SC651: Estimation on Lie Group
Assignment 2: Summary of van Goor's paper
VSLAM using Differential Geometry

Tanmay Dokania (200070083)

April 12, 2023

Contents

The paper deals with Constructive Observer Design for Visual Simultaneous Localisation and Mapping.

1 Model and Goal

The aim of the paper is to develop a VSLAM algorithm using the framework of Lie Groups.

1.1 What is VSLAM?

Simultaneous Localisation and Mapping (SLAM) refers to the problem of finding the location/pose of an agent (localize) and developing an understanding of the environment (map). This problem has a wide range of applications in robotics, as it enables us to send robots into unknown environments. Visual SLAM (VSLAM) refers to the use of visual data to develop maps and localize the agent. It covers a wide range of problems, ranging from when the camera sensor is monocular or stereo. The camera itself can be RGB-D (one with depth sensing).
In this paper, the focus is on developing an observer for the case of a monocular camera. The problem is highly nonlinear, and the current state-of-the-art solutions suffer from the problem of high computational complexity and lack of scalability.

1.2 What are we trying to do?

A novel non-linear equivariant observer is developed for the VSLAM Problem. The symmetry Lie-group $\mathbf{VSLAM}_{\mathrm{n}}(3)$ is fully developed with the equivariant actions on the state and output space. This allows us to represent measurements in a suitable framework and improve our estimates of states, in our case, the position & pose of the robot and the landmarks.
The observer works by having a state space and a measurement space on which a symmetry group acts to improve our estimate of the current state. These ideas will be formalized in the following sections.

2 Problem Formulation

2.1 VSLAM State Space

An absolute frame is defined, with respect to which the robot position and pose are denoted by $P \in \mathbf{SE}(3)$. There are n landmark positions, whose coordinates are denoted as $p_i \in \mathbb{R}^3, i \in \{1, 2, \dots, n\}$. The total state space is denoted as $\mathcal{T}_n^\circ(3) = \{(P, p_i) \in \mathbf{SE}(3) \times \mathbb{R}^3 \times \cdots \times \mathbb{R}^3\}$ for simplicity.

There is an inherent redundancy in the state space induced by the choices for ground reference, as we assume the landmarks to be stationary. Therefore, we use the action of $\mathbf{SE}(3)$ group to define equivalence between two objects in the state space, thereby defining an equivalence class. The group action, $S \in \mathbf{SE}(3)$ on $(P, p_i) \in \mathcal{T}_n^\circ(3)$ is defined as:

$$(P, p_i) \cdot S = (S^{-1}P', R_S^\top(p_i' - x_S))$$

The action of S corresponds to a shift in the system's reference coordinate.

The equivalence class can be defined as:

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

This equivalence class induces a quotient manifold structure on the total space ($\mathcal{T}_n^\circ(3)$), which can be defined as:

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

Hence, each element in $\mathcal{M}_n^\circ(3)$ corresponds to a set of configurations where the robot-centric coordinate of each of the landmarks is the same.

2.2 Kinematics of the total space

Elements in $\mathrm{T}\mathcal{T}_n^\circ(3)$, will correspond to a vector space with velocities ($\in \mathbb{R}^3$) of the each landmark, velocity (linear $\in \mathbb{R}^3$ and angular $\in \mathfrak{so}(3)$) of the robot ($\in \mathfrak{se}(3)$). However, as said earlier, we want to keep the landmarks fixed. Therefore we define a map $f : \mathcal{T}_n^\circ(3) \times \mathbb{V} \to \mathrm{T}\mathcal{T}_n^\circ(3)$, mapping velocities of the robot at a configuration to a change in configuration of the system.

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

2.3 Output Space

As the robot has a camera installed on it, it can measure the bearing of the landmark but not its distance. Some image processing needs to be done to identify the bearing of the landmark from the camera's images, but this is assumed to be done by other image processing algorithms.

Each landmark gives an element in S^2, therefore, the output space is defined as:

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

Hence, for n landmarks and a given element $(P, p_i) \in \mathcal{T}_n^\circ(3)$, a mapping can be defined on to n 2-spheres S^2. This is called the combined output function $h : \mathcal{T}_n(3)^\circ \to \mathcal{N}^n(3)$.

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

One quick check with intuition is that for any two "equivalent" elements in the total space, the output should be the same. Stating it formally, this combined output function must be well-defined as a map from the quotient manifold $\mathcal{M}_n^\circ(3)$ to the output space $\mathcal{N}^n(3)$. And it indeed is!

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

2.4 Novel Lie Group - $\mathbf{VSLAM_n(3)}$

A Lie group is defined as follows. It is easy to see that it will be a Lie group, as it is a finite cartesian product of different Lie groups.

$$\mathrm{VSLAM}_n(3) = \mathbf{SE}(3) \times (\mathbf{SO}(3) \times \mathbf{MR}(1))^n$$

Here, $\mathbf{MR}(1)$ is the set $(0, \infty)$, with the group operation of multiplication. It is clear to see that it is a group and a manifold, therefore, a Lie group. The identity element and group operation can be defined trivially, as done in the article.

2.4.1 Group action of $\mathbf{VSLAM_n(3)}$ on $\mathcal{T}_n^\circ(3)$

A mapping $\Upsilon : \mathrm{VSLAM}_n(3) \times \mathcal{T}_n^\circ(3) \to \mathcal{T}_n^\circ(3)$ defined by:

$$\Upsilon((A, (Q, a)_i), (P, p_i)) := (PA, a_i^{-1} R_{PA} Q_i^\top R_P^\top (p_i - x_P) + x_{PA})$$

This map is proven to be *right transitive.* It means that an element of the group $\mathrm{VSLAM}_n(3)$ can be found for any two elements of manifold $\mathcal{T}_n^\circ(3)$, such that the action of the group element on one of the manifold element gives the other manifold element. Alternatively, the group orbit over the manifold is the manifold itself.

The article also mentions that it is a symmetry of $\mathcal{T}_n^\circ(3)$. Here, symmetry is not in the traditional sense in which some parameter is invariant but rather in the sense of equivariance.

2.4.2 Group action of $\mathbf{VSLAM_n(3)}$ on $\mathcal{N}^n(3)$

The group action $\rho : \mathrm{VSLAM}_n(3) \times \mathcal{N}^n(3) \to \mathcal{N}^n(3)$ is defined by

$$\rho((A, (Q, a)_i), y_i) = Q_i^\top y_i$$

This definition implies that the action of the group just rotates each of the bearing measurements by $Q_i^\top$, different for each measurement. It can be clearly seen that this action is right in nature.

2.4.3 Equivariance

The output $h : \mathcal{T}_n^\circ(3) \to \mathcal{N}(3)$ is equivariant with respect to actions Υ and ρ. Implying, for any $X \in \mathrm{VSLAM}_n(3)$ and any $\xi \in \mathcal{T}_n^\circ(3)$,

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

Equivalently, with an "abuse of notation", we have:

$$h(X \cdot \xi) = X \cdot h(\xi))$$

As the group action is equivariant, it can be said that the order of group action does not depend and commutes with the output operation (h).

2.4.4 The intuition behind these definitions

This construction seems fairly unintuitive. However, after thinking about what these actions do and the figure in the paper, it becomes clear how this works.

The action of the group is elaborated in four steps (numbers corresponding to the operations in Figure 1):

1. P denotes the robot pose and position in SE(3)

2. p denotes the position of a landmark with respect to the absolute frame $\{0\}$; therefore, to get into the frame of the robot:

   - $(p - x_P)$: measuring the relative displacement
   - $R_P^\top(p - x_P)$: rotating the position vector to get into the orientation of the robot

3. $(Q, a) \in \mathbf{SO}(3) \times \mathbf{MR}(1)$: multiplication with a^{-1} denotes a re-scaling of the system and multiplication with $Q^\top$ denotes a rotation of the direction of the landmark's bearing in the frame fixed with the robot. Therefore, $a^{-1}Q^\top R_P^\top(p - x_P)$ becomes the new coordinate in the robot frame.

4. $A \in \mathbf{SE}(3)$: operation of A denotes a change in the frame of the robot, keeping the landmark fixed at the previously computed coordinate in the robot frame. Therefore, to get the new coordinate of the landmark, we need to operate it with the transformation associated with $P \cdot A$. Hence, we obtain the group action on the manifold.

The new landmark coordinate becomes:

$$p' = R_{PA} a^{-1} Q^\top R_P^\top (p - x_P) + x_{PA}$$

The new robot position-pose becomes:

$$P' = PA$$

The above construction implies that we can go from any element in $\mathcal{T}_n^\circ(3)$ to any element in $\mathcal{T}_n^\circ(3)$, using the action of an appropriately defined group

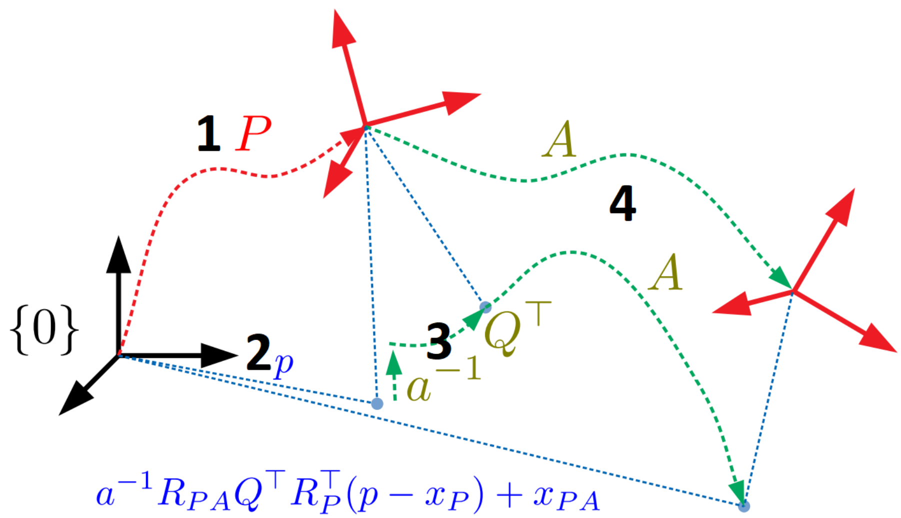

Figure 1: Action of $\mathbf{VSLAM_n}(3)$ on $\mathcal{T}_n^\circ(3)$

element. (As two elements can differ only in the relative bearing measurement with respect to robot and robot pose.) Given by:

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

Hence, we can conclude that the group action is transitive by the construction.

Coming to defining the group action on the output space, we can clearly see that the bearing measurements are rotated by $Q^\top$. Therefore, the natural choice of the group action is given by $\rho((A, (Q, a)_i), y_i) = Q_i^\top y_i$. Under this natural definition, it follows that the group operation is equivariant.

2.5 Lift of VSLAM kinematics

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

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

where $f((P,p_i),U)$ is given by (3).

In the above expression, we are considering the tangent map of the group action function. Basically, we want to find a tangent element in $\mathfrak{vslam}_n(3)$, which replicates the constraints imposed by the static landmarks.

The following definition follows from bookkeeping. The function $\Lambda : \mathcal{T}_n^\circ(3) \times \mathbb{V} \to \mathfrak{vslam}_n(3)$, defined by

$$\Lambda((P,p_i),U) := (U, (\Lambda_Q(U, R_P^\top(p_i - x_P)), \Lambda_a(U, R_P^\top(p_i - x_P))))$$

where $\Lambda_Q : \mathfrak{se}(3) \times (\mathbb{R}^3 \backslash \{0\}) \to \mathfrak{so}(3)$ is given by

$$\Lambda_Q(U,q) := (\Omega_U + \frac{q \times V_U}{|q|^2})^\times$$

and $\Lambda_a : \mathfrak{se}(3) \times (\mathbb{R}^3 \backslash \{0\}) \to \mathfrak{mr}(1)$ is given by

$$\Lambda_a(U,q) := \frac{q^\top V_U}{|q|^2}$$

is the required lift action.

The definition of f is such that it keeps the landmarks fixed but moves the robot.

The intuition behind Λ_a is that it maps the robot velocity and the relative position of the landmark with respect to the robot to the rate of change of the distance. This rate is represented in a multiplicative fashion rather than the usual addition fashion. Therefore, a dot product is used.

Similarly, Λ_Q maps these values to the rate of change of bearing of the landmark. The term Ω_U corresponds to the angular velocity of the robot, and the second term $\frac{q \times V_U}{|q|^2}$, basically gives the additional rate of change of bearing due to the velocity and the position. The cross product is used as the rate of change of the vector's direction given by that expression.

These definitions are obtained by using the fact that the landmarks are fixed.

The reason to consider the lift of the dynamics is to allow us to shift the dynamics of the estimate from $\mathcal{T}_n^\circ(3)$ to a lie group $\mathrm{VSLAM}_n(3)$; this will allow us to define the error between the estimate of measurement and the true measurement, as done in the next section.

3 Observer Design

3.1 Lifted System Kinematics

Let $\xi^\circ \in \mathcal{T}_n^\circ(3)$ be the initial estimate of the configuration. We initialize our observer $\hat{X} : \mathbb{R}^+ \to \mathrm{VSLAM}_n(3)$, with $\hat{X} = id$. Now, we update our estimate by updating $\hat{X}(t)$ appropriately. Therefore, our estimate of the system at time t is $\Upsilon(\hat{X}(t), \xi^\circ)$. Therefore, using the idea of lift, the dynamics of $\hat{X}$ can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{X} = \hat{X}\Lambda\left(\Upsilon\left(\hat{X}, \xi^\circ\right), U\right)$$

Here, U is the velocity of the robot. The multiplication of $\hat{X}$ is used to bring the lie algebra element to the tangent space at $\hat{X}$.

3.2 Measurement Error

Using the equivariant definition of group action on the measurement space, we can define the measurement error. Using the previous definition, the estimate of the measurement will be $\hat{y} = h(\Upsilon(\hat{X}(t), \xi^\circ))$.
Let the measurements be denoted by $y \in \mathcal{N}^n(3)$, then if we want to compare $\hat{y}$ and y, it is equivalent to comparing $\rho(\hat{X}^{-1}, \hat{y})$ and $\rho(\hat{X}^{-1}, y)$. However, from equivariance, we can say that

$$\begin{aligned} \rho(\hat{X}^{-1}, \hat{y}) &= \rho(\hat{X}^{-1}, h(\Upsilon(\hat{X}(t), \xi^\circ))) \\ &= h(\xi^\circ) \end{aligned}$$

which is a constant! Therefore, the problem of VSLAM can be equivalently be said to finding a function $\hat{X}$, such that $d = \rho(\hat{X}^{-1}, y)$ converges to $h(\xi^\circ)$.

3.3 Driving Estimates to True values

The previously defined dynamics of $\hat{X}$ don't incorporate the measurements. To bring in the effect of the disparity between estimated observations and true observations, we introduce a correction term. The aim of this term is to improve our estimates of the environment. A term $\Delta_{\hat{X}} = (\Delta, (\Gamma, \gamma)_i) \in \mathfrak{vslam}_n(3)$ is introduced for the same.

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

$$\tilde{r}_i = \frac{\hat{r}_i}{r_i}$$

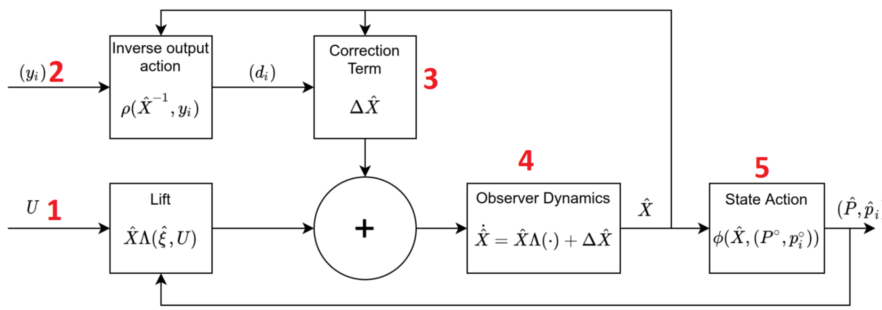

Figure 2: Block diagram of the Filter

The correction terms are defined as follows:

$$\Gamma_i := \left(\frac{d_i^\top \hat{Q}_i V_U}{\hat{r}_i\left(1 + d_i^\top y_i^\circ\right)} - \frac{k_i}{\left(1 + d_i^\top y_i^\circ\right)^2}\right)\left(d_i^\times y_i^\circ\right)^\times + \frac{1}{\hat{r}}\left(\left(y_i^\circ - d_i\right)^\times \hat{Q}_i V_U\right)^\times,$$

$$\gamma_i := \frac{\alpha_i}{\hat{r}_i^2}\left(\left(1 - d_i^\top y_i^\circ\right) d_i^\top \hat{Q}_i V_U - y_i^{\circ\top}\left(d_i \times \hat{Q}_i V_U\right)^\times d_i\right) + \frac{1}{\hat{r}_i}\left(y_i^\circ - d_i\right)^\top \hat{Q}_i V_U + \frac{\alpha_i}{\hat{r}_i}\beta_{\underline{c}}^\epsilon\left(\hat{r}_i\right),$$

Here, k_i and α_i are positive constants. $\beta_{\underline{c}}^\epsilon(\hat{r}_i)$ is a barrier function to ensure that the value of the estimated distance $\hat{r}_i$ does not get very close to zero and leaves the set $\mathbf{MR}(1)$. The barrier function is defined such that it does not make a difference when $\hat{r}_i$ is larger than $\underline{c}$. But when $\hat{r}_i$ gets close to ϵ, it blows up to infinity, ensuring that it is larger than ϵ always.

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

The choice of robot-pose correction term Δ does not have any effects on the convergence rates. However, we will next choose it in such a way that the landmark velocities are minimized.

We summarize the algorithm in five steps, as shown in the block diagram.

1. It corresponds to using the previous estimate $\hat{X}$ and the control input U to decide the rate of change of $\hat{X}$.

2. On the obtained measurements, we perform the action of $\hat{X}^{-1}$ to bring it into a better comparison term.

3. Using this new measurement, we define our correction term to drive the estimates to the true value. This corresponds to the Kalman gain element in a traditional setting.

4. Next, we use the previous two values to get $\dot{\hat{X}}$ and obtain the next value of $\hat{X}$.

5. The action of $\hat{X}$ on the initial estimate ($\hat{\xi}^\circ$) to get the current state estimate.

4 Stability of filter

Now, we will present the key theorem of the paper, which guarantees the convergence and stability of the filter.

4.1 Theorem

Consider the observer $\hat{X} \in \mathrm{VSLAM}_n(3)$. Assume that U is bounded, and that $y_i^\times y_i^\times V_U$ is persistently exciting, in the sense that there exist $T > 0$ and $\mu > 0$ such that

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

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

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

for all time. Then the landmark correction terms (19) define an almost semi-globally asymptotically stabilising (Def. A.1) correction term for the error dynamics of $d_i, \tilde{r}_i(16, 17)$ around the equilibrium $(y_i^\circ, 1)$ with an exception set

$$\chi = \left\{(d, \tilde{r})_i \in \left(\mathrm{S}^2 \times \mathbb{R}_+\right)^n \mid d_i = -y_i^\circ \ \text{ for some} \\ i \in [1, \dots, n]\right\}.$$

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

4.2 Proof Summary

The proof works in following steps:

1. Choosing parameters of the correction terms, as they depend on the initial conditions.
2. It is shown that the observer equations are well-defined, for all time.
3. A storage function is introduced, which is shown to be non-increasing in time for the error dynamics.
4. Persistence of excitation and Barbalat's lemma are used to that the storage function will converge to zero over time.
5. Convergence of error to zero, is equivalent to the convergence of the true and estimated states.

4.3 Defining the robot-pose correction term Δ

Following the proof of the theorem, one can see that conditions on Δ are free, i.e., we just need continuity of it to complete the proof. Hence, an ambiguity of the inertial frame occurs. This makes intuitive sense, as the choice of the ground frame should not affect the SLAM algorithm because all the measurements and state evolution are independent of the choice of the inertial frame.

As we know that landmarks are fixed, we can impose a condition on our algorithm to reinforce this on the estimates by choosing an inertial frame such that the sum of the squares of the velocities of the landmarks is minimum, i.e., $\sum_{i=1}^{n} \kappa_i ||\dot{\hat{p}}_i||^2$. This ensures convergence to a frame such that the landmarks are stationary. This has been done in the last proposition of the article.