Visual Inertial SLAM using Extended Kalman Filter

Rishabh Bhattacharya
Department of Mechanical & Aerospace Engineering
University of California, San Diego
ribhattacharya@ucsd.edu

***Index Terms*—kalman filter, bayesian filter, localization, mapping, SLAM, odometry, stereo camera**

I. Introduction

Any robot in a complex dynamic system has to be able to localize itself w.r.t the environment, in order to ensure operability and safety guarantees. In most cases, prior information about the environment is not known. So the task now involves estimating the position of the robot and simultaneously mapping the environment. This needs to be done in real time, which adds further computational challenges to this problem.

This can only be achieved by providing some sort of sensory information about the environment to the robot, and make some predictions about the robot state. Lidar and stereo cameras can provide obstacle data, while FOG, IMU (inertial measurement unit) and magnetic encoders can provide odometry related data for the robot (eg. linear accelerations, angular velocity, change in roll/pitch/yaw angles, total rotations of the wheel, etc).

In this project, we will use stereo camera data to map the environment. We will predict the state of the robot using IMU data (linear and angular velocity; linear velocity has been provided in the data instead of the regular linear acceleration data from IMU), while using the stereo camera observations to correct it.

A crucial aspect to note here is the fact that we avoid any sort of deterministic approaches, since any and all sensors have inherent errors. Also, any control input given to the robot (eg. move ahead by 1 m) will not be implemented perfectly since the distance travelled depends on the wheel diameters (which can vary according to tire pressure, temperature, etc). This is the reason we approach this problem, with a probabilistic approach.

Bayes filter is a probabilistic inference approach for estimating a state $\mathbf{x_t}$ given a set of controls $\mathbf{u}_{0:t}$ and observations $\mathbf{z}_{0:t}$. Particle filter is just a discrete formulation of the same problem. For the purposes of SLAM in this project, we will use a kalman filter implementation. Kalman filter is a special case of Bayes filter, with the assumptions that (i) prior pdf, motion model noise and observation model noise are **gaussian**, (ii) the motion model and observation model are also **linear in the state** $\mathbf{x_t}$ and (iii) the noises are **independent** to each other and to the state $\mathbf{x_t}$ for all time t. In the cases where the motion/observation models are non-linear in $\mathbf{x_t}$, the predicted and updated pdfs are **forced to be gaussian** via approximation, while keeping the other assumptions intact. This is called an Extended Kalman Filter, which has been implemented for this project.

II. Problem formulation

We break down the overall VI SLAM problem into 3 parts. Initially, we will implement an IMU prediction (II-A) step, which will help us get a trajectory of our robot poses w.r.t time. Next we assume this trajectory to be true, and implement a mapping only update step (II-B, landmarks are static, so no prediction is required). Finally, we combine the IMU prediction step from part (a) with the landmark update step from part (b) and implement an IMU update step based on the stereo-camera observation model to obtain a complete visual-inertial SLAM algorithm (II-C).

A. IMU Localization via EKF Prediction

Assumptions:

1) linear velocity $\mathbf{v}_t \in \mathbb{R}^3$ and angular velocity $\omega_t \in \mathbb{R}^3$ measurements are available

Objective: given IMU measurements $\mathbf{u}_{0:T}$ with $\mathbf{u}_t := \left[\mathbf{v}_t^\top, \omega_t^\top\right]^\top \in \mathbb{R}^6$, estimate the pose $T_t := {}_W T_{I,t} \in SE(3)$ of the IMU over time

B. Landmark Mapping via EKF Update

Assumptions:

1) the IMU pose $T_t := {}_W T_{l,t} \in SE(3)$ is known
2) the data association $\Delta_t : \{1, \dots, M\} \to \{1, \dots, N_t\}$ stipulating that landmark j corresponds to observation $\mathbf{z}_{t,i} \in \mathbb{R}^4$ with $i = \Delta_t(j)$ at time t is known or provided by an external algorithm
3) the landmarks $\mathbf{m}$ are static, i.e., it is not necessary to consider a motion model or a prediction step for $\mathbf{m}$

Objective: given the observations $z_t := \begin{bmatrix} z_{t,1}^\top & \cdots & z_{t,N_t}^\top \end{bmatrix}^\top \in \mathbb{R}^{4N_t}$ for $t = 0, \dots, T$, estimate the coordinates $\mathbf{m} := \begin{bmatrix} \mathbf{m}_1^\top & \dots & \mathbf{m}_M^\top \end{bmatrix}^\top \in \mathbb{R}^{3M}$ of the landmarks that generated them

C. Visual-Inertial SLAM

Assumptions:

1) linear velocity $\mathbf{v}_t \in \mathbb{R}^3$ and angular velocity $\omega_t \in \mathbb{R}^3$ measurements are available
2) known world-frame landmark coordinates $\mathbf{m} \in \mathbb{R}^{3M}$
3) the data association $\Delta_t : \{1, \dots, M\} \to \{1, \dots, N_t\}$ stipulating that landmark j corresponds to observation

$\mathbf{z}_{t,i} \in \mathbb{R}^4$ with $i = \Delta_t(j)$ at time t is known or provided by an external algorithm

Objective: given IMU measurements $\mathbf{u}_{0:T}$ with $\mathbf{u}_t := \left[\mathbf{v}_t^\top, \omega_t^\top\right]^\top \in \mathbb{R}^6$ and feature observations $\mathbf{z}_{0:T}$, estimate the pose $T_t := {}_W T_{I,t} \in SE(3)$ of the IMU over time and the landmark positions $\mathbf{m} \in \mathbb{R}^{3M}$ w.r.t world frame.

III. Technical Approach

This section presents a detailed approach to the problem at hand. We present the IMU prediction and update steps in (III-A), followed by the landmark update problem (III-B). We combine both the update formulations in (III-C). The overall SLAM problem would involve prediction steps from III-A and update steps from III-C. The quantity calculations are discussed at length in the corresponding sections.

A. IMU Localization via EKF Prediction & Update

Prior: $T_t \mid \mathbf{z}_{0:t}, \mathbf{u}_{0:t-1} \sim \mathcal{N}\left(\boldsymbol{\mu}_{t|t}, \Sigma_{t|t}\right)$ with $\boldsymbol{\mu}_{t|t} \in SE(3)$ and $\Sigma_{t|t} \in \mathbb{R}^{6\times 6}$. This means that

$$T_t = \boldsymbol{\mu}_{t|t} \exp\left(\delta\hat{\boldsymbol{\mu}}_{t|t}\right) \tag{1}$$

with $\delta\boldsymbol{\mu}_{t|t} \sim \mathcal{N}\left(0, \Sigma_{t|t}\right)$. $\Sigma_{t|t}$ is 6×6 because only the 6 degrees of freedom of T_t are changing

Motion Model: given by nominal kinematics of $\boldsymbol{\mu}_{t|t}$ and perturbation kinematics of $\delta\boldsymbol{\mu}_{t|t}$ with time discretization τ_t :

$$\begin{aligned} \boldsymbol{\mu}_{t+1|t} &= \boldsymbol{\mu}_{t|t} \exp\left(\tau_t \hat{\mathbf{u}}_t\right) \\ \delta\boldsymbol{\mu}_{t+1|t} &= \exp\left(-\tau_t \hat{\mathbf{u}}_t^{\curlywedge}\right) \delta\boldsymbol{\mu}_{t|t} + \mathbf{w}_t \end{aligned} \tag{2}$$

EKF Prediction Step: with $\mathbf{w}_t \sim \mathcal{N}(0, W)$.

$$\begin{aligned} \boldsymbol{\mu}_{t+1|t} &= \boldsymbol{\mu}_{t|t} \exp\left(\tau_t \hat{\mathbf{u}}_t\right) \\ \Sigma_{t+1|t} &= \exp\left(-\tau \hat{\mathbf{u}}_t^{\curlywedge}\right) \Sigma_{t|t} \exp\left(-\tau \hat{\mathbf{u}}_t^{\curlywedge}\right)^\top + W \end{aligned} \tag{3}$$

where,

$$\mathbf{u}_t := \begin{bmatrix} \mathbf{v}_t \\ \omega_t \end{bmatrix} \in \mathbb{R}^6, \quad \hat{\mathbf{u}}_t := \begin{bmatrix} \hat{\omega}_t & \mathbf{v}_t \\ \mathbf{0}^\top & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4}$$

$$\hat{\mathbf{u}}_t^{\curlywedge} := \begin{bmatrix} \hat{\boldsymbol{\omega}}_t & \hat{\mathbf{v}}_t \\ 0 & \hat{\boldsymbol{\omega}}_t \end{bmatrix} \in \mathbb{R}^{6\times 6}$$

Observation Model: with measurement noise $\mathbf{v}_t \sim \mathcal{N}(0, V)$ and world frame landmark coordinates $\mathbf{m_j} \in \mathbb{R}^3$,

$$\begin{aligned} \mathbf{z}_{t+1,i} &= h\left(T_{t+1}, \mathbf{m}_j\right) + \mathbf{v}_{t+1,i} \\ &:= K_s \pi\left({}_o T_l T_{t+1}^{-1} \underline{\mathbf{m}}_j\right) + \mathbf{v}_{t+1,i} \end{aligned} \tag{4}$$

where $\underline{\mathbf{m}}_j \in \mathbb{R}^4$ is the homogeneous coordinate

Predicted observation:

$$\tilde{\mathbf{z}}_{t+1,i} := K_s \pi\left(o T_l \mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j\right) \quad \text{for } i = 1, \ldots, N_{t+1}$$

Jacobian of $\tilde{\mathbf{z}}_{t+1,i}$ with respect to T_{t+1} evaluated at $\boldsymbol{\mu}_{t+1|t}$:

$$\begin{aligned} &H_{t+1,i} = \\ &-K_s \frac{d\pi}{d\mathbf{q}}\left(o T_I \mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j\right) o T_I \left(\mu_{t+1|t}^{-1} \underline{\mathbf{m}}_j\right)^{\odot} \in \mathbb{R}^{4\times 6} \end{aligned} \tag{5}$$

EKF Update Step:

$$\begin{aligned} K_{t+1} &= \Sigma_{t+1|t} H_{t+1}^\top \left(H_{t+1} \Sigma_{t+1|t} H_{t+1}^\top + I \otimes V\right)^{-1} \\ \mu_{t+1|t+1} &= \mu_{t+1|t} \exp\left(\left(K_{t+1}\left(\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1}\right)\right)^{\wedge}\right) \\ \Sigma_{t+1|t+1} &= \left(I - K_{t+1} H_{t+1}\right) \Sigma_{t+1|t} \end{aligned} \tag{6}$$

where,

$$H_{t+1} = \begin{bmatrix} H_{t+1,1} \\ \vdots \\ H_{t+1,N_{t+1}} \end{bmatrix} \in \mathbb{R}^{4N_t \times 6}$$

and

$$\begin{bmatrix} \mathbf{s} \\ 1 \end{bmatrix}^{\odot} := \begin{bmatrix} I & -\hat{\mathbf{s}} \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 6}$$

B. Landmark Mapping via EKF Update

Prior: $\mathbf{m} \mid \mathbf{z}_{0:t} \sim \mathcal{N}\left(\mu_t, \Sigma_t\right)$ with $\mu_t \in \mathbb{R}^{3M}$ and $\Sigma_t \in \mathbb{R}^{3M\times 3M}$, where $\mathbf{m}$ is the world coordinate of a landmark

Observation Model: with measurement noise $\mathbf{v}_{t,i} \sim \mathcal{N}(0, V)$

$$\mathbf{z}_{t,i} = h\left(T_t, \mathbf{m}_j\right) + \mathbf{v}_{t,i} := K_s \pi\left(o T_l T_t^{-1} \underline{\mathbf{m}}_j\right) + \mathbf{v}_{t,i}$$

where $\underline{\mathbf{m}}_j := \begin{bmatrix} \mathbf{m}_j \\ 1 \end{bmatrix}$ are the homogeneous world frame coordinates of the landmarks and

$$\pi(\mathbf{q}) := \frac{1}{q_3}\mathbf{q} \in \mathbb{R}^4, \quad \frac{d\pi}{d\mathbf{q}}(\mathbf{q}) = \frac{1}{q_3}\begin{bmatrix} 1 & 0 & -\frac{q_1}{q_3} & 0 \\ 0 & 1 & -\frac{q_2}{q_3} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{q_4}{q_3} & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}$$

where $\mathbf{q} \in \mathbb{R}^{3\times 4}$ is a vector. All observations are stacked as a $4N_t$ vector, at time t with notation abuse:

$$\mathbf{z}_t = K_s \pi\left(o T_I T_t^{-1} \underline{\mathbf{m}}\right) + \mathbf{v}_t \tag{7}$$

$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, I \otimes V), \quad I \otimes V := \begin{bmatrix} V & & \\ & \ddots & \\ & & V \end{bmatrix}$$

For our project,

$$K_s = \begin{bmatrix} 552.55 & 0 & 682.05 & 0 \\ 0 & 552.55 & 238.77 & 0 \\ 552.55 & 0 & 682.05 & -331.53 \\ 0 & 552.55 & 238.77 & 0 \end{bmatrix}$$

and

$$\begin{aligned} o T_I T_t^{-1} &= \left({}_{world} T_{imu} \times {}_{imu} T_{cam}\right)^{-1} \\ &= \left(T_t \times {}_{imu} T_{cam}\right)^{-1} \end{aligned}$$

Σ^{IMU}	$10^{-5} * \mathbf{I}^{6\times6}$
Σ^{MAP}	$10^{-5} * \mathbf{I}^{3M\times3M}$
W_p	$10^{-4} * \mathbf{I}^{6\times6}$
V	$10^{-2} * \mathbf{I}^{4\times4}$

TABLE I: Input parameters

where T_t is the IMU pose from prediction step and ${}_{imu}T_{cam}$ is given.

EKF Update: given a new observation $\mathbf{z}_{t+1} \in \mathbb{R}^{4N_{t+1}}$:

$$\begin{aligned}
\tilde{\mathbf{z}}_{t+1} &= K_s \pi \left({}_oT_I T_{t+1}^{-1} \underline{\mu}_t \right) + \mathbf{v}_t \in \mathbb{R}^4 \\
K_{t+1} &= \Sigma_t H_{t+1}^\top \left(H_{t+1} \Sigma_t H_{t+1}^\top + I \otimes V \right)^{-1} \\
\mu_{t+1} &= \mu_t + K_{t+1} (\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1}) \\
\Sigma_{t+1} &= (I - K_{t+1} H_{t+1}) \Sigma_t
\end{aligned} \tag{8}$$

$\tilde{\mathbf{z}}_{t+1} \in \mathbb{R}^{4N_{t+1}}$ is the predicted observation based on the landmark position estimates μ_t at time t.
We need the observation model Jacobian $H_{t+1} \in \mathbb{R}^{4N_t \times 3M}$ evaluated at μ_t with block elements $H_{t+1,i,j} \in \mathbb{R}^{4\times3}$:

$$H_{t+1,i,j} = \begin{cases} K_s \frac{d\pi}{dq} \left({}_oT_I T_{t+1}^{-1} \underline{\mu}_{t,j} \right) {}_oT_I T_{t+1}^{-1} P^\top & \text{if } \Delta_t(j) = i \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $P = \begin{bmatrix} I & 0 \end{bmatrix} \in \mathbb{R}^{3\times4}$

C. Visual-Inertial SLAM

In this part, we combine the above two update steps to form the overall approach to our problem. Instead of updating separately, we combine the update equations for the pose and map into a single formulation for $\mathbf{H}, \mathbf{K}$ & $\mathbf{\Sigma}$

Jacobian matrix:

$$\mathbf{H}_{t+1|t} = \begin{bmatrix} \mathbf{H}_{t+1|t}^{MAP} & \mathbf{H}_{t+1|t}^{IMU} \end{bmatrix} \in \mathbb{R}^{4N_t \times (3M+6)} \tag{10}$$

EKF Update:

$$\begin{aligned}
\mathbf{K}_{t+1|t} &= \mathbf{\Sigma}_{t+1|t} \mathbf{H}_{t+1|t}^T \left(\mathbf{H}_{t+1|t} \mathbf{\Sigma}_{t+1|t} \mathbf{H}_{t+1|t}^T + \mathbf{I} \otimes \mathbf{V} \right)^{-1} \\
\boldsymbol{\mu}_{t+1|t+1}^{IMU} &= \boldsymbol{\mu}_{t+1|t}^{IMU} + \mathbf{K}_{t+1|t} \left(\mathbf{z}_t - \tilde{\mathbf{z}}_t \right) \\
\boldsymbol{\mu}_{t+1|t+1}^{MAP} &= \boldsymbol{\mu}_{t+1|t}^{MAP} \exp \left(\left(\mathbf{K}_{t+1|t} \left(\mathbf{z}_{t+1} - \tilde{\mathbf{z}}_{t+1} \right) \right)^\wedge \right) \\
\mathbf{\Sigma}_{t+1|t+1} &= \left(\mathbf{I} - \mathbf{K}_{t+1|t} \mathbf{H}_{t+1|t} \right) \mathbf{\Sigma}_{t+1|t}
\end{aligned} \tag{11}$$

IV. Results

The following results were obtained using the parameters in Table I. Figures 1, 2 depict the results for IMU prediction steps. Figures 3 & 4 show the prediction + update steps for both the datasets. Figures 5 & 6 show the VI SLAM implementation for both the datasets provided. Figures 7 & 8 are a comparison between the prediction + update steps when the motion model noise is varied. Discussions have been provided along with hypothesis and possible reasoning beneath every figure.

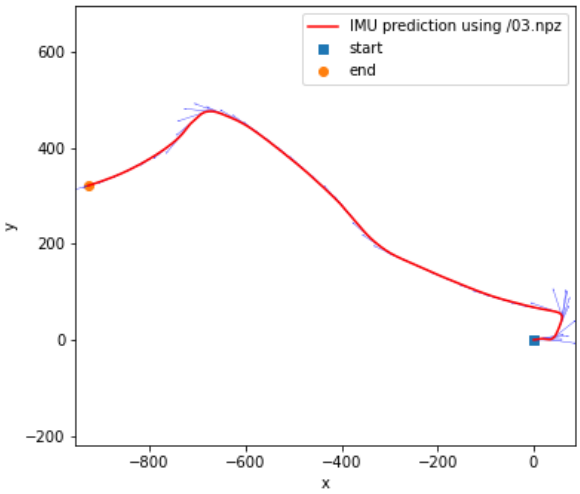

Fig. 1: **IMU prediction for 03.npz**: This is the IMU prediction step without any update steps. We see that the trajectory matches with the video made available.

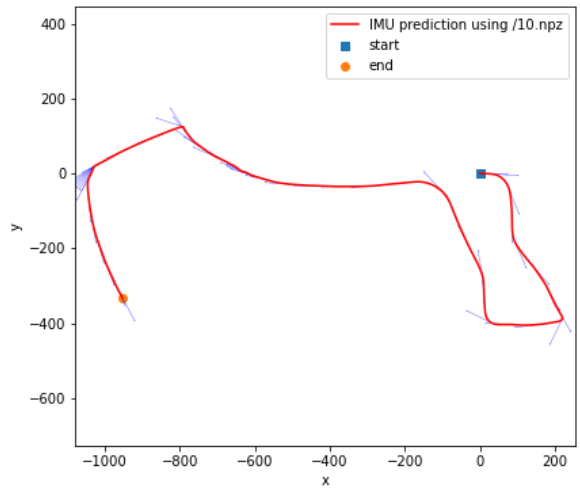

Fig. 2: **IMU prediction for 10.npz**: This is the IMU prediction step without any update steps. We see that the trajectory matches with the video made available.

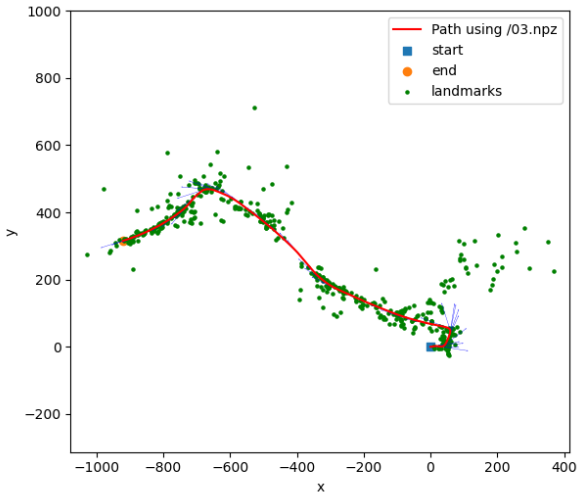

Fig. 3: **IMU prediction & map update for 03.npz; W = 10^{-3}, Using 500 features**: This plot was obtained by plotting the results from IMU prediction and map update independently. The IMU trajectory correlates well with 1.

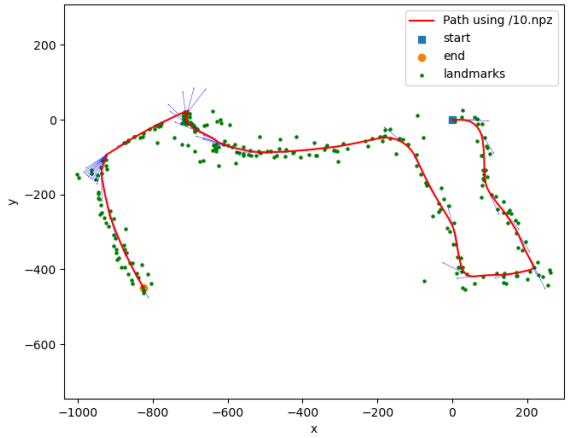

Fig. 4: **IMU prediction & map update for 10.npz; W = 10^{-3}, Using 500 features**: This plot was obtained by plotting the results from IMU prediction and map update independently. The IMU trajectory correlates well with 2.

V. Conclusions

We have attempted to implement a Visual Inertial SLAM problem using Extended Kalman Filter. Even though our results match for one dataset, we have not been able to solve the problem for the second result. This can be attributed to a variety of reasons, like non-invertible matrices, high variance, etc. The next stage in this project would be to better improve the runtime and accommodate more test cases to figure out the issue.

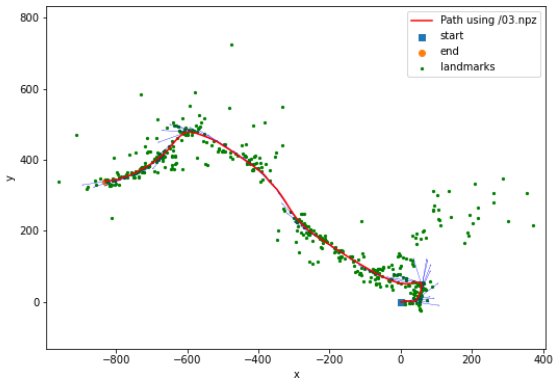

Fig. 5: **VI SLAM for 03.npz, Using 500 features**: We observe that there is a good correlation between the trajectory/map obtained from VI SLAM as compared to the dead-reckoning inputs (1). Even though the plots look similar, the VI SLAM approach is more correct since it includes the covariances amongst our odometry and observations. This might be a special case where IMU sensor data was quite accurate, but in general sensor data alone cannot be trusted.

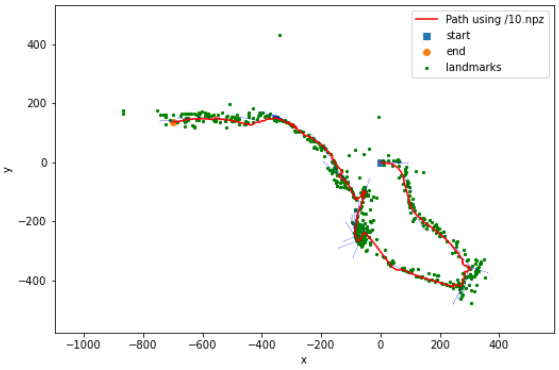

Fig. 6: **VI SLAM for 10.npz, Using 500 features**: This plot was obtained only for 2000 time steps out of 3000 steps due to runtime complexity. We see that the results are not as expected. This might be attributed to various reason, some of them being high variance, non-invertible Kalman gain matrix, etc.

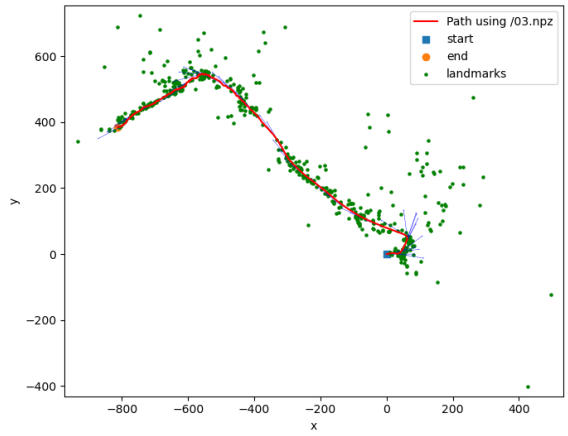

Fig. 7: **IMU prediction & map update for 03.npz; W =** 10^{-2}**, Using 500 features**: This plot was obtained by plotting the results from IMU prediction and map update independently. We can see that compared to 3, the plot doesn't have many changes. This can be attributed to the fact that 03.npz has only 1000 time steps, compared to 3000 for 10.npz. Thus the error does not have a enough time to accumulate and result in a deviation.

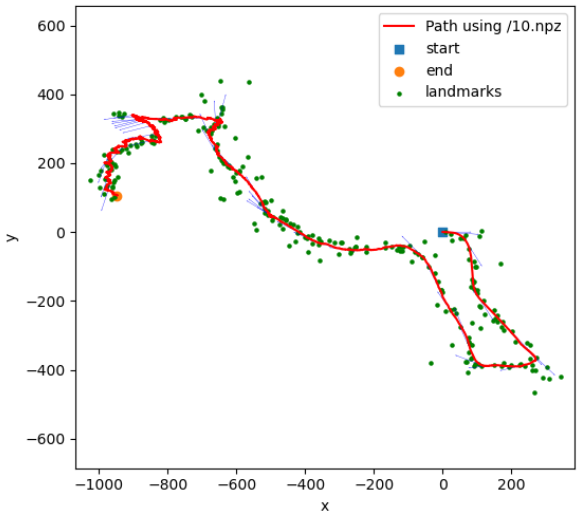

Fig. 8: **IMU prediction & map update for 10.npz; W =** 10^{-2}**, Using 500 features**: This plot was obtained by plotting the results from IMU prediction and map update independently. We can see that compared to 4, the plot changes a lot in the later time steps. This may be attributed to the fact that the error accumulates over a longer period in data 10.npz (3000 time steps) compared to data 03.npz (1000 time steps).