Q-Learning Integrated EKF-SLAM for a Unicycle Robot

Group 5

Arun Munagala
Anuj Prakash
Pranav Venkatadhri Pandalkudi Balaji

December 15, 2025

Abstract

This project presents an integrated Extended Kalman Filter (EKF) for Simultaneous Localization and Mapping (SLAM) combined with Q-Learning-based trajectory control for a unicycle robot. The robot navigates a racetrack environment with fifteen randomly placed landmarks using a range-bearing sensor affected by Gaussian noise. Q-Learning learns optimal angular velocity adjustments based on lateral and heading errors, improving trajectory following while EKF-SLAM performs localization and mapping. Simulation results demonstrate accurate landmark mapping, stable SLAM convergence, and improved control performance over multiple laps.

1 Introduction

Simultaneous Localization and Mapping (SLAM) is a fundamental problem in mobile robotics, where a robot must construct a map of an unknown environment while simultaneously estimating its own pose. Uncertainty in motion execution and sensor measurements makes SLAM inherently probabilistic, motivating the use of Bayesian state estimation techniques.

The Extended Kalman Filter (EKF) is a classical solution to SLAM problems involving Gaussian noise and nonlinear system dynamics. EKF-SLAM maintains a joint probability distribution over the robot pose and landmark locations, allowing consistent localization and mapping in moderately sized environments. Despite scalability limitations, EKF-SLAM remains widely used due to its interpretability and strong theoretical foundations.

Reinforcement Learning (RL), particularly model-free approaches such as Q-Learning, offers a complementary paradigm for control. Q-Learning enables a robot to learn control policies through interaction with the environment, without requiring an explicit model of system dynamics. When applied to trajectory tracking, Q-Learning can adaptively correct control actions in response to observed errors.

This project integrates EKF-SLAM with Q-Learning-based trajectory control for a unicycle robot navigating a racetrack environment. EKF-SLAM focuses on accurate localization and mapping, while Q-Learning learns angular velocity corrections to reduce lateral and heading errors. The integration of learning-based control with probabilistic state estimation results in improved navigation stability and robust SLAM performance over multiple laps.

2 Problem Formulation

2.1 Robot and Environment

The robot follows a unicycle motion model in a two-dimensional plane. The robot state at time t is defined as

$$\mathbf{x}_t = \begin{bmatrix} x_t & y_t & \theta_t \end{bmatrix}^T, \tag{1}$$

where (x_t, y_t) denotes the robot position and θ_t represents its orientation.

The environment contains $N = 15$ static landmarks distributed around an oval racetrack. Each landmark i is represented by its Cartesian coordinates

$$\mathbf{m}_i = \begin{bmatrix} m_{ix} & m_{iy} \end{bmatrix}^T. \tag{2}$$

2.2 Joint EKF-SLAM State

EKF-SLAM maintains a joint state vector consisting of the robot pose and all landmark positions:

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{x}_t & \mathbf{m}_1 & \cdots & \mathbf{m}_N \end{bmatrix}^T. \tag{3}$$

2.3 Control Inputs and Noise

The control input applied to the robot is

$$\mathbf{u}_t = \begin{bmatrix} v & \omega_t \end{bmatrix}^T, \tag{4}$$

where the linear velocity v is fixed at 1.0 m/s and the angular velocity ω_t is adjusted by the Q-Learning controller.

The system is affected by Gaussian process and measurement noise:

$$\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \tag{5}$$

with

$$\mathbf{Q} = \mathrm{diag}(0.05^2, 0.02^2), \quad \mathbf{R} = \mathrm{diag}(0.1^2, 0.05^2). \tag{6}$$

2.4 Sensor Model and Assumptions

The robot is equipped with a range-bearing sensor that provides distance and angular measurements to visible landmarks within a limited sensing range. Measurements are corrupted by zero-mean Gaussian noise, and landmarks outside the sensing range are not observed.

The following assumptions are made:

- Landmarks are static and uniquely identifiable.
- The environment is planar and free of dynamic obstacles.
- Process and measurement noise are Gaussian.
- Data association is performed using Mahalanobis distance gating.

These assumptions are standard in EKF-SLAM literature and allow a clear analysis of the interaction between learning-based control and probabilistic estimation.

3 Q-Learning Algorithm

3.1 State Representation

The Q-Learning state is defined using two error metrics:

$$e_{lat} = \text{signed distance to the reference path}, \tag{7}$$

$$e_\theta = \text{atan2}(t_y - y, t_x - x) - \theta. \tag{8}$$

These errors capture the robot's deviation from the desired trajectory in both position and orientation. The continuous error values are discretized into finite bins to construct a manageable state space.

3.2 Q-Update Rule

The Q-Learning update rule is given by

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right], \tag{9}$$

where α is the learning rate and γ is the discount factor.

The reward function penalizes tracking errors:

$$r = -\left(|e'_{lat}| + 0.5|e'_\theta| \right). \tag{10}$$

This formulation encourages smooth trajectory following while discouraging large heading deviations.

3.3 Learning Strategy

An ϵ-greedy exploration strategy is used to balance exploration and exploitation during training. Early episodes emphasize exploration to discover effective control actions, while later episodes favor exploitation of learned policies. The Q-Learning controller operates independently of the SLAM process, making it robust to moderate localization uncertainty.

4 EKF-SLAM Algorithm

4.1 Prediction Step

The robot motion model is

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t) + \mathbf{w}_t. \tag{11}$$

4.2 Update and Data Association

Following prediction, sensor measurements are incorporated using the EKF update step. Expected measurements are computed for existing landmarks, and the Mahalanobis distance is used to associate observations with the most likely landmark. Measurements that do not satisfy the gating criterion result in new landmark initialization.

To improve numerical stability, covariance updates are performed using the Joseph form, ensuring symmetry and positive semi-definiteness of the covariance matrix.

5 Implementation Details

The system is implemented in Python using NumPy for matrix operations. The simulation runs for 1000 time steps on a figure-eight trajectory. All angular quantities are normalized to $[-\pi, \pi]$, and small diagonal regularization terms are added to the covariance matrix to prevent numerical singularities.

Algorithm 1 EKF-SLAM Prediction Step

```
**Input:** $\mathbf{x}_{t-1}, \mathbf{\Sigma}_{t-1}, \mathbf{u}_t, \Delta t$
**Output:** $\bar{\mathbf{x}}_t, \bar{\mathbf{\Sigma}}_t$
Extract $\theta$ from $\mathbf{x}_{t-1}$
$|\omega_t| < 10^{-6}$
$\Delta x \leftarrow v_t \cos(\theta)\Delta t$
$\Delta y \leftarrow v_t \sin(\theta)\Delta t$
$\Delta\theta \leftarrow 0$
$\Delta x \leftarrow \frac{v_t}{\omega_t}(\sin(\theta + \omega_t\Delta t) - \sin(\theta))$
$\Delta y \leftarrow \frac{v_t}{\omega_t}(-\cos(\theta + \omega_t\Delta t) + \cos(\theta))$
$\Delta\theta \leftarrow \omega_t\Delta t$
$\bar{\mathbf{x}}_t \leftarrow \mathbf{x}_{t-1} + [\Delta x, \Delta y, \Delta\theta]^T$
Normalize $\bar{\mathbf{x}}_t[2]$ to $[-\pi, \pi]$
Compute Jacobian $\mathbf{G}_t$
$\bar{\mathbf{\Sigma}}_t \leftarrow \mathbf{G}_t\mathbf{\Sigma}_{t-1}\mathbf{G}_t^T + \mathbf{Q}$
$\bar{\mathbf{x}}_t, \bar{\mathbf{\Sigma}}_t$
```

Algorithm 2 EKF-SLAM Update with Data Association

```
**Input:** $\bar{\mathbf{x}}_t, \bar{\mathbf{\Sigma}}_t, \mathbf{Z}_t$
each measurement $\mathbf{z}_i \in \mathbf{Z}_t$
$d_{min} \leftarrow \infty$, $j_{best} \leftarrow -1$
$k = 1$ to $N$
Compute $\hat{\mathbf{z}}_k$, $\mathbf{H}_k$
$\mathbf{S}_k \leftarrow \mathbf{H}_k\bar{\mathbf{\Sigma}}_t\mathbf{H}_k^T + \mathbf{R}$
$d^2 \leftarrow (\mathbf{z}_i - \hat{\mathbf{z}}_k)^T\mathbf{S}_k^{-1}(\mathbf{z}_i - \hat{\mathbf{z}}_k)$
$d^2 < d_{min}$
$d_{min} \leftarrow d^2$, $j_{best} \leftarrow k$
$d_{min} > \gamma$ **or** $N = 0$
Initialize landmark $\mathbf{m}_{N+1}$
Augment state and covariance
$N \leftarrow N + 1$
$\mathbf{K}_t \leftarrow \bar{\mathbf{\Sigma}}_t\mathbf{H}_{best}^T\mathbf{S}_{best}^{-1}$
$\bar{\mathbf{x}}_t \leftarrow \bar{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{z}_i - \hat{\mathbf{z}}_{best})$
$\bar{\mathbf{\Sigma}}_t \leftarrow (\mathbf{I} - \mathbf{K}_t\mathbf{H}_{best})\bar{\mathbf{\Sigma}}_t$
```

Algorithm 3 Q-Learning Control Loop

```
Initialize $Q(s, a) \leftarrow 0$
each episode
Reset robot pose
each time step
Compute $e_{lat}, e_\theta$
$s_t \leftarrow \text{Discretize}(e_{lat}, e_\theta)$
random() $< \epsilon$
$a_t \leftarrow$ random_action()
$a_t \leftarrow \arg\max_a Q(s_t, a)$
Apply control $\mathbf{u}_t$
Observe $s_{t+1}$, reward $r_t$
Update $Q(s_t, a_t)$
```

6 Results

6.1 Trajectory Tracking

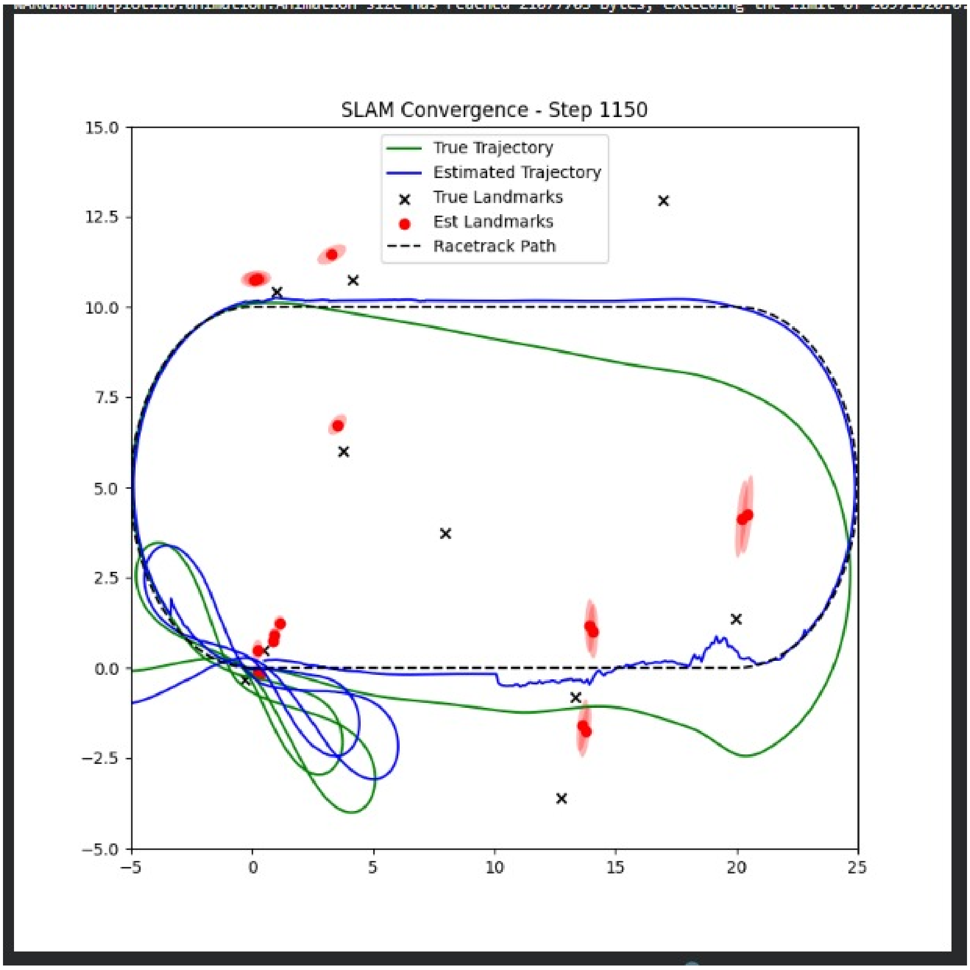

Figure 1: Trajectory tracking performance using Q-Learning.

6.2 Qualitative Performance Analysis

The learned Q-Learning policy significantly reduces lateral and heading errors, particularly in curved sections of the racetrack. Improved trajectory tracking results in more consistent landmark observations, indirectly enhancing SLAM performance. Over multiple laps, landmark estimates converge and pose uncertainty reduces.

7 Future Work and Improvements

Several extensions can further enhance the proposed Q-Learning integrated EKF-SLAM system. The current controller relies on discretized state and action spaces, which limits control resolution. Future work could explore continuous reinforcement learning methods such as Deep Deterministic Policy Gradient (DDPG) or Soft Actor-Critic (SAC) to enable smoother and more precise control actions.

From a SLAM perspective, EKF-SLAM does not scale well with an increasing number of landmarks due to the growth of the covariance matrix. Sparse or submapping-based SLAM approaches could be investigated to improve scalability while maintaining estimation accuracy.

Additionally, the current system assumes static landmarks and fixed noise models. Extending the framework to handle dynamic environments, adaptive noise estimation, or tighter coupling between control and localization uncertainty would improve robustness. Finally, validating the approach on real robotic hardware would provide valuable insight into practical deployment challenges.

8 Conclusion

The integrated Q-Learning EKF-SLAM system demonstrates robust trajectory tracking and accurate localization and mapping. Learning-based control improves navigation stability, while EKF-SLAM ensures consistent state estimation under uncertainty. The results highlight the benefits of combining reinforcement learning with probabilistic SLAM frameworks.

9 References

- Sutton, R. S. and Barto, A. G., *Reinforcement Learning: An Introduction*, MIT Press, 2018.
- Thrun, S., Burgard, W., and Fox, D., *Probabilistic Robotics*, MIT Press, 2005.