Optimal Control and Estimation

Heng Yang

2023-06-11

Contents

Preface

This is the textbook for Harvard ES/AM 158: Introduction to Optimal Control and Estimation. Information about the offerings of the class is listed below.

2023 Fall

Time: Mon/Wed 2:15 - 3:30pm

Location: Science and Engineering Complex, Room TBD

Instructor: Heng Yang

Teaching Fellow: Weiyu Li

Syllabus

Acknowledgment

Chapter 1

The Optimal Control Formulation

1.1 The Basic Problem

Consider a discrete-time dynamical system

$$x_{k+1} = f_k(x_k, u_k, w_k), \quad k = 0, 1, \dots, N-1 \tag{1.1}$$

where

- $x_k \in \mathbb{X} \subseteq \mathbb{R}^n$ is the *state* of the system,
- $u_k \in \mathbb{U} \subseteq \mathbb{R}^m$ is the *control* we wish to design,
- $w_k \in \mathbb{W} \subseteq \mathbb{R}^p$ a random *disturbance* or noise (e.g., due to unmodelled dynamics) which is described by a probability distribution $P_k(\cdot \mid x_k, u_k)$ that may depend on x_k and u_k but not on prior disturbances $w_0, \dots, w_{k-1}$,
- k indexes the discrete time,
- N denotes the horizon,
- f_k models the transition function of the system (typically $f_k \equiv f$ is time-invariant, especially for robotics systems; we use f_k here to keep full generality).

Remark (Deterministic v.s. Stochastic). When $w_k \equiv 0$ for all k, we say the system (1.1) is *deterministic*; otherwise we say the system is *stochastic*. In the following we will deal with the stochastic case, but most of the methodology should carry over to the deterministic setup.

We consider the class of *controllers* (also called *policies*) that consist of a sequence of functions

$$\pi = \{\mu_0, \dots, \mu_{N-1}\},$$

where $\mu_k(x_k) \in \mathbb{U}$ for all x_k, i.e., μ_k is a *feedback* controller that maps the state to an admissible control. Given an initial state x_0 and an admissible policy π, the state *trajectory* of the system is a sequence of random variables that evolve according to

$$x_{k+1} = f_k(x_k, \mu_k(x_k), w_k), \quad k = 0, \dots, N-1 \tag{1.2}$$

where the randomness comes from the disturbance w_k.

We assume the state-control trajectory $\{u_k\}_{k=0}^{N-1}$ and $\{x_k\}_{k=0}^{N}$ induce an *additive cost*

$$g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k) \tag{1.3}$$

where $g_k, k = 0, \dots, N$ are some user-designed functions.

With (1.2) and (1.3), for any admissible policy π, we denote its induced *expected cost* with initial state x_0 as

$$J_\pi(x_0) = \mathbb{E}\left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k)) \right\}, \tag{1.4}$$

where the expectation is taken over the randomness of w_k.

Definition 1.1 (Discrete-time, Finite-horizon Optimal Control)**.** Find the best admissible controller that minimizes the expected cost in (1.4)

$$\pi^\star \in \arg\min_{\pi \in \Pi} J_\pi(x_0), \tag{1.5}$$

where Π is the set of all admissible controllers. The cost attained by the optimal controller, i.e., $J^\star = J_{\pi^\star}(x_0)$ is called the optimal *cost-to-go*, or the optimal *value function.*

1.2 Open-Loop v.s. Closed-Loop

An important feature of the basic problem in Definition 1.1

Chapter 2

Stability Analysis

Lemma 2.1 (Barbalat's Lemma)**.** *Let $f(t)$ be differentiable, if*

- $\lim_{t\to\infty} f(t)$ *is finite, and*
- $\dot{f}(t)$ *is uniformly continuous,*[1]

then

$$\lim_{t\to\infty} \dot{f}(t) = 0.$$

Theorem 2.1 (Barbalat's Stability Certificate)**.** *If a scalar function $V(x,t)$ satisfies*

- $V(x,t)$ *is lower bounded,*
- $\dot{V}(x,t)$ *is negative semidefinite*
- $\dot{V}(x,t)$ *is uniformly continuous*

then $\dot{V}(x,t) \to 0$ as $t \to \infty$.

Proof. $V(x,t)$ is lower bounded and $\dot{V}$ is negative semidefinite implies the limit of V as $t \to \infty$ is finite (note that $V(x,t) \le V(x(0),0)$). Then the theorem clearly follows from Barbalat's Lemma 2.1. $\square$

[1]A sufficient condition for this to hold is that $\ddot{f}$ exists and is bounded.

Chapter 3

Adaptive Control

Lemma 3.1 (Basic Lemma)**.** *Let two signals $e(t)$ and $\phi(t)$ be related by*

$$e(t) = H(p)[k\phi(t)^T v(t)] \tag{3.1}$$

where $e(t)$ a scalar output signal, $H(p)$ a strictly positive real (SPR) transfer function, k an unknown real number with known sign, $\phi(t) \in \mathbb{R}^m$ a control signal, and $v(t) \in \mathbb{R}^m$ a measurable input signal.

If the control signal $\phi(t)$ satisfies

$$\dot{\phi}(t) = -\operatorname{sgn}(k)\gamma e(t)v(t) \tag{3.2}$$

with $\gamma > 0$ a positive constant, then $e(t)$ and $\phi(t)$ are globally bounded. Moreover, if $v(t)$ is bounded, then

$$\lim_{t\to\infty} e(t) = 0.$$

Proof. Let the state-space representation of (3.1) be

$$\dot{x} = Ax + b[k\phi^T v], \quad e = c^T x. \tag{3.3}$$

Since $H(p)$ is SPR, it follows from the Kalman-Yakubovich Lemma A.1 that there exist $P, Q \succ 0$ such that

$$A^T P + PA = -Q, \quad Pb = c.$$

Let

$$V(x, \phi) = x^T P x + \frac{|k|}{\gamma}\phi^T \phi,$$

clearly V is positive definite (i.e., $V(0,0) = 0$, and $V(x, \phi) > 0$ for all $x \neq 0, \phi \neq 0$). The time derivative of V along the trajectory defined by (3.3) with ϕ chosen as in (3.2) is

$$\dot{V} = \frac{\partial V}{\partial x}\dot{x} + \frac{\partial V}{\partial \phi}\dot{\phi} \tag{3.4}$$

$$= x^T(PA + A^T P)x + 2x^T Pb(k\phi^T v) + \frac{2|k|}{\gamma}\phi^T(-\text{sgn}(k)\gamma ev) \tag{3.5}$$

$$= -x^T Qx + 2(x^T c)(k\phi^T v) - 2\phi^T(ev) \tag{3.6}$$

$$= -x^T Qx \leq 0. \tag{3.7}$$

As a result, we know x and ϕ must be bounded ($V(x(t), \phi(t)) \leq V(x(0), \phi(0))$ is bounded). Since $e = c^T x$, we know e must be bounded as well.

If the input signal v is also bounded, then $\dot{x}$ is bounded as seen from (3.3). Because $\ddot{V} = -2x^T Q\dot{x}$ is now bounded, we know $\dot{V}$ is uniformly continuous. Therefore, by Barbalat's stability certificate (Theorem 2.1), we know $\dot{V}$ tends to zero as t tends to infinity, which implies $\lim_{t\to\infty} x(t) = 0$ and hence $\lim_{t\to\infty} e(t) = 0$. □

3.1 First-Order Systems

3.1.1 Linear Systems

Consider the first-order single-input single-output (SISO) system

$$\dot{x} = -ax + bu \tag{3.8}$$

where a and b are unknown groundtruth parameters. However, we do assume that the sign of b is known. What if the sign of b is unknown too?

Let $r(t)$ be a reference trajectory, e.g., a step function or a sinusoidal function, and $x_d(t)$ be a desired system trajectory that tracks the reference

$$\dot{x}_d = -a_d x_d + b_d r(t), \tag{3.9}$$

where $a_d, b_d > 0$ are user-defined constants. Note that the transfer function from r to x_d is

$$x_d = \frac{b_d}{p + a_d} r$$

and the system is stable. Review basics of transfer function.

The goal of adaptive control is to design a control law and an adaptation law such that the tracking error of the system $x(t) - x_d(t)$ converges to zero.

Control law. We design the control law as

$$u = \hat{a}_r(t)r + \hat{a}_x(t)x \tag{3.10}$$

where $\hat{a}_r(t)$ and $\hat{a}_x(t)$ are time-varying feedback gains that we wish to adapt. The closed-loop dynamics of system (3.8) with the controller (3.10) is

$$\dot{x} = -ax + b(\hat{a}_r r + \hat{a}_x x) = -(a - b\hat{a}_x)x + b\hat{a}_r r.$$

With the equation above, the reason for choosing the control law (3.10) is clear: if the system parameters (a, b) were known, then choosing

$$a_r^\star = \frac{b_d}{b}, \quad a_x^\star = \frac{a - a_d}{b} \tag{3.11}$$

leads to the closed-loop dynamics $\dot{x} = -a_d x + b_d r$ that is exactly what we want in (3.9).

However, in adaptive control, since the true parameters (a, b) are not revealed to the control designer, an adaptation law is needed to dynamically adjust the gains $\hat{a}_r$ and $\hat{a}_x$ based on the tracking error $x(t) - x_d(t)$.

Adaptation law. Let $e(t) = x(t) - x_d(t)$ be the tracking error, and we develop its time derivative

$$\begin{aligned}
\dot{e} &= \dot{x} - \dot{x}_d && (3.12)\\
&= -a_d(x - x_d) + (a_d - a + b\hat{a}_x)x + (b\hat{a}_r - b_d)r && (3.13)\\
&= -a_d e + b\underbrace{(\hat{a}_x - \hat{a}_x^\star)}_{=:\tilde{a}_x} x + b\underbrace{(\hat{a}_r - \hat{a}_r^\star)}_{=:\tilde{a}_r} r && (3.14)\\
&= -a_d e + b(\tilde{a}_x x + \tilde{a}_r r) && (3.15)
\end{aligned}$$

where $\tilde{a}_x$ and $\tilde{a}_r$ are the gain errors w.r.t. the optimal gains in (3.11) if the true parameters were known. The error dynamics (3.15) is equivalent to the following transfer function

$$e = \frac{1}{p + a_d} b(\tilde{a}_x x + \tilde{a}_r r) = \frac{1}{p + a_d}\left(b\begin{bmatrix}\tilde{a}_x\\ \tilde{a}_r\end{bmatrix}^T \begin{bmatrix}x\\ r\end{bmatrix}\right), \tag{3.16}$$

which is in the form of (3.1). Therefore, we choose the adaptation law

$$\begin{bmatrix}\dot{\tilde{a}}_x\\ \dot{\tilde{a}}_r\end{bmatrix} = -\mathrm{sgn}(b)\gamma e \begin{bmatrix}x\\ r\end{bmatrix}. \tag{3.17}$$

Tracking convergence. With the control law (3.10) and the adaptation law (3.17), we can prove that the tracking error converges to zero, using Lemma 3.1. With $\tilde{a} = [\tilde{a}_x, \tilde{a}_r]^T$, let

$$V(e, \tilde{a}) = e^2 + \frac{|b|}{\gamma}\tilde{a}^T\tilde{a}$$

be a positive definite Lyapunov function candidate with time derivative

$$\dot{V} = -2a_d e^2 \leq 0.$$

Clearly, e and $\tilde{a}$ are both bounded. Assuming the reference trajectory r is bounded, we know x_d is bounded (due to (3.9)) and hence x is bounded (due to $e = x - x_d$ being bounded). Consequently, from the error dynamics (3.15) we know $\dot{e}$ is bounded, which implies $\ddot{V} = -4a_d e\dot{e}$ is bounded and $\dot{V}$ is uniformly continuous. By Barbalat's stability certificate 2.1, we conlude $e(t) \to 0$ as $t \to \infty$.

It is always better to combine mathematical analysis with intuitive understanding. Can you explain intuitively why the adaptation law (3.17) makes sense? (Hint: think about how the control should react to a negative/positive tracking error.)

Parameter convergence. We have shown the control law (3.10) and the adaptation law (3.17) guarantee to track the reference trajectory. However, is it guaranteed that the gains of the controller (3.10) also converge to the optimal gains in (3.11)?

We will now show that the answer is indefinite and it depends on the reference trajectory $r(t)$. Because the tracking error e converges to zero, and e is the output of a stable filter (3.16), we know the input $b(\tilde{a}_x x + \tilde{a}_r r)$ must also converge to zero. On the other hand, the adaptation law (3.17) shows that both $\dot{\tilde{a}}_x$ and $\dot{\tilde{a}}_r$ converge to zero (due to e converging to zero and x, r being bounded). As a result, we know $\tilde{a} = [\tilde{a}_x, \tilde{a}_r]^T$ converges to a constant that satisfies

$$v^T \tilde{a} = 0, \quad v = \begin{bmatrix} x \\ r \end{bmatrix}, \tag{3.18}$$

which is a single linear equation of $\tilde{a}$ with time-varying coeffients.

- **Constant reference: no guaranteed convergence**. Suppose $r(t) \equiv r_0 \neq 0$ for all t. From (3.9) we know $x = x_d = \alpha r_0$ when $t \to \infty$, where α is the constant DC gain of the stable filter. Therefore, the linear equation (3.18) reduces to

  $$\alpha \tilde{a}_x + \tilde{a}_r = 0.$$

  This implies that $\tilde{a}$ does not necessarily converge to zero. In fact, it converges to a straight line in the parameter space.

- **Persistent excitation: guaranteed convergence**. However, when the signal v satisfies the so-called *persistent excitation* condition, which states that for any t, there exists $T, \beta > 0$ such that

  $$\int_t^{t+T} v v^T d\tau \geq \beta I, \tag{3.19}$$

  then $\tilde{a}$ is guaranteed to converge to zero. To see this, we multiply (3.18) by v and integrate it from t to $t + T$, which gives rise to

  $$\left(\int_t^{t+T} v v^T d\tau \right) \tilde{a} = 0.$$

By the persistent excitation condition (3.19), we infer that $\tilde{a} = 0$ is the only solution.

It remains to understand under what conditions of the reference trajectory $r(t)$ can we guarantee the persistent excitation of v. We leave it as an exercise for the reader to show, if $r(t)$ contains at least one sinusoidal component, then the persistent excitation condition of v is guaranteed.

3.1.2 Nonlinear Systems

Appendix A

The Kalman-Yakubovich Lemma

Lemma A.1 (Kalman-Yakubovich)**.** *Consider a controllable linear time-invariant system*

$$\dot{x} = Ax + buy = c^T x.$$

The transfer function

$$h(p) = c^T (pI - A)^{-1} b$$

is strictly positive real (SPR) if and only if there exist positive definite matrices P and Q such that

$$A^T P + PA = -QPb = c.$$