# Control for Robotics

*From Optimal Control to Reinforcement Learning*

`Short Solutions for Exercises`

Angela P. Schoellig and Siqi Zhou

*with contributions from*

Lukas Brunke
Martin Schuck
Adam W. Hall
Haocheng Zhao
Ralf Römer
Oliver Hausdörfer

Release Version: 2026 Summer

# Chapter 1

# Introduction to Optimal Control

**Exercise 1.1** (Short Questions)**.**

a)

Choice (B) is correct.

b)

Choices (A), (B), and (D) are correct.

c)

Choices (A) and (D) are correct.

d)

Choices (B) and (C) are correct.

e)

Choice (A) is correct.

f)

Choices (A), (B), and (C) are correct.

g)

Choices (A) and (C) are correct.

**Exercise 1.2** (Shortest Path Problem)**.**

The shortest path from node A is $A \rightarrow B \rightarrow C$, with a total cost of 5. The shortest path from node A is $B \rightarrow C$, with a total cost of 3.

**Exercise 1.3** (Shortest Path Problem)**.**

The shortest path from the start node to the terminal node is $S \rightarrow B \rightarrow D \rightarrow E \rightarrow T$, with a total cost of 10.

**Exercise 1.4** (Mobile Robot Goal-Reaching Problem)**.**

a)

The cost function penalizes any nonzero states and inputs along the path according to the gains $q$ and $r$. If $q < r$, inputs will be penalized more than states, and smaller inputs will be favoured along the trajectory. If $r < q$, then larger inputs will be desired to drive the system to $x_2 = 0$. The terminal cost at $x_2$ helps to drive the system to $x_2 = 0$.

b)

The optimal policy is $u_1 = -\frac{1}{2}x_1, u_0 = -\frac{3}{4}x_0$, with $J(x_0) = J_0 = \frac{13}{4}$.

c)

The optimal policy is $u_1 = \frac{-x_1-1}{2}, u_0 = \frac{-6x_0-7}{8}$, with $J(x_0) = J_0 = \frac{111}{16}$.

d)

Comparing the two cases starting from $x_0 = -1$, without the disturbance, $J_0 = 13/4$, and $u_0 = 6/8$ and with the disturbance $J_0 = 111/16$, and $u_0 = -1/8$. We see that with the disturbance, the input to the system, which is supposed to drive the system to equilibrium, is actually in the opposite direction of what we would expect and the opposite to what we see in the case with no disturbance! This is because the expectation of the terrain disturbances is actually in the positive direction and mainly contributes to pushing the robot forward. To compensate for this, the first controller action is in the downward direction. This would be like trying to position the robot exactly at some point on a hill slopped downwards in the direction of travel.

**Exercise 1.5** (Constrained Mobile Robot Goal-Reaching Problem)**.**

a)

The optimal policy is $u_1^*(x_1) = 1$ if $x_1 < 0$, $-1$ if $x_1 > 0$, and either $\pm 1$ if $x_1 = 0$. For the initial stage, $u_0^* = 1$, yielding a minimum cost of $J_0 = 11/2$.

b)

The optimal policy is $u_1^*(x_1) = -0.5x_1$ for $x_1 \leq -1$, and $u_1^*(x_1) = -0.5 - x_1$ for $-1 < x_1 \leq -0.5$. At the first stage, $u_0^* = 0.5$, which places the robot at the boundary $x_1 = -0.5$, yielding a total cost of $J_0 = 29/8$.

# Chapter 2

# Linear Quadratic Optimal Control

**Exercise 2.1** (Short Questions)**.**

a)

> Choices (A), (B), and (C) are correct.

b)

> Choice (C) is correct.

c)

> Choices (A), (B), and (D) are correct.

d)

> Choices (B) and (C) are correct.

e)

> Choices (B), (C), and (D) are correct.

f)

Choices (A) and (D) are correct.

**Exercise 2.2** (HJB Derivation)**.**

In our derivation below, we first discretize the continuous-time equations into $N$ steps, then apply the dynamic programming algorithm, and finally take the limit of the discretization time $\delta t = \frac{T}{N}$ to zero. We can generally relate the continuous-time and discrete-time variables as $t = k\delta t$.
Note that, here, we motivate the HJB Equation from dynamic programming, but there exist other approaches to arrive at the same result.

- **Discretization:** We first define discrete-time signals $x_k = x(k\delta t)$ for all $k = \{0, 1, ..., N\}$ and $u_k = u(k\delta t)$ for all $k = \{0, 1, ..., N-1\}$. We then write the discretized dynamics and cost based on their first-order approximations
  - Dynamics
  $$\frac{x_{k+1} - x_k}{\delta t} = f(k\delta t, x_k, u_k) \quad \Leftrightarrow \quad x_{k+1} = x_k + \delta t \cdot f(k\delta t, x_k, u_k) \tag{2.1}$$
  - Cost
  $$\ell_T(x(N\delta t)) + \delta t \sum_{k=0}^{N-1} \ell(k\delta t, x_k, u_k) \tag{2.2}$$
- **Apply Dynamic Programming:** Denote $J^*(t, x)$ and $\widetilde{J}^*(t, x)$ as the optimal cost-to-go at time $t$ and state $x$ for the original continuous problem and the discretized problem, respectively.
  - Initialization:
  $$\widetilde{J}^*(N\delta t, x_N) = \ell_T(x(N\delta t)) \tag{2.3}$$

- Recursion:

$$\widetilde{J}^*(k\delta t, x_k) = \min_{u_k \in \mathcal{U}} \begin{pmatrix} \ell(k\delta t, x_k, u_k)\,\delta t \\ +\,\widetilde{J}^*\big((k+1)\,\delta t, x_k \\ +\,f(k\delta t, x_k, u_k)\,\delta t\big) \end{pmatrix} \tag{2.4}$$

- We can approximate the cost-to-go by computing the Taylor expansion of $\widetilde{J}^*\big((k+1)\delta t,\ x_k + f(k\delta t, x_k, u_k)\ \delta t\big)$ around $(k\delta t,\ x_k)$:

$$\begin{aligned} &\widetilde{J}^*((k+1)\,\delta t,\ x_k + f(k\delta t, x_k, u_k)\,\delta t) \\ =&\widetilde{J}^*(k\delta t, x_k) + \nabla_t\,\widetilde{J}^*(k\delta t, x_k)\delta t \\ &+ \nabla_x\,\widetilde{J}^*(k\delta t, x_k) f(k\delta t, x_k, u_k)\delta t + O(\delta t), \end{aligned} \tag{2.5}$$

  where $O(\delta t)$ denotes the higher order terms that satisfies $\lim_{\delta \to 0} \frac{O(\delta)}{\delta} = 0$.

- By substituting the approximation back in (2.4), dividing the expression by $\delta t$, and making the substitution $t = k\delta t$, we obtain

$$0 = \min_{u \in \mathcal{U}} \frac{1}{\delta t} \begin{pmatrix} \ell(t, x, u)\,\delta t + \nabla_t\,\widetilde{J}^*(t, x)\,\delta t \\ +\,\nabla_x\,\widetilde{J}^*(t, x)\,f(t, x, u)\,\delta t + O(\delta t) \end{pmatrix} \tag{2.6}$$

$$= \min_{u \in \mathcal{U}} \begin{pmatrix} \ell(t, x, u) + \nabla_t\,\widetilde{J}^*(t, x) \\ +\,\nabla_x{}^T \widetilde{J}^*(t, x)\,f(t, x, u) + \dfrac{O(\delta t)}{\delta t} \end{pmatrix}. \tag{2.7}$$

  Note that the explicit time dependence of $x$ and $u$ is dropped for clarity.

- **Take Limit to Zero:** Now let $\delta t \to 0$. Assume $\widetilde{J}^* \to J^*$. We then arrive at the HJB equation:

$$\begin{aligned} 0 = \min_{u \in \mathcal{U}} \Big(\ell(t, x, u) + \nabla_t\,J^*(t, x) + \nabla_x\,J^*(t, x)\,f(t, x, u)\Big), \\ \forall x \in \mathcal{X},\ t \in [0, T], \end{aligned} \tag{2.8a}$$

$$\text{subject to} \quad J^*(T, x) = \ell_T(x), \quad \forall x \in \mathcal{X}. \tag{2.8b}$$

  The minimizing $u = \mu(t, x)$ is an optimal policy.

**Exercise 2.3** (Finite-Horizon CARE Derivation).

By substituting the linear dynamics and quadratic cost function into (**??**)–(**??**), and moving $\nabla_t J(t,x)$ in (**??**) to the left-hand, we obtain

$$-\nabla_t J(t,x) = \min_{u \in \mathcal{U}} \left( x^T Q x + u^T R u + \nabla_x^T J(t,x)\big(Ax + Bu\big) \right), \quad \forall x, t, \tag{2.9a}$$

$$J(T,x) = x^T Q_T x, \quad \forall\, x. \tag{2.9b}$$

Note that the time dependencies of $A$, $B$, $Q$, $R$, $x$, and $u$ are omitted in the derivation. Given the quadratic form of the cost function, we assume that the optimal cost-to-go has a quadratic form

$$J(t,x) = x^T S(t) x, \tag{2.10}$$

where $S$ is a symmetric time-varying matrix. The gradient of $J(t,x)$ with respect to $t$ and $x$ are $\nabla_t J(t,x) = x^T \dot{S}(t) x$ and $\nabla_x J(t,x) = 2S(t)x$, respectively. With the quadratic parametrization of $J(t,x)$, we can rewrite the terminal value problem as

$$-x^T \dot{S}(t) x = \min_{u \in \mathcal{U}} \left( x^T \big(Q + 2S(t)A\big) x + u^T R u + 2u^T B(t)^T S(t) x \right), \quad \forall\, x,\ t, \tag{2.11a}$$

$$J(T,x) = x^T Q_T x, \quad \forall\, x. \tag{2.11b}$$

Note that the quadratic term $x^T 2S(t)Ax$ can be equally write as $x^T\big(S(t)A + A^T S(t)\big)x$ by noting that the term is just a scalar and the transpose equals to itself. This is a common trick to ensure that the matrix in the quadratic term is symmetric. We can find the minimizing $u$ by differentiating the right-hand side of (2.11a) with respect to $u$ and setting the derivative to zero:

$$2Ru + 2B^T S(t) x = 0, \tag{2.12}$$

which gives the optimal policy

$$u(t) = \underbrace{-R^{-1}B^T S(t)}_{=K(t)} x(t) = K(t)x(t) = \mu^*\big(t, x(t)\big). \tag{2.13}$$

Then, (2.11a) becomes

$$-x^T \dot{S}(t)x = x^T \left(Q + S(t)A + A^T S(t) - S(t)BR^{-1}B^T S(t)\right) x, \quad \forall\, x,\, t. \tag{2.14}$$

Since (2.14) holds for all $x$, we have

$$\dot{S} = -Q - SA - A^T S + SBR^{-1}B^T S, \quad \forall\, t, \tag{2.15a}$$
$$S(T) = Q(T). \tag{2.15b}$$

The differential equation in (2.15) is the CARE for the continuous-time linear quadratic optimal control problem.

**Exercise 2.4** (Finite-Horizon LQR [**?** , Example 13.2]).

a)

$$K_k = -\frac{S_{k+1}}{r + S_{k+1}}, \quad S_k = \frac{r + (r+1)S_{k+1}}{r + S_{k+1}}, \quad S_N = q$$

b)

| $k$ | 4 | 3 | 2 | 1 | 0 |
|---|---|---|---|---|---|
| $S_k$ | $q$ | $\frac{1+2q}{1+q}$ | $\frac{3+5q}{2+3q}$ | $\frac{8+13q}{5+8q}$ | $\frac{21+34q}{13+21q}$ |
| $K_k$ | | $-\frac{q}{1+q}$ | $-\frac{1+2q}{2+3q}$ | $-\frac{3+5q}{5+8q}$ | $-\frac{8+13q}{13+21q}$ |

c)

$$\begin{aligned} q = 0: &\quad x_1 = 0.769, \quad x_2 = 0.308, \quad x_3 = 0.154, \quad x_4 = 0.154. \\ q \to \infty: &\quad x_1 = 0.762, \quad x_2 = 0.286, \quad x_3 = 0.095, \quad x_4 = 0. \end{aligned}$$

d)

$$\begin{aligned} q = 0: &\quad J^* = 2 \cdot \frac{21}{13} = 3.231. \\ q \to \infty: &\quad J^* = \frac{1}{2} x_0^2 S_0 = 2 \cdot \frac{34}{21} = 3.238. \end{aligned}$$

**Exercise 2.5** (Relationship between LQR and PD Control)**.**

a)

With the state defined as $x(t) = [p(t), v(t)]^T$, we have

$$u(t) = \begin{bmatrix} -k_p & -k_v \end{bmatrix} \begin{bmatrix} p(t) \\ v(t) \end{bmatrix} = \underbrace{\begin{bmatrix} -4 & -5 \end{bmatrix}}_{K} x(t).$$

b)

$$S = \begin{bmatrix} s_{11} & 4 \\ 4 & 5 \end{bmatrix}$$

c)

$$\text{for } R=1, Q=\begin{bmatrix}16 & 0\\ 0 & 17\end{bmatrix} \qquad \text{for } R=2, Q=\begin{bmatrix}32 & 0\\ 0 & 34\end{bmatrix}$$

**Exercise 2.6** (Mobile Robot Output-Based Regulation)**.**

a)

$$x_{k+1} = \frac{1}{2}x_k - \frac{1}{2}v_k + w_k$$

b)

Consider the expectation of the closed-loop system

$$\begin{aligned}
\mathbb{E}_{v_k,w_k}[x_{k+1}] &= \mathbb{E}_{v_k,w_k}[\frac{1}{2}x_k - \frac{1}{2}v_k + w_k] && (2.16)\\
&= \mathbb{E}_{v_k,w_k}[\frac{1}{2}x_k] - \mathbb{E}_{v_k,w_k}[\frac{1}{2}v_k] + \mathbb{E}_{v_k,w_k}[w_k] && (2.17)\\
&= \mathbb{E}_{v_k,w_k}[\frac{1}{2}x_k]. && (2.18)
\end{aligned}$$

There are two ways to see this system is stable. First, from discrete-systems analysis, we know that the eigenvalue of this system is $1/2$, which lies within the unit circle, and thus this system is stable. Second, one can also deduce this simply from the fact that this recursion expanded out is

$$\begin{aligned}
x_1 &= \frac{1}{2}x_0 && (2.19)\\
x_2 &= \frac{1}{2}x_1 = \frac{1}{2}(\frac{1}{2}x_0) && (2.20)\\
&\vdots && (2.21)\\
x_N &= \left(\frac{1}{2}\right)^N x_0, && (2.22)
\end{aligned}$$

which converges to 0 as $N \to \infty$, meaning the system is stable.

c)

$$J = \frac{3}{2}$$

d)

$$\alpha = -\frac{1}{2}$$

**Exercise 2.7** (Infinite-Horizon LQR [**?** , Example 13.3])**.**

a)

$$S = \frac{1}{2} + \sqrt{\frac{1}{4} + r}$$

b)

$$K = -\frac{S}{r+S} = -\frac{1+\sqrt{1+4r}}{2r+1+\sqrt{1+4r}}$$

c)

$$x_{k+1} = x_k + u_k = (1+K)x_k = \frac{2r}{2r+1+\sqrt{1+4r}}x_k$$

Since $\frac{2r}{2r+1+\sqrt{1+4r}} < 1$, the equilibrium $x_{\text{eq}} = 0$ is asymptotically stable.

d)

- For $r = 0$: $K = -1$, i.e., Dead-Beat Control: Reaches equilibrium $x_{\text{eq}} = 0$ in one time step
- For $r \to \infty$: $K = 0$ closed-loop behaviour is equal to open-loop behaviour, only marginally stable

**Exercise 2.8** (Infinite-Horizon LQR)**.**

a)

$$S = \begin{bmatrix} 3.7913 & 0 \\ 0 & 2.7913 \end{bmatrix}$$

b)

Given that $(A, B)$ is stabilizable and $(A, \sqrt{Q})$ is detectable, we can find a unique solution to the DARE and the optimal policy is stabilizing. This implies that the eigenvalues of $(A + BK)$ lie strictly within the unit circle. The only possible eigenvalues are thus $\lambda_1 = 0.3$ and $\lambda_3 = 0.25 + 0.25i$.

c)

$$\begin{aligned}
K &= -(B^T SB + r)^{-1} B^T SA \\
&= -\left( \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} c+1 & b \\ b & c \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \right)^{-1} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} c+1 & b \\ b & c \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\
&= -(c+2)^{-1} \begin{bmatrix} -b & c+1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & \frac{3-\sqrt{21}}{2} \end{bmatrix}
\end{aligned}$$

d)

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k = (A+BK)x_k \\ &= \left( \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & \frac{3-\sqrt{21}}{2} \end{bmatrix} \right) x \\ &= \begin{bmatrix} 0 & \frac{5-\sqrt{21}}{2} \\ -1 & 0 \end{bmatrix} x \end{aligned}$$

$$\operatorname{eig}(A+BK) = \pm\sqrt{-\frac{5-\sqrt{21}}{2}} = \pm 0.4569i \qquad \Rightarrow \text{asymptotically stable}$$

**Exercise 2.9** (LQR for Mobile Robot Steering Problem)**.**

a)

The nonlinear dynamics of the system is

$$\dot{x}(t) = f\big(x(t), u(t)\big) = \begin{bmatrix} \sin\theta(t) \\ u(t) \end{bmatrix}. \tag{2.23}$$

The linearized dynamics is

$$\delta\dot{x} = A\tilde{x} + B\tilde{u} \tag{2.24}$$

with

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{2.25}$$

b)

Note that the matrix $A$ has an upper triangular form. The eigenvalues of the matrix are the diagonal elements $\lambda = \{0, 0\}$. To check the stabilizability of the system, we need to check whether $[A - \lambda_i I,\ B]$ has full rank for all unstable eigenvalues. For our problem, we have

$$[A - \lambda_i I,\ B] = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{for} \quad \lambda_i = 0, \tag{2.26}$$

which has full rank (i.e., a rank of 2). The system is thus stabilizable. Intuitively, stabilizability implies that we have the control authority to stabilize all unstable subspaces or modes of the unforced system.

c)

To ensure that the optimal control problem is solvable and has a unique solution, we require $(A, \sqrt{Q})$ to be detectable. The linearized system with zero input has two poles at 0. The matrix $Q$ needs to be a non-zero symmetric positive semidefinite matrix. We additionally require $r$ to be positive for the optimal control input to be bounded.

d)

$$u^*(t) = \begin{bmatrix} -1 & -2 \end{bmatrix} \tilde{x}(t)$$

# Chapter 3

# Optimization Fundamentals

**Exercise 3.1** (Short Questions)**.**

a)

Choice (C) is correct.

b)

Choice (C) is correct.

c)

Options (A) and (C) are correct.

d)

Choices (A), (B) and (C) are correct.

e)

Choice (D) is correct.

f)

Choices (A) and (E) are correct.

g)

Choices (A), (B), and (C) are correct.

h)

Choices (B) and (D) are correct.

i)

Choices (A) and (B) are correct.

j)

Choices (A) and (B) are correct.

**Exercise 3.2** (Constrained Optimization)**.**

a)

$$x_1^* = \frac{1}{3}, \quad x_2^* = \frac{2}{3}, \quad \lambda_1^* = \frac{4}{3} \text{ with } f(x_1^*, x_2^*) = \frac{2}{3}.$$

b)

$$x_1^* = \frac{5}{3}, \quad x_2^* = \frac{1}{3} \text{ with } f(x_1^*, x_2^*) = 1.$$

c)

$$x_1^* = -0.5, \quad x_2^* = 4.5 \text{ with } f(x_1^*, x_2^*) = 4.75.$$

**Exercise 3.3** (Constrained Open-Loop Optimal Control).

a)

$$u_0^* = 1, \quad u_1^* = 0.5 \text{ with cost } J(u_0^*, u_1^*) = 0.$$

1 step to reach reference.

b)

$$u_0^* = 0.75, \quad u_1^* = 0.625 \text{ with cost } J(u_0^*, u_1^*) = \frac{1}{16}.$$

2 steps to reach reference.

**Exercise 3.4** (Constraint Softening [? ]).

a)

$$x^* = 1, \quad \mu^* = 4 \text{ with } f(x^*) = 6.$$

b)

$$x^* = 2, \quad \epsilon^* = 1 \text{ with } f(x^*) + l(\epsilon^*) = 4.$$

c)

$$x^* = 1, \quad \epsilon^* = 0 \text{ with } f(x^*) + l(\epsilon^*) = 6.$$

d)

$$x^* = 1, \quad \epsilon^* = 0 \text{ with } f(x^*) + l(\epsilon^*) = 6.$$