

Project 3: Infinite-Horizon Stochastic Optimal Control

Collaboration in the sense of discussion is allowed, however, the assignment is individual and the work you do should be entirely your own. See the collaboration and academic integrity statement here: <https://natanaso.github.io/ece276b>. Books, websites, and papers may be consulted but not copied from. It is absolutely forbidden to copy or even look at anyone else's code. **Please acknowledge in writing people you discuss the problems with.**

Submission

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

1. **Programming assignment:** upload all code you have written for the project and a README file with a clear, concise description of the main file and how to run your code.
2. **Report:** upload your report in pdf format. You are encouraged but not required to use an IEEE conference template¹ for your report.

Problems

In square brackets are the points assigned to each part.

1. This project will consider safe trajectory tracking for a ground differential-drive robot. Consider a robot whose state $\mathbf{x}_t := (\mathbf{p}_t, \theta_t)$ at discrete-time $t \in \mathbb{N}$ consists of its position $\mathbf{p}_t \in \mathbb{R}^2$ and orientation $\theta_t \in [-\pi, \pi)$. The robot is controlled by a velocity input $\mathbf{u}_t := (v_t, \omega_t)$ consisting of linear velocity $v_t \in \mathbb{R}$ and angular velocity (yaw rate) $\omega_t \in \mathbb{R}$. The discrete-time kinematic model of the differential-drive robot obtained from Euler discretization of the continuous-time kinematics with time interval $\tau > 0$ is:

$$\mathbf{x}_{t+1} = \begin{bmatrix} \mathbf{p}_{t+1} \\ \theta_{t+1} \end{bmatrix} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) = \underbrace{\begin{bmatrix} \mathbf{p}_t \\ \theta_t \end{bmatrix}}_{\mathbf{x}_t} + \underbrace{\begin{bmatrix} \tau \cos(\theta_t) & 0 \\ \tau \sin(\theta_t) & 0 \\ 0 & \tau \end{bmatrix}}_{\mathbf{G}(\mathbf{x}_t)} \underbrace{\begin{bmatrix} v_t \\ \omega_t \end{bmatrix}}_{\mathbf{u}_t} + \mathbf{w}_t, \quad t = 0, 1, 2, \dots \quad (1)$$

where $\mathbf{w}_t \in \mathbb{R}^3$ models the motion noise with Gaussian distribution $\mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\sigma})^2)$ with standard deviation $\boldsymbol{\sigma} = [0.04, 0.04, 0.004] \in \mathbb{R}^3$. The motion noise is assumed to be independent across time and of the robot state \mathbf{x}_t . The kinematics model in (1) defines the probability density function $p_f(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{u}_t)$ of \mathbf{x}_{t+1} conditioned on \mathbf{x}_t and \mathbf{u}_t as the density of a Gaussian distribution with mean $\mathbf{x}_t + \mathbf{G}(\mathbf{x}_t)\mathbf{u}_t$ and covariance $\text{diag}(\boldsymbol{\sigma})^2$. The control input \mathbf{u}_t of the vehicle is limited to an allowable set of linear and angular velocities $\mathcal{U} := [0, 1] \times [-1, 1]$.

The objective is to design a control policy for the differential-drive robot in order to track a desired reference position trajectory $\mathbf{r}_t \in \mathbb{R}^2$ and orientation trajectory $\alpha_t \in [-\pi, \pi)$ while avoiding collisions with obstacles in the environment. There are two circular obstacles \mathcal{C}_1 centered at $(-2, -2)$ with radius 0.5 and \mathcal{C}_2 centered at $(1, 2)$ with radius 0.5. Let $\mathcal{F} := [-3, 3]^2 \setminus (\mathcal{C}_1 \cup \mathcal{C}_2)$ denote the free space in the environment. The environment, the obstacles, and the reference trajectory are illustrated in Fig. 1.

It will be convenient to define an error state $\mathbf{e}_t := (\tilde{\mathbf{p}}_t, \tilde{\theta}_t)$, where $\tilde{\mathbf{p}}_t := \mathbf{p}_t - \mathbf{r}_t$ and $\tilde{\theta}_t := \theta_t - \alpha_t$ measure the position and orientation deviation from the reference trajectory, respectively. The equations of motion for the error dynamics are:

$$\mathbf{e}_{t+1} = \begin{bmatrix} \tilde{\mathbf{p}}_{t+1} \\ \tilde{\theta}_{t+1} \end{bmatrix} = g(t, \mathbf{e}_t, \mathbf{u}_t, \mathbf{w}_t) = \underbrace{\begin{bmatrix} \tilde{\mathbf{p}}_t \\ \tilde{\theta}_t \end{bmatrix}}_{\mathbf{e}_t} + \underbrace{\begin{bmatrix} \tau \cos(\tilde{\theta}_t + \alpha_t) & 0 \\ \tau \sin(\tilde{\theta}_t + \alpha_t) & 0 \\ 0 & \tau \end{bmatrix}}_{\tilde{\mathbf{G}}(\mathbf{e}_t)} \underbrace{\begin{bmatrix} v_t \\ \omega_t \end{bmatrix}}_{\mathbf{u}_t} + \begin{bmatrix} \mathbf{r}_t - \mathbf{r}_{t+1} \\ \alpha_t - \alpha_{t+1} \end{bmatrix} + \mathbf{w}_t. \quad (2)$$

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

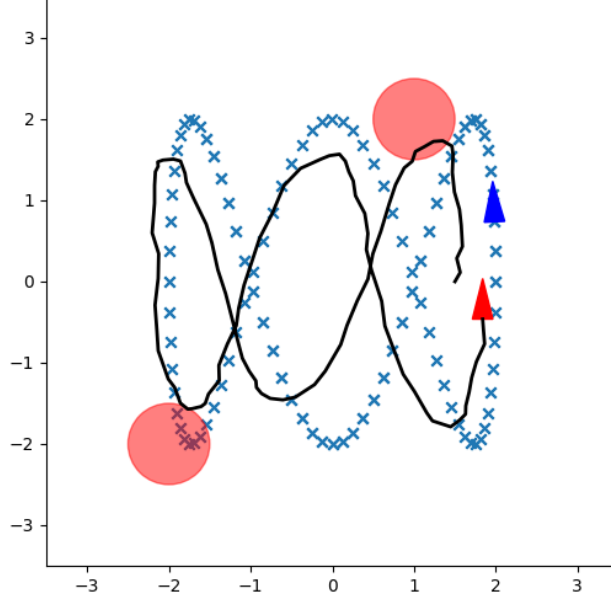


Figure 1: The objective of this project is to design a control policy for a differential-drive robot (red triangle) to track a reference trajectory (cyan crosses), while avoiding two circular obstacles (red). A baseline proportional-error controller is provided but the trajectory (black curve) it generates has a large tracking error and does not avoid obstacles.

We formulate the trajectory tracking with initial time τ and initial tracking error \mathbf{e} as a discounted infinite-horizon stochastic optimal control problem:

$$\begin{aligned}
 V^*(\tau, \mathbf{e}) = \min_{\pi} \mathbb{E} & \left[\sum_{t=\tau}^{\infty} \gamma^t \left(\tilde{\mathbf{p}}_t^\top \mathbf{Q} \tilde{\mathbf{p}}_t + q(1 - \cos(\tilde{\theta}_t))^2 + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t \right) \mid \mathbf{e}_\tau = \mathbf{e} \right] \\
 \text{s.t. } & \mathbf{e}_{t+1} = g(t, \mathbf{e}_t, \mathbf{u}_t, \mathbf{w}_t), \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\sigma})^2), \quad t = \tau, \tau + 1, \dots \\
 & \mathbf{u}_t = \pi(t, \mathbf{e}_t) \in \mathcal{U} \\
 & \tilde{\mathbf{p}}_t + \mathbf{r}_t \in \mathcal{F}
 \end{aligned} \tag{3}$$

where $\mathbf{Q} \in \mathbb{R}^{2 \times 2}$ is a symmetric positive-definite matrix defining the stage cost for deviating from the reference position trajectory \mathbf{r}_t , $q > 0$ is a scalar defining the stage cost for deviating from the reference orientation trajectory α_t , and $\mathbf{R} \in \mathbb{R}^{2 \times 2}$ is a symmetric positive-definite matrix defining the stage cost for using excessive control effort. We will compare two different approaches for solving the problem in (3): (a) receding-horizon certainty equivalent control (CEC) and (b) generalized policy iteration (GPI).

Part 1: CEC is a suboptimal control scheme that applies, at each stage, the control that would be optimal if the noise variables \mathbf{w}_t were fixed at their expected values (zero in our case). The main attractive characteristic of CEC is that it reduces a stochastic optimal control problem to a deterministic optimal control problem, which can be solved more effectively. Receding-horizon CEC, in addition, approximates an infinite-horizon problem by repeatedly solving the following discounted finite-horizon deterministic optimal control problem at each time step:

$$\begin{aligned}
 V^*(\tau, \mathbf{e}) = \min_{\mathbf{u}_\tau, \dots, \mathbf{u}_{\tau+T-1}} & q(\mathbf{e}_{\tau+T}) + \sum_{t=\tau}^{\tau+T-1} \gamma^t \left(\tilde{\mathbf{p}}_t^\top \mathbf{Q} \tilde{\mathbf{p}}_t + q(1 - \cos(\tilde{\theta}_t))^2 + \mathbf{u}_t^\top \mathbf{R} \mathbf{u}_t \right) \\
 \text{s.t. } & \mathbf{e}_{t+1} = g(t, \mathbf{e}_t, \mathbf{u}_t, \mathbf{0}), \quad t = \tau, \dots, \tau + T - 1 \\
 & \mathbf{u}_t \in \mathcal{U} \\
 & \tilde{\mathbf{p}}_t + \mathbf{r}_t \in \mathcal{F}
 \end{aligned} \tag{4}$$

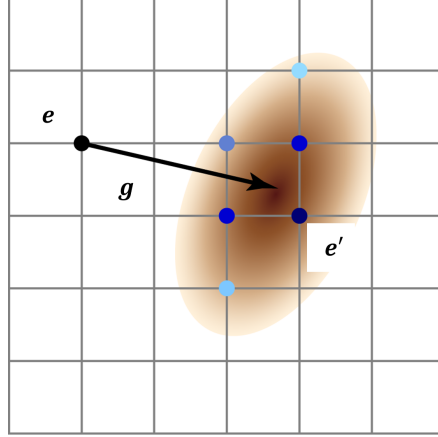


Figure 2: To create transition probabilities in the discretized MDP, for each discrete state \mathbf{e} and control \mathbf{u} , we choose the next grid points \mathbf{e}' around the mean $g(t, \mathbf{e}, \mathbf{u}, 0)$, evaluate the likelihood of $\mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\sigma})^2)$ at the chosen grid points, and normalize so that the outgoing transition probabilities sum to 1.

where $q(\mathbf{e})$ is a suitably chosen terminal cost. The receding-horizon CEC problem is now a non-linear program (NLP) of the form:

$$\begin{aligned} \min_{\mathbf{U}} \quad & c(\mathbf{U}, \mathbf{E}) \\ \text{s.t.} \quad & \mathbf{U}_{lb} \leq \mathbf{U} \leq \mathbf{U}_{ub} \\ & \mathbf{h}_{lb} \leq \mathbf{h}(\mathbf{U}, \mathbf{E}) \leq \mathbf{h}_{ub}, \end{aligned} \tag{5}$$

where $\mathbf{U} := [\mathbf{u}_\tau^\top, \dots, \mathbf{u}_{\tau+T-1}^\top]^\top$ and $\mathbf{E} := [\mathbf{e}_\tau^\top, \dots, \mathbf{e}_{\tau+T}^\top]^\top$. An NLP program can be solved by an NLP solver, such as CasADi. Details about CasADi and examples can be found at <https://web.casadi.org/>. CasADi can be installed easily via:

```
pip install casadi
```

but other options are available on the website. Once a control sequence $\mathbf{u}_\tau, \dots, \mathbf{u}_{\tau+T-1}$ is obtained, CEC applies the first control \mathbf{u}_τ to the system, obtains the new error state $\mathbf{e}_{\tau+1}$ at time $\tau + 1$, and repeats the optimization in (4) online in order to determine the control input $\mathbf{u}_{\tau+1}$. This online re-planning is necessary because the CEC formulation does not take the effect of the motion noise into account.

Part 2: In this part, we will use the GPI algorithm to solve (3) directly but we need to discretize the state and control spaces since they are continuous. We can discretize the state and control spaces into $(n_t, n_x, n_y, n_\theta)$ and (n_v, n_ω) number of grid points, respectively. Be careful that $\tilde{\theta}$ is an angle and wrap-around should be enforced, while $n_t = 100$ since the provided reference trajectory is periodic with period of 100. The position error $\tilde{\mathbf{p}}$ and control input \mathbf{u} may be discretized over the regions $[-3, 3]^2$ and \mathcal{U} . To create transition probabilities in the discretized MDP, for each discrete state \mathbf{e} and each discrete control \mathbf{u} , we can choose the next grid points \mathbf{e}' around the mean $g(t, \mathbf{e}, \mathbf{u}, 0)$, evaluate the likelihood of $\mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\sigma})^2)$ at the chosen grid points, and normalize so that the outgoing transition probabilities sum to 1. See Fig. 2 for an example. You should also be careful to account for the safety constraint $\tilde{\mathbf{p}}_t + \mathbf{r}_t \in \mathcal{F}$, either by not allowing transitions to states that may collide or by introducing an additional stage-cost term to penalize collisions. Generally, denser discretization produces more accurate solutions in the original continuous-space problem but this might lead to a very large state space for the GPI algorithm. If you have time, you may explore ideas such as adaptive discretization of \mathbf{e} with a finer grid closer to the reference trajectory and a coarser grid further away. Alternatively, you may consider representing the Value function with a linear combination of features, $V^*(t, \mathbf{e}) \approx \tilde{V}(t, \mathbf{e}; \boldsymbol{\theta}) := \boldsymbol{\theta}^\top \boldsymbol{\phi}(t, \mathbf{e})$, as discussed in Lecture 14². Exploring these ideas is optional. It is sufficient to implement a discrete-space Value Iteration or Policy Iteration algorithm for this part.

²https://natanaso.github.io/ece276b/ref/ECE276B_14_ValueFunctionApproximation.pdf

Write a project report describing your approach to the trajectory tracking problem. Include the following sections:

- (a) [5 pts] **Introduction:** present a brief overview of the trajectory tracking problem and the techniques used in this project.
- (b) [5 pts] **Problem Statement:** state the problem you are trying to solve in mathematical terms. The project description already defines the variables, constraints, and objective in (1) - (3), and it is acceptable to restate this formulation.
- (c) [30 pts] **Technical Approach:** describe your approach in two sections: (a) dedicated to the receding-horizon CEC control and (b) dedicated to the GPI algorithm. Carefully describe the NLP formulation in part (a), including the optimization variables, constraints, and objective function. Describe the state discretization, handling collision avoidance, and how you managed to scale the GPI algorithm to a large state space in part (b).
- (d) [30 pts] **Results:** show qualitative and quantitative results for both the CEC and GPI trajectory tracking techniques and compare their performance in terms of computational complexity, collision avoidance, and tracking error. You may discuss the effect of ignoring the noise in the CEC formulation and of the discretization in the GPI formulation. Discuss also the performance and any interesting details you observe about the two techniques.
- (e) [30 pts] **Code:** upload your code to Gradescope. It will be evaluated based on correctness and efficiency. Correctness refers to whether your code implements the algorithms you present in your report and whether the algorithms you present actually solve the trajectory tracking problem. Efficiency refers to your code being able to execute and produce control inputs for the vehicle in reasonable time. The results presented in the report should be consistent with the output of your code. The code should have a clear structure and comments. **Copying, rephrasing, or even looking at anyone else's code is strictly forbidden. Writing your own code is the best way to avoid plagiarism discussions or consequences.**