Project 3: Visual-Inertial SLAM

Collaboration in the sense of discussion is allowed, however, the work you turn in should be your own - you should not split parts of the assignments with other students and you should certainly not copy other students' solutions or code. See the collaboration and academic integrity statement here: https://natanaso.github.io/ece276a. *Books may be consulted but not copied from.*

Submission

You should submit the following files on **Gradescope** by the deadline shown at the top right corner.

1. Theoretical problems: upload your solutions to Problem 1-3. You may use latex, scanned handwritten notes (write legibly!), or any other method to prepare a pdf file. Do not just write the final result. Present your work in detail, explaining your approach at every step.

2. Programming assignment: upload all code you have written for the project (do not include the training and test datasets) and a README file with clear description of each file.

3. Report: upload your report. You are encouraged but not required to use an IEEE conference template[1] for your report.

Problems

In square brackets are the points assigned to each problem.

1. [26 pts] Consider the bicycle shown in Fig. 1, in which the diameter of the front wheel is twice that of the rear wheel. Frames $\{a\}$ and $\{b\}$ are attached, respectively, to the centers of the wheels with the axes $\hat{y}_a$ and $\hat{y}_b$ aligned. Frame $\{c\}$ is attached to the top of the front wheel and its distance to frame $\{b\}$ is D in the $\hat{x}$ direction. Assuming that the bike moves forward in the $\hat{y}$-direction, find the transformation ${}_{\{a\}}T_{\{c\}}$ from frame $\{c\}$ to frame $\{a\}$ as a function of the front wheel's rotation angle θ. Assume that $\theta = 0$ at the instant shown in Fig. 1.

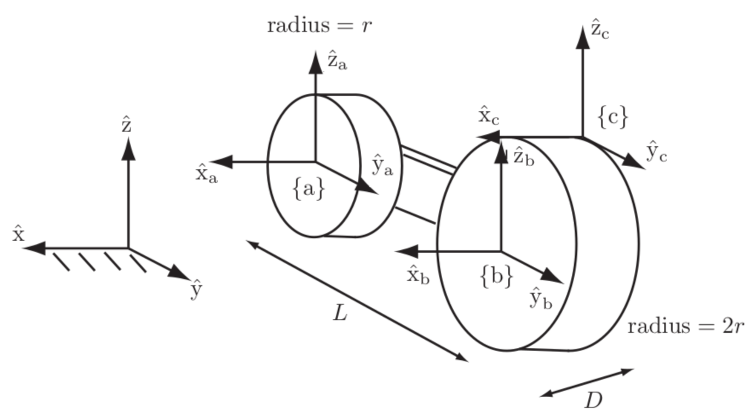

Figure 1: A bicycle with a large front wheel.

2. [26 pts] In class, we developed expressions for the matrix exponential for spatial motions mapping elements from $\mathfrak{so}(3)$ to $SO(3)$ and elements of $\mathfrak{se}(3)$ to $SE(3)$. Similarly, we showed how to compute the matrix logarithm going the other direction. In this problem, your objective is to write down the explicit matrix exponential from $\mathfrak{so}(2)$ to $SO(2)$ and from $\mathfrak{se}(2)$ to $SE(2)$ and the logarithm function going back. For the $\mathfrak{so}(2)$ to $SO(2)$ case there is a single exponential coordinate, while for the $\mathfrak{se}(2)$ to $SE(2)$ case there are three exponential coordinates, corresponding to a planar translation and angle of rotation.

[1] https://www.ieee.org/conferences_events/conferences/publishing/templates.html

3. [26 pts] Consider a mobile robot moving around on the xy-plane with motion model:

$$\mathbf{s}_{t+1} := \begin{bmatrix} x_{t+1} \\ y_{t+1} \\ \theta_{t+1} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \\ \theta_t \end{bmatrix} + \tau \begin{bmatrix} \cos\theta_t & 0 \\ \sin\theta_t & 0 \\ 0 & 1 \end{bmatrix} \left(\begin{bmatrix} v_t \\ \omega_t \end{bmatrix} + \mathbf{w}_t \right)$$

where $\mathbf{s}_t \in SE(2)$ is the robot state at time t, $\mathbf{u}_t := [v_t,\ \omega_t]^T \in \mathbb{R}^2$ is the control input, $\tau > 0$ is the time discretization, and $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{W})$ is Gaussian control noise with covariance $\mathbf{W} \in \mathbb{R}^{2\times 2}$. Suppose that the robot is equipped with a sensor that measures the range and bearing to the origin, according to the observation model:

$$\mathbf{z}_t := \begin{bmatrix} r_t \\ \phi_t \end{bmatrix} = \begin{bmatrix} \sqrt{x_t^2 + y_t^2} \\ \mathbf{atan2}\left(-y_t, -x_t\right) - \theta_t \end{bmatrix} + \boldsymbol{\eta}_t$$

where $\mathbf{z}_t$ is the sensor observation at time t and $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \mathbf{V})$ is Gaussian measurement noise with covariance $\mathbf{V} \in \mathbb{R}^{2\times 2}$. Let the prior distribution of the robot state at time t be $\mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t}\right)$ and assume that the robot moves before observing, i.e., the motion model is applied before the observation model.

(a) Work out the equations of the prediction step of the Extended Kalman Filter, needed to compute the distribution $\mathcal{N}\left(\boldsymbol{\mu}_{t+1|t}, \boldsymbol{\Sigma}_{t+1|t}\right)$. Make sure that the equations you write are specific to the robot motion model specified above.

(b) Work out the equations of the update step of the Extended Kalman Filter, needed to compute the distribution $\mathcal{N}\left(\boldsymbol{\mu}_{t+1|t+1}, \boldsymbol{\Sigma}_{t+1|t+1}\right)$. Make sure that the equations you write are specific to the robot observation model specified above.

(c) Derive the prediction and update steps for the information matrix. In other words, let $\boldsymbol{\Omega}_{t|t} := \boldsymbol{\Sigma}_{t|t}^{-1}$ and derive the prediction equations for $\boldsymbol{\Omega}_{t+1|t}$ as well as the update equations for $\boldsymbol{\Omega}_{t+1|t+1}$.

4. [140 pts] Implement visual-inertial simultaneous localization and mapping (SLAM) using the Extended Kalman Filter. You are provided with synchronized measurements from an IMU and a stereo camera as well as the intrinsic camera calibration and the extrinsic calibration between the two sensors, specifying the transformation from the IMU to the left camera frame. The data includes:

- **IMU Measurements**: linear velocity $\mathbf{v}_t \in \mathbb{R}^3$ and angular velocity $\boldsymbol{\omega}_t \in \mathbb{R}^3$ measured in the body frame of the IMU
- **Stereo Camera Images**: pixel coordinates $\mathbf{z}_t \in \mathbb{R}^{4\times M}$ of detected visual features with precomputed correspondences between the left and the right camera frames (see Fig. 2). The landmarks i that were not observable at time t have a measurement of $\mathbf{z}_{t,i} = \begin{bmatrix} -1 & -1 & -1 & -1 \end{bmatrix}^T$. Videos showing the feature tracking over time are provided in the data.
- **Time Stamps**: time stamps τ in UNIX standard seconds-since-the-epoch January 1, 1970
- **Intrinsic Calibration**: stereo baseline b and camera calibration matrix:

$$\mathbf{K} := \begin{bmatrix} fs_u & 0 & c_u \\ 0 & fs_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

- **Extrinsic Calibration**: the transformation ${}_CT_I \in SE(3)$ from the IMU to left camera frame

The goal of the project is to use the EKF equations with a prediction step based on $SE(3)$ kinematics and an update step based on the stereo camera observation model to perform localization and mapping. In detail, you should complete the following tasks:

(a) **IMU Localization via EKF Prediction**: Implement the EKF prediction step based on the $SE(3)$ kinematics to estimate the pose $T_t \in SE(3)$ of the IMU over time t

Figure 2: Visual features matched across the left-right camera frames (left) and across time (right).

(b) **Landmark Mapping via EKF Update**: assume that the predicted IMU trajectory from part (a) above is correct and focus on estimating the landmark positions. In detail, you should implement an EKF with the unknown landmark positions $\mathbf{m} \in \mathbb{R}^{3\times M}$ as a state and perform EKF update step after every visual observation $\mathbf{z}_t$ in order to keep track of the mean and covariance of $\mathbf{m}$. Note that we are assuming that the landmarks are static so it is not necessary to implement a prediction step. Moreover, since the sensor does not move sufficiently along the z-axis, the estimation for the z coordinate of the landmarks will not be very good. You can assume that the z coordinates for all landmarks are 0 and focus only on estimating their xy coordinates.

(c) **Extra Credit: Visual-Inertial SLAM**: combine the IMU prediction step from part (a) with the landmark update step from part (b) and an IMU update step based on the stereo camera observation model to obtain a complete visual-inertial SLAM algorithm. Compare the accuracy of the SLAM implementation versus the individual parts obtained in part (a) and part (b).

5. [32 pts] Write a project report describing your approach to the visual-inertial SLAM problem. Your report should include the following sections:
   - **Introduction**: discuss why the problem is important and present a brief overview of your approach
   - **Problem Formulation**: state the problem you are trying to solve in mathematical terms. This section should be short and clear and should rigorously define the quantities you are interested in.
   - **Technical Approach**: describe your approach to visual-inertial localization and mapping.
   - **Results**: present your results, and discuss them – what worked, what did not, and why. Make sure your results include plots clearly showing the estimated robot trajectory as well as the estimated 2-D positions of the visual features. If you have videos do include them in the zip file and refer to them in your report!