industrial cameras, visual tags mounted on the robot, and ad hoc software in select parts of the trajectory. This ground truth is complemented by what is called *extended* ground truth, available for all the trajectory, which is computed using scan matching from the laser scanners in the robot.

The BO-SLAM algorithm herein proposed uses only directions, linear, and angular velocities, therefore only one of these cameras, odometry and the inertial measurement unit are necessary. It is possible to obtain the angular velocity (or angular displacements) from both the wheel encoders and the inertial measurement unit, and therefore it is necessary to decide which of the sources to use. For that purpose, the difference between both angular velocities and the one derived from the extended ground truth was analysed. As can be seen in Figure 4, even though the error has zero mean, it reaches consistently very high levels, with standard deviations of 1.398°/s (IMU) and 2.087°/s (odometry). This results in a very distorted trajectory, when dead-reckoning the body-fixed linear velocity provided by the odometry and either angular velocities. Bearing-only (and range-only) SLAM algorithms rely greatly on the quality of the measurements that drive them, in this case the linear and angular velocities. This is particularly relevant when the field of view of the camera is limited to a region in the front of the vehicle, because, when turning, the camera will quickly lose track of features that help correct motion information. The available benchmark solution to monocular SLAM provided in the dataset [50], based on the algorithms in [11] and [51] shows precisely both how the monocular camera filtering results can be poor, due to the limited field-of-view and the large distance, and how the odometry is extremely deficient in turns. For these reasons, the authors

decided to use an artificial measurement of angular velocity, taking the measurement provided by the scan matching procedure and adding artificial noise with a realistic standard deviation of 0.15°/s – obtained from the data in the experiments detailed by the authors in [28] and [38]. This noise level can easily be found nowadays in off-the-shelf IMUs like the *Microstrain 3DM-GX3-25*.

5.2. The algorithm at work

The bearing-only SLAM algorithm herein presented is a sensor-based filter, and, as such, its main product is the landmark map in the body-fixed frame. Furthermore, in the experiments to be detailed, these landmarks are naturally extracted from the environment and not artificially placed in predefined position. Therefore, there is not a clear ground truth to validate directly the results of the sensor-based algorithm, and the map can only be evaluated quantitatively through the architectural drawings of the area. For that purpose, consider Figure 5. There is depicted the estimated map, in black, with the architectural drawings in the background. From this comparison, it can be seen that the resulting map clearly fits the architectural drawings, and, shows the convergence of the landmark state which was, as explained throughout the paper, initialized without any special care. Even though some landmarks are too far away for not having had enough time to converge, either because they were only observed too few frames or because the association process did not recognize them, they help demonstrate the convergence, as they contrast with the other landmarks that fulfilled all the requirements for convergence. It is also interesting to remark that both corridors that connect the two buildings have more dispersed landmarks that the rest of the building. This may be due to the fact that both corridors are very dark and that mostly only distant lights from the neighbouring buildings are detected as landmarks (see Figure 3(g)). Furthermore, the bottom corridor that connects the two buildings has fewer (and more dispersed) landmarks than the top one, which can be explained by the fact that the robot circulates around the bottom left part of the library facing the corridor twice, and falsely loop-closed landmarks from the area of the corridor are re-estimated as being part of the building, thus moving away from the corridor. To further validate the performance of the filter, even if indirectly, the sensor-based map produced by the Kalman filter was used to obtain the pose of the vehicle, through the Earth-fixed trajectory and map (ETM) estimation procedure also used in [28]. The result of this methodology is depicted in blue in Figure 6, along with extended ground truth in yellow and the dead-reckoned odometry path in red. This shows that the sensor-based map in conjunction with the ETM procedure is able to partially correct the error introduced by the odometry, its norm being kept under 1 meter for most of the run. Note that these results do not represent directly the performance of the sensor-based BO-SLAM filter proposed in this paper, but do help on its evaluation. Finally, the spatial evolution of a selected set of landmarks rotated to the Earth-fixed frame using the ground truth is shown in Figure 7. This enables to better visualize the convergence of the filter and effects of repeated sightings. Each small blob represents the sampled history (every 20 seconds) of a landmark, from very light colours (old) to stronger (more recent).

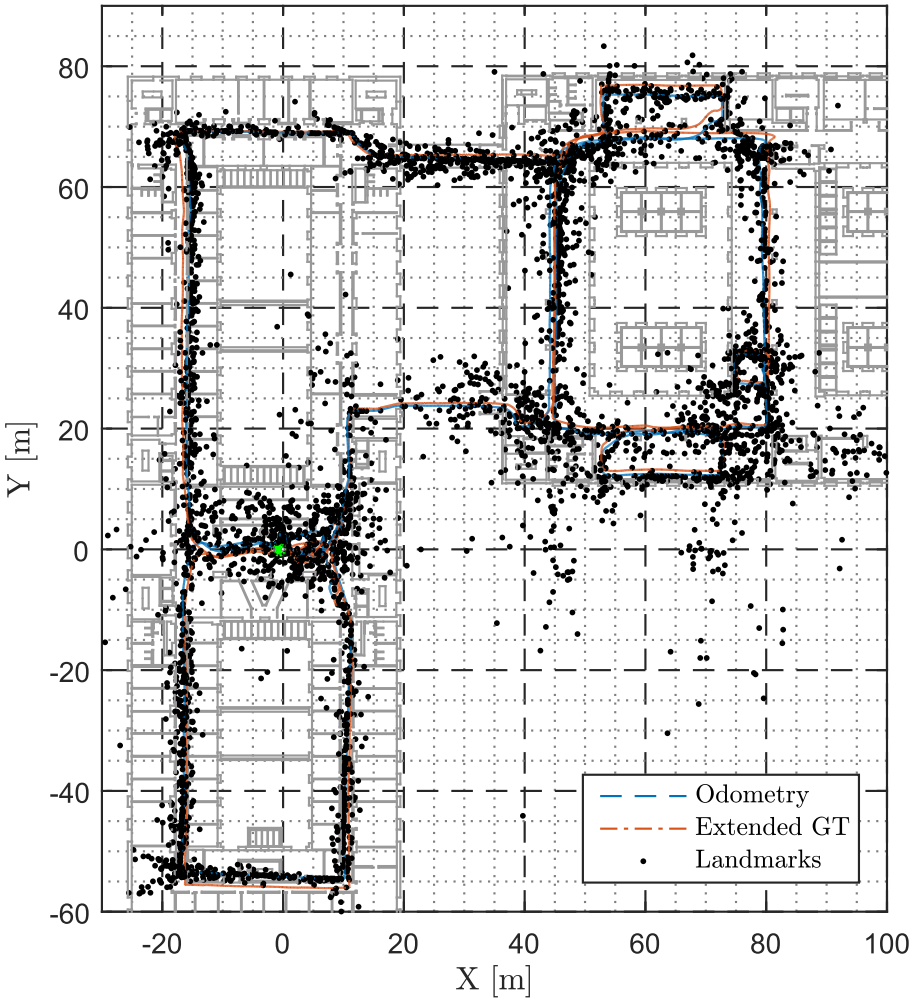


Figure 5: The estimated map at the end of the experiment, along with the dead-reckoned path and extended ground-truth.

With this figure it is possible to evaluate the dispersion of the estimates through time and space, and assess zones where convergence was slower or erroneous associations were made. For instance, the corridor at around $Y = 25$ m shows a blur of grey that is due to the landmarks therein slowly converging to their final positions.

In a sensor-based approach, all the processes exogenous to the filter, that build upon its results or upon which the filter relies, can avoid the inherent nonlinearity that stems from the transformation of the naturally body-fixed

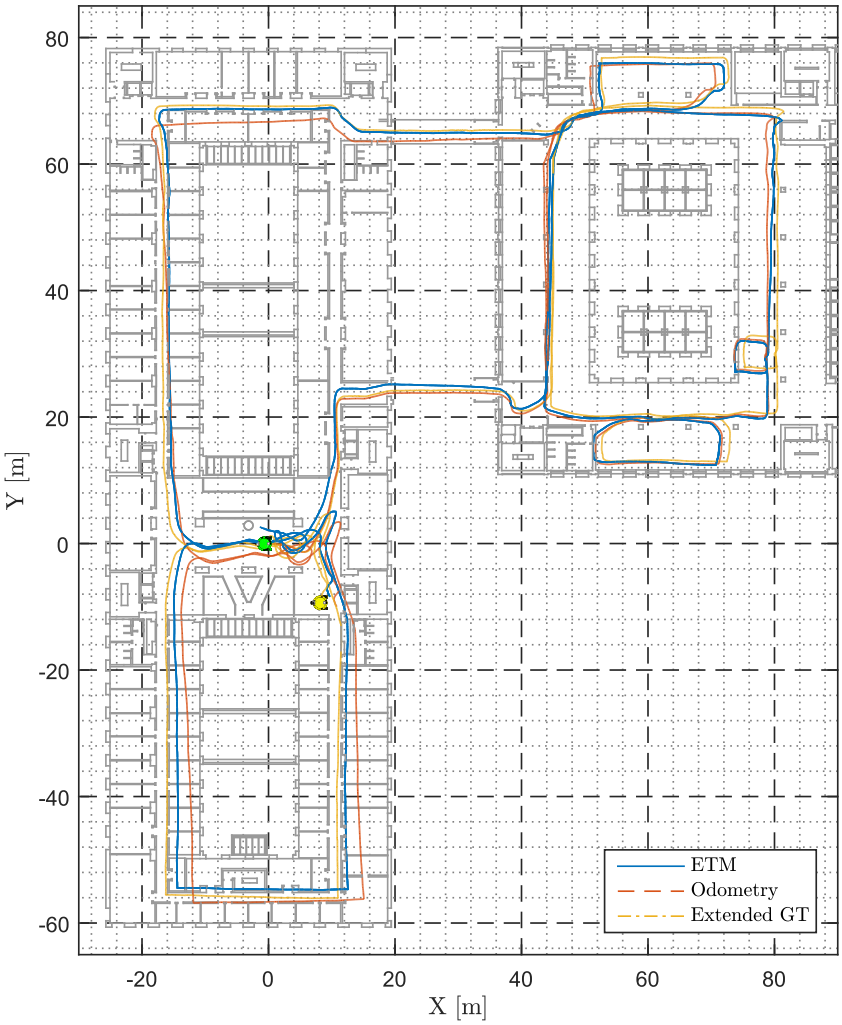


Figure 6: The estimated path at the end of the experiment, along with the dead-reckoned path and extended ground-truth.

measurements to an inertial frame. For example, by performing landmark association, loop-closing, control, and decision procedures in the sensor-based frame, it is possible to reduce the effects of nonlinearities in the consistency of the filter. Consistency is a very important aspect of any estimation algorithm, and simultaneous localization and mapping is not an exception. This can be evaluated by checking how the uncertainty indicated by the algorithm compares to the actual estimation error. When this approach cannot be pursued, there are other ways to evaluate it. One of the best measures of the consistency and

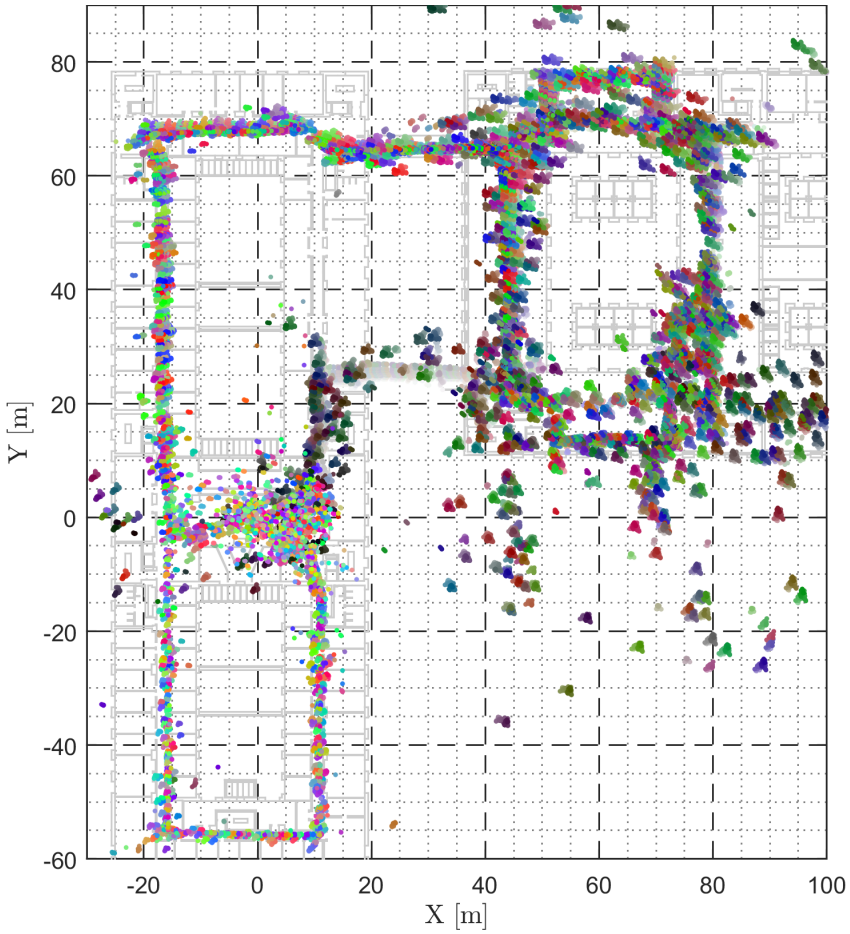
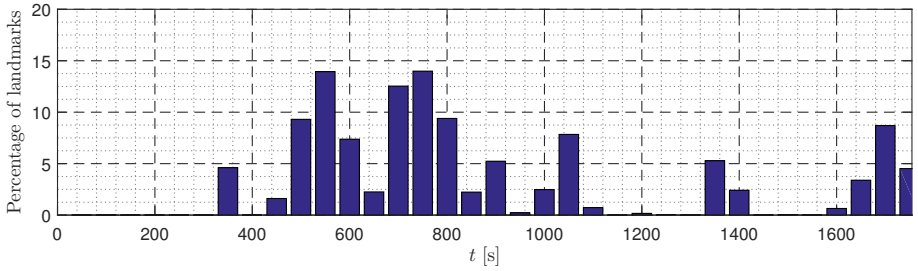
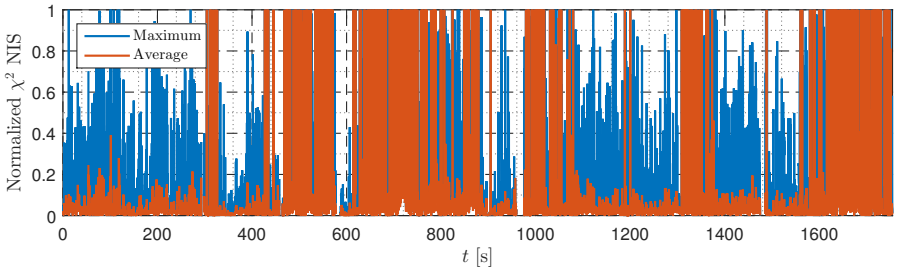


Figure 7: The estimated map throughout the whole of the experiment, with the spatial evolution of each landmark shown in intervals of 20 seconds.

validity of a SLAM algorithm is the way it handles loop closures, i.e., if it does recognize previously visited places. The algorithm proposed in this filter does not include a specially tailored loop closing procedure, even though one of the many existent in the literature that avail themselves of other techniques such as place recognition from images can be applied on top of the proposed BO-SLAM filter. In this case, loop closures will occur naturally from association of new observations with old landmarks that fit the association criteria. Even though association/loop closing is based on the image descriptors, the search for associations is limited to the field-of-view predicted by the BO-SLAM filter as explained in Section 4.2. Furthermore, the decision to associate or add a new landmark depends on the Mahalanobis distance of the measurement. Hence, the loop closings depend both on image and filter information. Loop closures



(a) Percentage of loop-closed landmarks.



(b) Normalized innovation square.

Figure 8: Algorithm performance statistics.

can be clearly observed when there is a sharp reduction of the uncertainty of a landmark after some time increasing. For the purpose of providing a better visualization of this process and its results, the landmarks re-observed after a long time, henceforth denoted as loop-closed landmarks, were bundled in slots of 50 seconds as shown in Figure 8(a). The number of loop-closed landmarks is very large around the 600-800 seconds mark, which corresponds to the time when the robot revisits the bottom part of the building in the right side of the map. Then there are peaks around the 1000 – the robot revisits the top left corner of the right-side building, 1400 – the robot revisits the starting place, and 1600 seconds marks – the robot revisits the leftmost corridor of the map. Details about all these can be seen in Figures 9 and 10. The maps in both figures depict the maps at the end of each of the chosen time slots, along with the loop-closed landmarks highlighted in red. The pose of the robot at the start of the time slot is represented by the green solid circle, and the yellow circle stands for the pose of the robot at the end of the time slot. In Figure 9 frames of the camera stream used in the filter are also presented. The bottom ones are the frames which contain the most loop-closed landmarks of the corresponding time slot, and the top ones are the frames where more of the loop-closed landmarks were observed the first time. The red circles represent detected features, the green circles mark the features selected for the filter, and the blue circles depict the features that were associated with existing landmarks. The evenly dotted circles represent loop-closed landmarks in the actual frame, and the other dotted

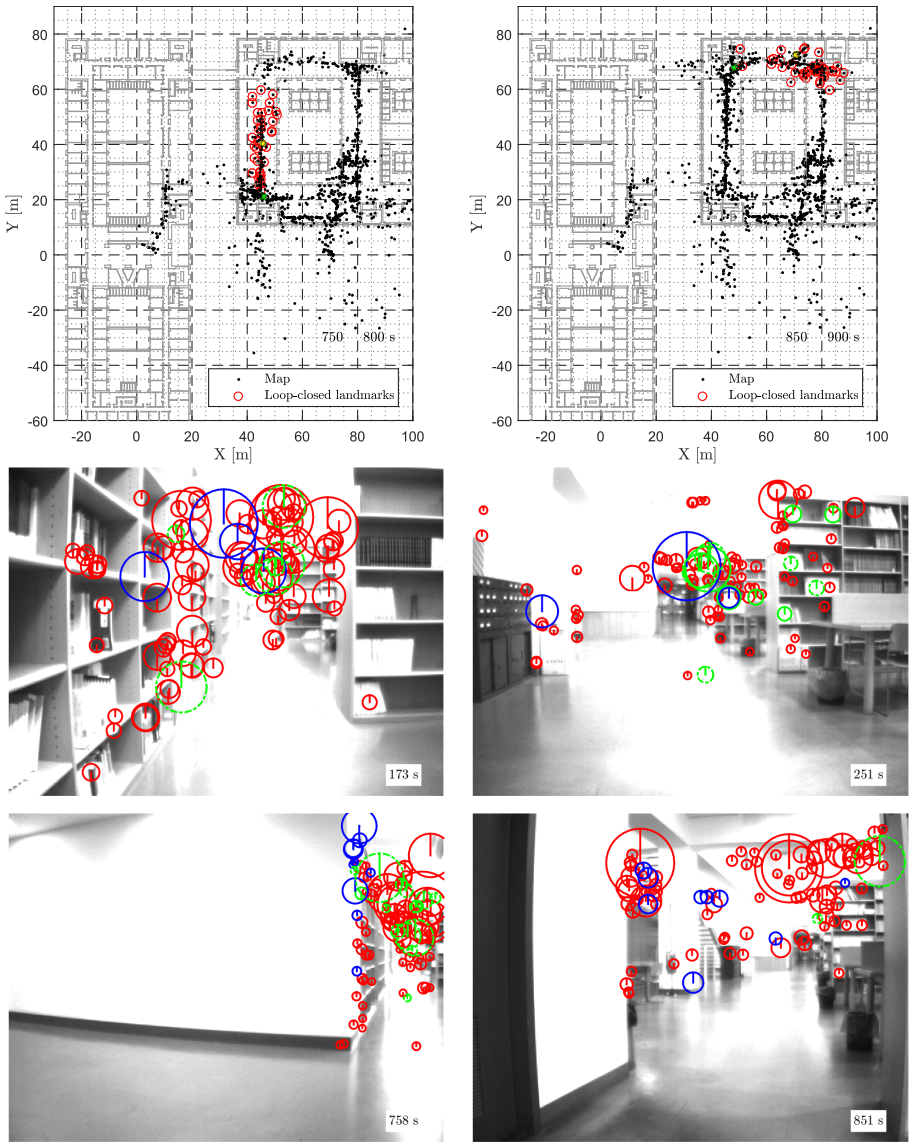


Figure 9: At the top: loop-closed landmarks in the first loop closing peaks of Figure 8(a), with loop-closed landmarks highlighted in red. The middle and bottom pictures are frames of the camera stream with detected (red), selected (green) and associated (blue) features. The bottom one contains the most loop-closed landmarks of the corresponding time slot and the top one shows the one where more loop-closed landmarks where first observed.

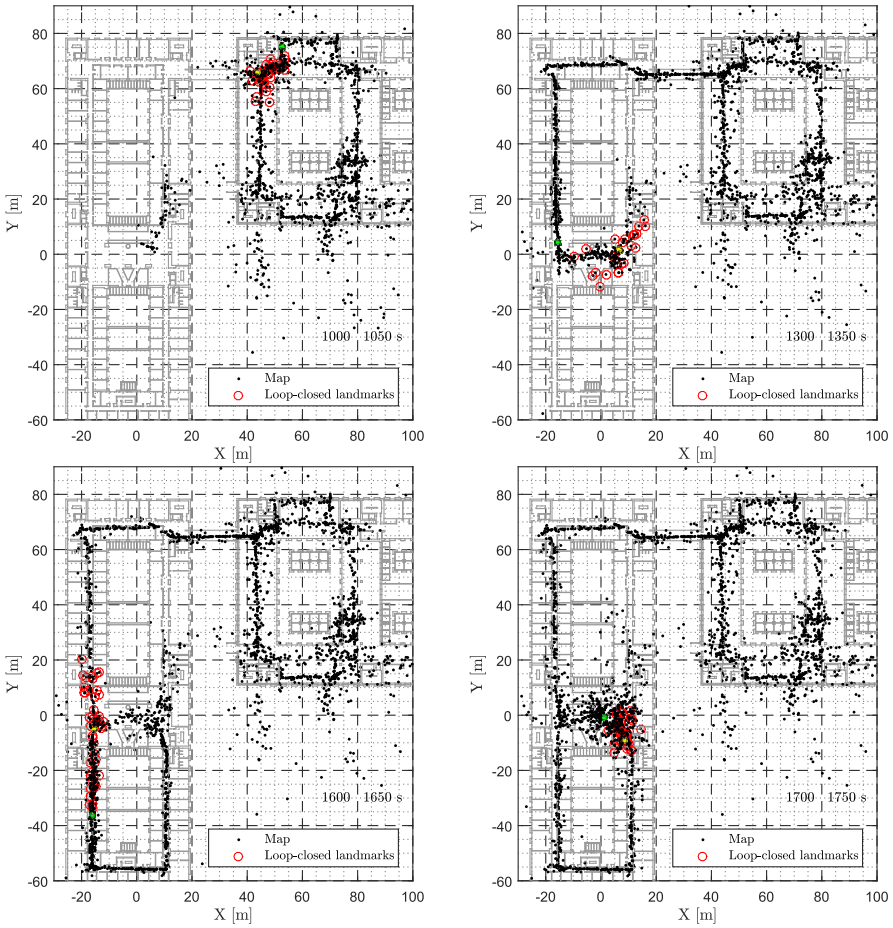


Figure 10: Maps corresponding to loop closing peaks in Figure 8(a), with loop-closed landmarks highlighted in red.

circles show landmarks that closed a loop sometime in the run, not exactly in this frame. The vertical line in each feature indicates the sign of the laplacian. It can be seen that, even though the loop closing is not based directly on visual place recognition, the algorithm is able to discern whether a block of features in different frames come in fact from the same place, thus demonstrating the good performance of the algorithm, as well as its consistency.

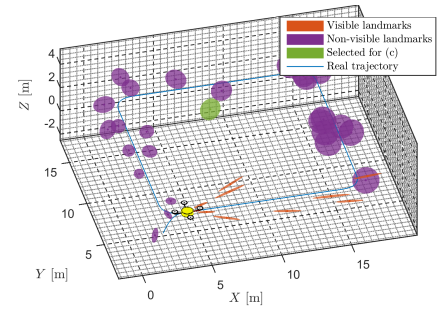
Another measure of the consistency of a SLAM algorithm is the normalized innovation squared (NIS) value for each set of associations, given by the joint Mahalanobis distance of the observation. This value is employed as an association index, and is known to follow a chi-squared distribution with 3 degrees of freedom for 3-D measurements. It is compared to the 95% percentile threshold to determine if an association is valid or not. In Figure 8(b), the average and

maximum NIS values are also normalized with the value corresponding to the 95% percentile. It is observed that although the maximum values approach the threshold, the average association is well below it, indicating that the landmark associations and the overall sensor-based algorithm are consistent. A clear exception to this behaviour is observed in the situations where more loop closings occur, which makes sense as the uncertainty of older landmarks is larger, and the influence of the drift of the odometry sensors is also more felt than when associating with newer landmarks.

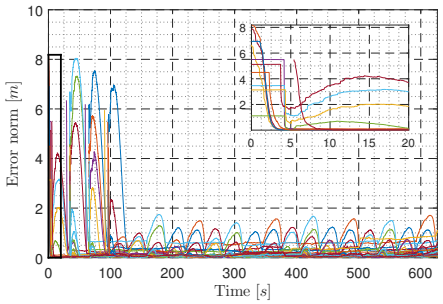
5.3. Simulated results and analysis

In order to allow a better understanding of some of the properties of the algorithm, a simulated environment was devised and several simulations were performed. This environment emulates the fifth floor of the North Tower at IST [32], and consists of a 16 by 16 by 3 m closed corridor. 36 landmarks were put in notable places such as corners and doors. The vehicle starts stopped at the ground, and after taking off makes several laps around the corridor. It completes a loop of 55 m in 124 seconds, and the total trajectory is 5 loops at 0.440 m/s of average speed. The field of view of the vehicle was limited to 90° horizontally and vertically with a range of up to 20 m. Furthermore, the effect of walls was taken into account. This means that the landmarks are only visible during a limited period of time in each loop. The bearing measurements are obtained by rotating the true bearing about random vectors of a random zero-mean angle with Gaussian distribution with standard deviation of 1°. The remaining measurements are corrupted with additive zero mean white noise. The standard deviation of the noise corrupting the linear velocity is 0.01 m/s, and that of the angular velocity is 0.15°/s. All measurements are obtained at 20 Hz.

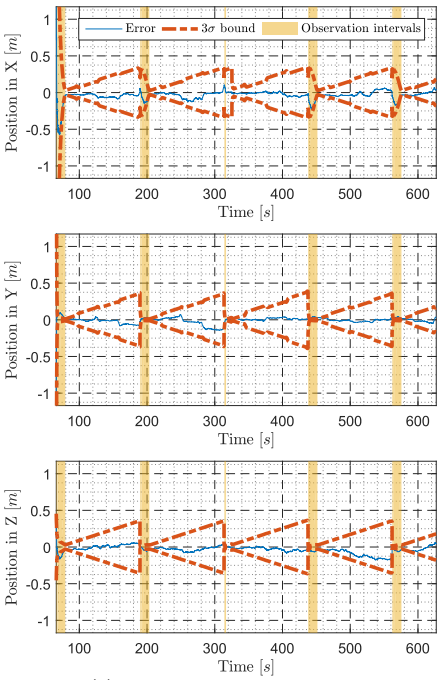
Figure 11 condenses the results of a typical run. At the top left, Figure 11(a) shows the mapped landmarks with their 3σ uncertainty ellipsoids and the trajectory of the vehicle in blue. The landmarks that were currently being observed in that instant are depicted in red and the rest are depicted in purple. One selected landmark is shown in green, and the time evolution of its estimates is shown in detail in Figure 11(c). There the estimation error is shown in blue for each coordinate and the Kalman filter uncertainty bounds are depicted in red. To enable a better understanding of the behaviour of the convergence observed therein, the intervals where the landmark was observed are shown in vertical yellow strips. As expected, the estimates converge quite fast when the landmark is first observed and then the uncertainty grows with the motion of the vehicle away from the landmark, converging again at a new sighting. Finally, Figure 11(b) shows the convergence of the norm of the estimation error for all landmarks. At the right top of that figure a detail of the first 20 seconds is depicted. It can be seen that, as predicted by the theoretical results, even when initial estimate is far off (up to 8 meters of error), the error converges quite fast, depending on how much time each landmark was observed. The simulation results detailed above confirm the theoretical results of the paper and provide quantitative data for its evaluation, but do not allow a thorough understanding



(a) 3-D map.



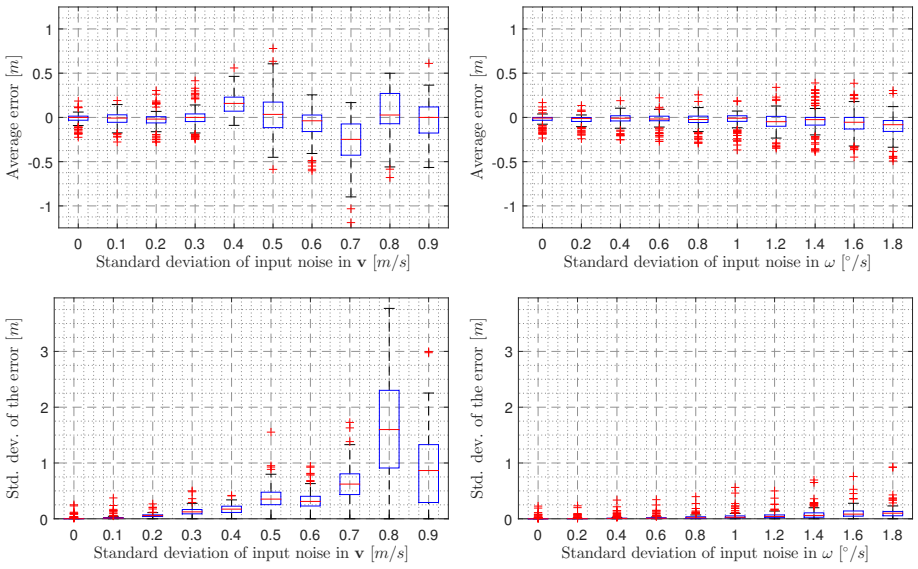
(b) Norm of the error and detail of initial convergence.



(c) Detail of one landmark.

Figure 11: Results of a typical simulation.

of the performance of the algorithm in the presence of different input noise levels. For that reason, and to illustrate how the algorithm deals with noise levels such as the ones in Figure 4, a series of trials were run with varying levels of noise. The results are depicted in Figure 12, where the distribution of the estimation error for 10 different noise levels is presented. The top figures show the distribution of the average landmark estimation error for all coordinates after initial convergence, and the bottom figures show the distribution of the standard deviation. In each column, one type of input noise was varied, while the others were kept constant. For both, the bearing measurement noise is the same as in Figure 11. For Figure 12(a) the standard deviation of the noise in the linear velocity measurements was increased from 0 to 0.9 m/s while the standard deviation of angular rate noise was constant at 0.15°/s, and for Figure 12(b) the standard deviation of the noise in the angular velocity measurements was increased from 0 to 1.8 °/s while the standard deviation of linear velocity noise was constant at 0.01 m/s. These figures show the impact of the noise on the estimation error, and how differently noise in the linear and angular velocities affects the estimates. In fact, the filter deals quite well with large levels of noise in the angular rates, and is much more sensitive to very high levels of noise in the linear velocity measurements. However, the average error for each landmark



(a) Varying noise levels in linear velocity (b) Varying noise levels in angular velocity

Figure 12: Sensitivity of the estimation error with respect to noise in the linear velocity and angular velocity measurements. The red horizontal lines represent the median, the blue boxes delimits the 25th and 75th percentiles, the black whiskers are extended at 2.7σ, and the red crosses are all the individual occurrences beyond that bound. On the left, angular velocity noise standard deviation is fixed at $0.15°/s$, and on the right the linear velocity noise standard deviation is fixed at 0.01 m/s.

coordinate is kept within 1 meter for both cases, and the standard deviation is mostly contained below 2 meters. Note that the parameters of the filter were not adjusted to each noise level, which indicates that these results could, in theory, be better.

The observability results of Section 3 provide insight that should be taken into account when designing experiments. Even though that was not the case here, the observability conditions seem to be fulfilled for the most part of the landmarks, as they converge to recognizable positions. For that reason, these experimental results accompanied by an analysis in simulation served the purpose of demonstrating the capabilities of the sensor-based BO-SLAM filter herein proposed both in terms of performance and consistency.

6. Conclusions and future work

A novel sensor-based globally exponentially stable filter for bearing-only simultaneous localization and mapping was proposed in this paper. The design of a linear time-varying system that mimics the dynamics of the underlying nonlinear system was facilitated by the use of a state augmentation and a simple output transformation on a nonlinear system, while disposing of non-visible

landmarks. A thorough and constructive observability analysis was performed, leading to the establishment of physically-grounded necessary and sufficient conditions for observability, stability and convergence of the Kalman filter that followed, as well as conditions for the observability of the underlying nonlinear system. These conditions are interesting for trajectory design or motion planning. The sensor-based BO-SLAM filter was prepared for practical implementation with a camera as the only exteroceptor, and thus the theoretical work which is the main focus of this paper was validated through experimental results based on a widely available dataset, that allowed to expose the good performance of the algorithm, as well as its consistency in an indoor environment with a large loop. These experimental results also helped to confirm the convergence of the error dynamics of the Kalman filter for LTV systems, as predicted by the theoretical results herein presented.

With respect to future work, the authors identify one main course of action, consisting of the optimization of the implementation for real-time, which is of paramount importance for achieving a truly online filter that can be used with autonomous ground or aerial vehicles.

AppendixA. Proof of Theorem 4

As was done in the proof of Theorem 1, the transformed system (7) will be used to simplify the analysis, and the obtained results can be applied to the system (6).

A pair $(\boldsymbol{\mathcal{A}}(t,\mathbf{y}(t)),\boldsymbol{\mathcal{C}}(t))$ is uniformly completely observable if and only if there exist positive constants δ and α such that for all $t \geq t_0$ and for all unit vectors $\mathbf{c}$ the quadratic form $\mathbf{c}^T\boldsymbol{\mathcal{W}}(t,t+\delta)\mathbf{c}$ is greater than or equal to α, i.e., if and only if

$$\underset{\substack{\delta>0\\ \alpha>0}}{\exists} \quad \underset{t\geq t_0}{\forall} \quad \underset{\substack{\mathbf{c}\in\mathbb{R}^{n_z}\\ \|\mathbf{c}\|=1}}{\forall} : \quad \mathbf{c}^T\boldsymbol{\mathcal{W}}(t,t+\delta)\mathbf{c} \geq \alpha, \tag{A.1}$$

meaning that, in contrast with the observability definition used in Theorem 1, the Gramian must have uniform bounds at all times. The proof follows by exhaustion, by analysing the quadratic form in the previous expression for all the possible cases of unit vectors $\mathbf{c}$ for all time. [52, Proposition 4.2] is helpful in that analysis. It states that, given a vector function $\mathbf{g}(\tau,t_0)$ if it is possible to find a positive constant β such that $\|\frac{\partial^i}{\partial\tau^i}\mathbf{g}(\tau,t_0)\| > \beta$ then there exists a $\gamma>0$ such that $\|\mathbf{g}(t_0,t_0+\delta)\| > \gamma$ as long as $\frac{\partial^j}{\partial\tau^j}\mathbf{g}(\tau,t_0)|_{\tau=t_0} = 0$ for all $j<i$ and the norm of the $i+1$-th derivative is upper bounded. This proposition applies to the quadratic form in study, and therefore it suffices to show that the norm of $\mathbf{f}(\tau,t)$, given by

$$\|\mathbf{f}(\tau,t)\|^2 = \left\|\mathbf{c}_p - {}^E\mathbf{b}_1(\tau)c_r\right\|^2 \tag{A.2}$$

is lower bounded uniformly in time by some $\alpha^*>0$ at some time $\tau\in\mathcal{T}_\delta$ for all the possible cases of unit vectors $\mathbf{c}$. It is necessary to address $\|\mathbf{f}(\tau,t)\|$ differently depending on $\mathbf{c}_p$ and $\mathbf{c}_r$. For the case where $\|\mathbf{c}_p\| \geq \alpha_p$ and $|c_r| < \alpha_r$ for some

positive constants $\alpha_p < 1$ and $\alpha_r < 1$, (A.2) can be simplified at $\tau = t$ to yield

$$\begin{aligned}\|\mathbf{f}(t,t)\|^2 &\geq \|\mathbf{c}_p\| \left(\|\mathbf{c}_p\| - 2|c_r|\right) \\ &\geq \alpha_p(\alpha_p - 2\alpha_r).\end{aligned}$$

If α_p is chosen so that $\alpha_p \geq 4\alpha_r$, this becomes

$$\|\mathbf{f}(t,t)\|^2 \geq \frac{\alpha_p^2}{2}.$$

For the next case, where $|c_r| \geq \alpha_r$ and $0 < \|\mathbf{c}_p\| < 1$, the condition (15) of the theorem will be necessary. Consider the norm of $\mathbf{f}(\tau, t)$ once more evaluated at $\tau = t$,

$$\begin{aligned}\|\mathbf{f}(t,t)\|^2 \geq& \left\| \mathbf{c}_p - {}^E\mathbf{b}_1(t_1)c_r + \int_t^{t_1} {}^E\dot{\mathbf{b}}_1(\tau)\mathrm{d}\tau\ c_r \right\|^2 \\ \geq& \left\| \int_t^{t_1} {}^E\dot{\mathbf{b}}_1(\tau)\mathrm{d}\tau\ c_r \right\|^2 \\ & - 2\left\|\mathbf{c}_p - {}^E\mathbf{b}_1(t_1)c_r\right\| \left\| \int_t^{t_1} {}^E\dot{\mathbf{b}}_1(\tau)\mathrm{d}\tau\ c_r \right\| \\ \geq& \alpha_b\alpha_r \left(\alpha_b\alpha_r - 2\left\|\mathbf{c}_p - {}^E\mathbf{b}_1(t_1)c_r\right\|\right).\end{aligned}$$

If $\|\mathbf{c}_p - {}^E\mathbf{b}_1(t_1)c_r\| < \alpha_1$ with $\alpha_1 < \frac{1}{4}\alpha_b\alpha_r$ this becomes

$$\|\mathbf{f}(t,t)\|^2 > \frac{\alpha_b^2\alpha_r^2}{2}.$$

If, on the other hand, $\|\mathbf{c}_p - {}^E\mathbf{b}_1(t_1)c_r\|$ is lower bounded by α_1, then

$$\|\mathbf{f}(t_1,t)\|^2 \geq \alpha_1^2.$$

This concludes the enumeration of all the possible cases. Positive lower bounds were found for $\|\mathbf{f}(\tau,t)\|$ on different instants of $\mathcal{T}_\delta$ depending on the case, which, using [52, Proposition 4.2] implies that the integral in (9) is in fact lower bounded for unit vectors $\mathbf{c}$, and hence (A.1) is true. Thus, the pair $(\boldsymbol{\mathcal{A}}(t), \boldsymbol{\mathcal{C}}(t))$ is uniformly completely observable, and, as its associated system is related to system (6) through a Lyapunov transformation, the pair $(\mathbf{A}(t), \mathbf{C}(t))$ is also uniformly completely observable, thus concluding the proof of the sufficiency part of the theorem.

The proof of the necessity part of the theorem follows by contraposition. The objective is to negate the conditions of the theorem, and show that under that hypothesis the pair $(\mathbf{A}(t), \mathbf{C}(t))$ is not uniformly completely observable. For that purpose consider the negation of the condition of the theorem, given

by

$$\underset{\substack{\delta>0\\ \beta>0}}{\forall}\quad \underset{t\geq t_0}{\exists}\quad \underset{t_1\in\mathcal{T}_\delta}{\forall}: \quad \left\| \int_t^{t_1} {}^E\dot{\mathbf{b}}_1(\tau)\mathrm{d}\tau \right\| < \beta, \tag{A.3}$$

and let us analyse $\mathbf{c}^T\boldsymbol{\mathcal{W}}(t,t+\delta)\mathbf{c}$ for a particular $\mathbf{c}$.

$$\begin{aligned}\mathbf{c}^T\boldsymbol{\mathcal{W}}(t,t+\delta)\mathbf{c} &= \int_t^{t+\delta} \left\|\mathbf{c}_p - {}^E\mathbf{b}_1(\tau)c_r\right\|^2 \mathrm{d}\tau \\ &= \int_t^{t+\delta} \left\| \mathbf{c}_p - \left({}^E\mathbf{b}_1(t) + \int_t^{\tau} {}^E\dot{\mathbf{b}}_1(\sigma)\mathrm{d}\sigma \right) c_r \right\|^2 \mathrm{d}\tau\end{aligned}$$

Choosing $\mathbf{c}_p = {}^E\mathbf{b}_1(t)c_r$ and $c_r = \frac{\sqrt{2}}{2}$, it is possible to write

$$\mathbf{c}^T\boldsymbol{\mathcal{W}}(t,t+\delta)\mathbf{c} = \frac{1}{2}\int_t^{t+\delta} \left\| \int_t^{\tau} {}^E\dot{\mathbf{b}}_1(\sigma)\mathrm{d}\sigma \right\|^2 \mathrm{d}\tau. \tag{A.4}$$

Recall now the negation of the conditions of the theorem, (A.3). It is clear that the norm therein is upper-bounded for any t_1 between t and $t+\delta$ for a given $t \geq t_0$. In particular, this is also true for $t_1 = \tau$, which means that there exists a $t \geq t_0$ such that the norm in the integral in (A.4) is also upper-bounded, i.e.,

$$\underset{\substack{\delta>0\\ \epsilon>0}}{\forall}\quad \underset{t\geq t_0}{\exists}\quad \underset{\substack{\mathbf{c}\in\mathbb{R}^{n_z}\\ \|\mathbf{c}\|=1}}{\exists}: \qquad \mathbf{c}^T\boldsymbol{\mathcal{W}}(t,t+\delta)\mathbf{c} < \frac{\beta^2\delta}{2} := \epsilon, \tag{A.5}$$

where $\beta = \sqrt{\frac{2\epsilon}{\delta}}$. Therefore, when the conditions of the theorem do not hold, it is possible to find a unit $\mathbf{c}$, e.g. $\mathbf{c} = \begin{bmatrix} \frac{\sqrt{2}}{2}{}^E\mathbf{b}_1^T(t) & \frac{\sqrt{2}}{2} \end{bmatrix}^T$, such that (A.5) is true, i.e., the pair $(\mathbf{A}(t), \mathbf{C}(t))$ is not uniformly completely observable. By contraposition, the conditions of the theorem are necessary for the uniform complete observability of the system (6). □

AppendixB. Proof of Lemma 1

Consider the update equation [47] for the covariance of the Kalman filter, given by

$$\boldsymbol{\Sigma}_{k|k} = \left(\mathbf{I} - \boldsymbol{\Sigma}_{k|k-1}\mathbf{C}_{F_k}^T \boldsymbol{\Sigma}{\nu_k}^{-1} \right) \boldsymbol{\Sigma}_{k|k-1}$$

where $\mathbf{C}_{F_k}$ is the discrete version of (5), and the innovation covariance $\boldsymbol{\Sigma}\nu_k$ is given by

$$\boldsymbol{\Sigma}\nu_k = \mathbf{C}_{F_k}\boldsymbol{\Sigma}_{k|k-1}\mathbf{C}_{F_k}^T + \boldsymbol{\Theta}_k.$$

Assume that $\boldsymbol{\Sigma}_{k|k-1}$ has the structure in (19). In this case, if the measurement covariance $\boldsymbol{\Theta}_k$ is block diagonal, i.e., there are no cross-covariances between landmark observations, it is a matter of tedious computation to arrive at

$$\boldsymbol{\Sigma}\nu_{kij} = \begin{cases} \boldsymbol{\Sigma}_{L_i}^{k|k-1} - \mathbf{b}_{i_k}\boldsymbol{\Sigma}_{R_iL_i}^{k|k-1} - \left(\boldsymbol{\Sigma}_{L_iR_i}^{k|k-1} - \mathbf{b}_{i_k}\boldsymbol{\Sigma}_{R_i}^{k|k-1}\right)\mathbf{b}_{i_k}^T + \boldsymbol{\Theta}_{i_k}, & i = j \\ \mathbf{0}, & i \neq j \end{cases}$$

which means that the innovation covariance is also block-diagonal and its inverse can be written as $\boldsymbol{\Sigma}_{\nu_k}^{-1} = \text{diag}(\boldsymbol{\Sigma}_{\nu_{k1}}^{-1}, \ldots, \boldsymbol{\Sigma}_{\nu_{kN_O}}^{-1})$. Using this structure it is a matter of computation to arrive at a covariance matrix with the structure in (19), where the individual elements are

$$\begin{cases} \boldsymbol{\Sigma}_{L_i}^{k|k} = \boldsymbol{\Sigma}_{L_i}^{k|k-1} - \left(\boldsymbol{\Sigma}_{L_i}^{k|k-1} - \boldsymbol{\Sigma}_{L_iR_i}^{k|k-1}\mathbf{b}_i^T\right)\boldsymbol{\Sigma}_{\nu_i}^{-1}\left(\boldsymbol{\Sigma}_{L_i}^{k|k-1} - \boldsymbol{\Sigma}_{L_iR_i}^{k|k-1}\mathbf{b}_i^T\right)^T \\ \boldsymbol{\Sigma}_{R_i}^{k|k} = \boldsymbol{\Sigma}_{R_i}^{k|k-1} - \left(\boldsymbol{\Sigma}_{R_iL_i}^{k|k-1} - \boldsymbol{\Sigma}_{R_i}^{k|k-1}\mathbf{b}_i^T\right)\boldsymbol{\Sigma}_{\nu_i}^{-1}\left(\boldsymbol{\Sigma}_{R_iL_i}^{k|k-1} - \boldsymbol{\Sigma}_{R_i}^{k|k-1}\mathbf{b}_i^T\right)^T \\ \boldsymbol{\Sigma}_{L_iR_i}^{k|k} = \boldsymbol{\Sigma}_{L_iR_i}^{k|k-1} - \left(\boldsymbol{\Sigma}_{L_i}^{k|k-1} - \boldsymbol{\Sigma}_{L_iR_i}^{k|k-1}\mathbf{b}_i^T\right)\boldsymbol{\Sigma}_{\nu_i}^{-1}\left(\boldsymbol{\Sigma}_{R_iL_i}^{k|k-1} - \boldsymbol{\Sigma}_{R_i}^{k|k-1}\mathbf{b}_i^T\right)^T \end{cases}.$$

This shows that the structure is maintained by the update process of the Kalman filter when applied to the system at hand. Recall the discrete dynamics matrix $\mathbf{F}_{F_k}$ defined in (17). The predicted covariance for the instant $k+1$ is then given by

$$\boldsymbol{\Sigma}_{k+1|k} = \boldsymbol{\Xi}_k + \begin{bmatrix} \mathbf{R}_{k+1}^T\mathbf{R}_k\boldsymbol{\Sigma}_{L_1}^{k|k}\mathbf{R}_k^T\mathbf{R}_{k+1} & \mathbf{0} & \mathbf{R}_{k+1}^T\mathbf{R}_k\boldsymbol{\Sigma}_{L_1R_1}^{k|k} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{k+1}^T\mathbf{R}_k\boldsymbol{\Sigma}_{L_2}^{k|k}\mathbf{R}_k^T\mathbf{R}_{k+1} & \mathbf{0} & \mathbf{R}_{k+1}^T\mathbf{R}_k\boldsymbol{\Sigma}_{L_2R_2}^{k|k} \\ \boldsymbol{\Sigma}_{R_1L_1}^{k|k}\mathbf{R}_k^T\mathbf{R}_{k+1} & \mathbf{0} & \boldsymbol{\Sigma}_{R_1}^{k|k} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{R_2L_2}^{k|k}\mathbf{R}_k^T\mathbf{R}_{k+1} & \mathbf{0} & \boldsymbol{\Sigma}_{R_2}^{k|k} \end{bmatrix},$$

where the number of landmarks was limited to 2 to avoid cluttering the reading. Thus, using the final condition of the lemma, that requires the process covariance $\boldsymbol{\Xi}_k$ to have the same structure as (19), one complete step of the Kalman filter for system (16) maintains throughout the update and predict stages the structure of the covariance matrix if it was (19). Therefore, if the initial covariance of the filter is set to have the structure in (19) and the conditions of the lemma hold, then the filter covariance will maintain that structure for all time. □

Acknowledgments

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT) through ISR under LARSyS UID/EEA/50009/2013, and through IDMEC, under LAETA UID/EMS/50022/2013 contracts, by the University of Macau Project MYRG2015-00126-FST, and by the Macao Science and Tech-

nology Development Fund under Grant FDCT/048/2014/A1. The work of P. Lourenço was supported by the PhD. Student Grant SFRH/BD/89337/2012 from FCT.

References

[1] R. Smith, P. Cheeseman, On the representation and estimation of spatial uncertainty, International Journal of Robotics Research 5 (4) (1986) 56–68. doi:10.1177/027836498600500404.

[2] J. Leonard, H. Durrant-Whyte, Simultaneous map building and localization for an autonomous mobile robot, in: Proc. of the IEEE/RSJ International Workshop on Intelligent Robots and Systems (IROS), Vol. 3, 1991, pp. 1442–1447.

[3] H. Durrant-Whyte, T. Bailey, Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms, IEEE Robotics & Automation Magazine 13 (2) (2006) 99–110.

[4] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping (SLAM): Part II, IEEE Robotics & Automation Magazine 13 (3) (2006) 108–117.

[5] S. Huang, G. Dissanayake, A critique of current developments in simultaneous localization and mapping, International Journal of Advanced Robotic Systems 13 (5) (2016) 1–13. doi:10.1177/1729881416669482.

[6] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J. J. Leonard, Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age, IEEE Transactions on Robotics 32 (6) (2016) 1309–1332. doi:10.1109/TRO.2016.2624754.

[7] T. Lemaire, S. Lacroix, J. Solà, A practical 3D bearing-only SLAM algorithm, in: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005, pp. 2449–2454.

[8] T. Bailey, Constrained initialisation for bearing-only SLAM, in: Proceedings of the 2003 IEEE International Conference on Robotics and Automation, Vol. 2, IEEE, 2003, pp. 1966–1971.

[9] N. M. Kwok, G. Dissanayake, An efficient multiple hypothesis filter for bearing-only SLAM, in: Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vol. 1, 2004, pp. 736–741.

[10] J. Solà, A. Monin, M. Devy, T. Lemaire, Undelayed initialization in bearing only SLAM, in: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005, pp. 2499–2504.

[11] J. Civera, A. J. Davison, J. M. M. Montiel, Inverse depth parametrization for monocular slam, IEEE Transactions on Robotics 24 (5) (2008) 932–945. doi:10.1109/TRO.2008.2003276.

[12] K. Choi, J. Park, Y.-H. Kim, H.-K. Lee, Monocular SLAM with undelayed initialization for an indoor robot, Robotics and Autonomous Systems 60 (6) (2012) 841–851. doi:10.1016/j.robot.2012.02.002.

[13] K. Bekris, M. Click, E. Kavraki, Evaluation of algorithms for bearing-only SLAM, in: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, 2006, pp. 1937–1943.

[14] P. Jensfelt, D. Kragic, J. Folkesson, M. Bjorkman, A framework for vision based bearing only 3D SLAM, in: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, 2006, pp. 1944–1950.

[15] H. Huang, F. D. Maire, N. Keeratipranon, Bearing-only simultaneous localization and mapping for vision-based mobile robots, in: G. Obinata, A. Dutta (Eds.), Vision Systems-Applications, I-Tech Education and Publishing, Vienna, Austria, 2007, pp. 335–360.

[16] C. Becker, D. Ribas, P. Ridao, Simultaneous Sonar Beacon Localization & AUV Navigation, in: Proceedings of the 9th IFAC Conference on Manoeuvring and Control of Marine Craft (MCMC), Arenzano, Italy, 2012, pp. 200–205.

[17] A. Davison, I. Reid, N. Molton, O. Stasse, MonoSLAM: Real-Time Single Camera SLAM, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 1052–1067.

[18] A. Chiuso, P. Favaro, H. Jin, S. Soatto, 3-D motion and structure from 2-D motion causally integrated over time: Implementation, in: Proceedings of the European Conference on Computer Vision, Springer Berlin Heidelberg, 2000, pp. 734–750.

[19] A. Kim, R. M. Eustice, Real-Time Visual SLAM for Autonomous Underwater Hull Inspection Using Visual Saliency, IEEE Transactions on Robotics 29 (3) (2013) 719–733. doi:10.1109/TRO.2012.2235699.

[20] R. Mur-Artal, J. M. M. Montiel, J. D. Tardós, ORB-SLAM: A Versatile and Accurate Monocular SLAM System, IEEE Transactions on Robotics 31 (5) (2015) 1147–1163. doi:10.1109/TRO.2015.2463671.

[21] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in: 2011 International Conference on Computer Vision, 2011, pp. 2564–2571. doi:10.1109/ICCV.2011.6126544.

[22] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-Up Robust Features (SURF), Computer Vision and Image Understanding 110 (3) (2008) 346–359.

[23] G. Dubbelman, B. Browning, COP-SLAM: Closed-Form Online Pose-Chain Optimization for Visual SLAM, IEEE Transactions on Robotics 31 (5) (2015) 1194–1213. doi:10.1109/TRO.2015.2473455.

[24] S. Ullman, The interpretation of structure from motion, Proceedings of the Royal Society of London B: Biological Sciences 203 (1153) (1979) 405–426. arXiv:http://rspb.royalsocietypublishing.org/content/203/1153/405.full.pdf, doi:10.1098/rspb.1979.0006.
URL http://rspb.royalsocietypublishing.org/content/203/1153/405

[25] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon, Vision Algorithms: Theory and Practice, Vol. 1883 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2000, Ch. Bundle Adjustment — A Modern Synthesis, pp. 298–372. doi:10.1007/3-540-44480-7_21.

[26] H. Strasdat, J. Montiel, A. J. Davison, Visual slam: Why filter?, Image and Vision Computing 30 (2) (2012) 65–77. doi:10.1016/j.imavis.2012.02.009.

[27] N. Trawny, S. I. Roumeliotis, A unified framework for nearby and distant landmarks in bearing-only SLAM, in: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, IEEE, 2006, pp. 1923–1929.

[28] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, Simultaneous Localization and Mapping for Aerial Vehicles: a 3-D sensor-based GAS filter, Autonomous Robots 40 (2016) 881–902. doi:10.1007/s10514-015-9499-z.

[29] J. Neira, J. Tardós, Data Association in Stochastic Mapping Using the Joint Compatibility Test, IEEE Transactions on Robotics and Automation 17 (6) (2001) 890–897.

[30] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, J. Tardós, A comparison of loop closing techniques in monocular SLAM, Robotics and Autonomous Systems 57 (12) (2009) 1188–1197. doi:10.1016/j.robot.2009.06.010.

[31] P. Batista, C. Silvestre, P. Oliveira, Globally exponentially stable filters for source localization and navigation aided by direction measurements, Systems & Control Letters 62 (11) (2013) 1065–1072.

[32] P. Lourenço, P. Batista, P. Oliveira, C. Silvestre, A Globally Exponentially Stable filter for Bearing-Only Simultaneous Localization and Mapping in 3-D, in: Proc. of the 2015 European Control Conference, Linz, Austria, 2015, pp. 2817–2822.

[33] A. Bonarini, W. Burgard, G. Fontana, M. Matteucci, D. G. Sorrenti, J. D. Tardos, Rawseeds: Robotics advancement through web-publishing of sensorial and elaborated extensive data sets, in: Proceedings of IROS'06 Workshop on Benchmarks in Robotics Research, 2006.

[34] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. G. Sorrenti, P. Taddei, Rawseeds ground truth collection systems for indoor self-localization and mapping, Autonomous Robots 27 (4) (2009) 353–371. doi:10.1007/s10514-009-9156-5.

[35] J. Castellanos, R. Martinez-Cantin, J. Tardós, J. Neira, Robocentric map joining: Improving the consistency of EKF-SLAM, Robotics and Autonomous Systems 55 (1) (2007) 21–29.

[36] T. Vidal-Calleja, M. Bryson, S. Sukkarieh, A. Sanfeliu, J. Andrade-Cetto, On the Observability of Bearing-only SLAM, in: Proceedings of the 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 4114–4119.

[37] B. J. Guerreiro, P. Batista, C. Silvestre, P. Oliveira, Globally Asymptotically Stable Sensor-based Simultaneous Localization and Mapping, IEEE Transactions on Robotics 29 (6) (2013) 1380–1395.

[38] P. Lourenço, P. Batista, P. Oliveira, C. Silvestre, C. L. P. Chen, Sensor-based Globally Exponentially Stable Range-Only Simultaneous Localization and Mapping, Robotics and Autonomous Systems 68 (2015) 72–85. doi:10.1016/j.robot.2015.01.010.

[39] T. A. Johansen, E. Brekke, Globally Exponentially Stable Kalman Filtering for SLAM with AHRS, in: Proceedings of the 19th International Conference on Information Fusion (FUSION 2016), Heidelberg. Germany, 2016, pp. 909–916.

[40] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, 3-D Inertial Trajectory and Map Online Estimation: Building on a GAS Sensor-based SLAM filter, in: Proc. of the 2013 European Control Conference, Zurich, Switzerland, 2013, pp. 4214–4219.

[41] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, Uncertainty Characterization of the Orthogonal Procrustes Problem with Arbitrary Covariance Matrices, Pattern Recognition 61 (2017) 210–220. doi:10.1016/j.patcog.2016.07.037.

[42] P. Batista, C. Silvestre, P. Oliveira, A GES Attitude Observer with Single Vector Observations, Automatica 48 (2) (2012) 388–395.

[43] P. Batista, C. Silvestre, P. Oliveira, Single range aided navigation and source localization: Observability and filter design, Systems & Control Letters 60 (8) (2011) 665–673.

[44] R. Brockett, Finite Dimensional Linear Systems, Series in decision and control, John Wiley & Sons, 1970.

[45] H. Khalil, Nonlinear Systems, 3rd Edition, Prentice Hall, 2002.

[46] B. D. O. Anderson, Stability properties of Kalman-Bucy filters, Journal of the Franklin Institute 291 (2) (1971) 137–144.

[47] A. Gelb, Applied Optimal Estimation, MIT Press, 1974.

[48] J. Heikkila, O. Silven, A four-step camera calibration procedure with implicit image correction, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997, 1997, pp. 1106–1112. doi:10.1109/CVPR.1997.609468.

[49] J.-Y. Bouguet, Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/parameters.html, online: version of October 14th, 2015 (Oct. 2015).

[50] J. Civera, O. García, J. M. M. Montiel, Benchmark Solution "Inverse Depth EKF based Visual Odometry using 1Point RANSAC", http://www.rawseeds.org/rs/solutions/view/48 (Nov. 2009).

[51] J. Civera, O. G. Grasa, A. J. Davison, J. M. M. Montiel, 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry, Journal of Field Robotics 27 (5) (2010) 609–631. doi:10.1002/rob.20345.

[52] P. Batista, C. Silvestre, P. Oliveira, On the observability of linear motion quantities in navigation systems, Systems & Control Letters 60 (2) (2011) 101–110.