

# Optimal Control and Estimation

Heng Yang

2023-07-19



# Contents

<b>Preface</b>	<b>5</b>
<b>1 The Optimal Control Formulation</b>	<b>7</b>
1.1 The Basic Problem . . . . .	7
1.2 Dynamic Programming and Principle of Optimality . . . . .	9
<b>2 Exact Dynamic Programming</b>	<b>13</b>
2.1 Linear Quadratic Regulator . . . . .	13
<b>3 Approximate Dynamic Programming</b>	<b>15</b>
<b>4 Stability Analysis</b>	<b>17</b>
4.1 Autonomous Systems . . . . .	18
<b>5 Output Feedback</b>	<b>25</b>
5.1 State Observer . . . . .	25
5.2 Observer Feedback . . . . .	44
<b>6 Adaptive Control</b>	<b>45</b>
6.1 Model-Reference Adaptive Control . . . . .	45
6.2 Certainty-Equivalent Adaptive Control . . . . .	55
<b>A Linear System Theory</b>	<b>57</b>
A.1 Stability . . . . .	57
<b>B Convex Analysis and Optimization</b>	<b>61</b>
<b>C The Kalman-Yakubovich Lemma</b>	<b>63</b>
<b>D Feedback Linearization</b>	<b>65</b>
<b>E Sliding Control</b>	<b>67</b>



# 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

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

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.,  $\mu_k$  is a *feedback* controller that maps the state to an admissible control. Given an initial state  $x_0$  and an admissible policy  $\pi$ , the state *trajectory* of the system is a sequence of random variables that evolve according to

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

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*

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

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

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

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

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)

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

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

*Remark* (Open-loop v.s. Closed-loop). An important feature of the basic problem in Definition 1.1 is that the problem seeks *feedback policies*, instead of numerical values of the controls, i.e.,  $u_k = \mu_k(x_k)$  is in general a function of the state  $x_k$ . In other words, the controls are executed sequentially, one at a time after observing the state at each time. This is called closed-loop control, and is in general better than open-loop control

$$\min_{u_0, \dots, u_{N-1}} \mathbb{E} \left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k) \right\}$$

where all the controls are planned at  $k = 0$ . Intuitively, a closed-loop policy is able to utilize the extra information received at each timestep (i.e., it observes  $x_{k+1}$  and hence also observes the disturbance  $w_k$ ) to obtain a lower cost than an open-loop controller. Example 1.2.1 in (?) gives a concrete application where a closed-loop policy attains a lower cost than an open-loop policy.



In deterministic control (i.e., when  $w_k \equiv 0, \forall k$ ), however, a closed-loop policy has no advantage over an open-loop controller. This is obvious because at  $k = 0$ , even the open-loop controller predicts perfectly the consequences of all its actions and there is no extra information to be observed at later time steps. In fact, even in stochastic problems, a closed-loop policy may not be advantageous, see Exercise 1.27 in (?).

## 1.2 Dynamic Programming and Principle of Optimality

We now introduce a general and powerful algorithm, namely *dynamic programming* (DP), for solving the optimal control problem 1.1. The DP algorithm builds upon a quite simple intuition called the *Bellman principle of optimality*.

**Theorem 1.1** (Bellman Principle of Optimality). *Let  $\pi^* = \{\mu_0^*, \mu_1^*, \dots, \mu_{N-1}^*\}$  be an optimal policy for the optimal control problem 1.1. Assume that when using  $\pi^*$ , a given state  $x_i$  occurs at timestep  $i$  with positive probability (i.e.,  $x_i$  is reachable at time  $i$ ).*

*Now consider the following subproblem where we are at  $x_i$  at time  $i$  and wish to minimize the cost-to-go from time  $i$  to time  $N$*

$$\min_{\mu_i, \dots, \mu_{N-1}} \mathbb{E} \left\{ g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k)) \right\}.$$

*Then the truncated policy  $\{\mu_i^*, \mu_{i+1}^*, \dots, \mu_{N-1}^*\}$  must be optimal for the subproblem.*

Theorem 1.1 can be proved intuitively by contradiction: if the truncated policy  $\{\mu_i^*, \mu_{i+1}^*, \dots, \mu_{N-1}^*\}$  is not optimal for the subproblem, say there exists a different policy  $\{\mu'_i, \mu'_{i+1}, \dots, \mu'_{N-1}\}$  that attains a lower cost for the subproblem starting at  $x_i$  at time  $i$ . Then the combined policy  $\{\mu_0^*, \dots, \mu_{i-1}^*, \mu'_i, \dots, \mu'_{N-1}\}$  must attain a lower cost for the original optimal control problem 1.1 due to the additive cost structure, contradicting the optimality of  $\pi^*$ .

The Bellman principle of optimality is more than just a principle, it is also an algorithm. It suggests that, to build an optimal policy, one can start by solving the last-stage subproblem to obtain  $\{\mu_{N-1}^*\}$ , and then proceed to solve the subproblem containing the last two stages to obtain  $\{\mu_{N-2}^*, \mu_{N-1}^*\}$ . The recursion continues until optimal policies at all stages are computed. The following theorem formalizes this concept.

**Theorem 1.2** (Dynamic Programming). *The optimal value function  $J^*(x_0)$  of the optimal control problem 1.1 (starting from any given initial condition  $x_0$ ) is*

equal to  $J_0(x_0)$ , which can be computed backwards and recursively as

$$J_N(X_N) = g_N(x_N) \quad (1.6)$$

$$J_k(x_k) = \min_{u_k \in \mathbb{U}} \mathbb{E}_{w_k \sim P_k(\cdot | x_k, u_k)} \{g_k(x_k, u_k) + J_{k+1}(f_k(x_k, u_k, w_k))\}, \quad k = N-1, \dots, 1, 0. \quad (1.7)$$

Moreover, if  $u_k^* = \mu_k^*(x_k)$  is a minimizer of (1.7) for every  $x_k$ , then the policy  $\pi^* = \{\mu_0^*, \dots, \mu_{N-1}^*\}$  is optimal.

*Proof.* For any admissible policy  $\pi = \{\mu_0, \dots, \mu_{N-1}\}$ , denote  $\pi^k = \{\mu_k, \dots, \mu_{N-1}\}$  the last- $(N-k)$ -stage truncated policy. Consider the subproblem consisting of the last  $N-k$  stages starting from  $x_k$ , and let  $J_k^*(x_k)$  be its optimal cost-to-go. Mathematically, this is

$$J_k^*(x_k) = \min_{\pi^k} \mathbb{E}_{w_k, \dots, w_{N-1}} \left\{ g_N(x_N) + \sum_{i=k}^{N-1} g_i(x_i, \mu_i(x_i)) \right\}, \quad k = 0, 1, \dots, N-1. \quad (1.8)$$

We define  $J_N^*(x_N) = g(x_N)$  for  $k = N$ .

Our goal is to prove the  $J_k(x_k)$  computed by dynamic programming from (1.7) is equal to  $J_k^*(x_k)$  for all  $k = 0, \dots, N$ . We will prove this by induction.

Firstly, we already have  $J_N^*(x_N) = J_N(x_N) = g(x_N)$ , so  $k = N$  holds automatically.

Now we assume  $J_{k+1}^*(x_{k+1}) = J_{k+1}(x_{k+1})$  for all  $x_{k+1}$ , and we wish to induce

$J_k^*(x_k) = J_k(x_k)$ . To show this, we write

$$J_k^*(x_k) = \min_{\pi^k} \mathbb{E}_{w_k, \dots, w_{N-1}} \left\{ g_N(x_N) + \sum_{i=k}^{N-1} g_i(x_i, \mu_i(x_i)) \right\} \quad (1.9)$$

$$= \min_{\mu_k, \pi^{k+1}} \mathbb{E}_{w_k, \dots, w_{N-1}} \left\{ g_k(x_k, \mu_k(x_k)) + g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i)) \right\} \quad (1.10)$$

$$= \min_{\mu_k} \left[ \min_{\pi^{k+1}} \mathbb{E}_{w_k, \dots, w_{N-1}} \left\{ g_k(x_k, \mu_k(x_k)) + g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i)) \right\} \right] \quad (1.11)$$

$$= \min_{\mu_k} \mathbb{E}_{w_k} \left\{ g_k(x_k, \mu_k(x_k)) + \min_{\pi^{k+1}} \left[ \mathbb{E}_{w_{k+1}, \dots, w_{N-1}} \left\{ g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i, \mu_i(x_i)) \right\} \right] \right\} \quad (1.12)$$

$$= \min_{\mu_k} \mathbb{E}_{w_k} \{ g_k(x_k, \mu_k(x_k)) + J_{k+1}^*(f_k(x_k, \mu_k(x_k), w_k)) \} \quad (1.13)$$

$$= \min_{\mu_k} \mathbb{E}_{w_k} \{ g_k(x_k, \mu_k(x_k)) + J_{k+1}(f_k(x_k, \mu_k(x_k), w_k)) \} \quad (1.14)$$

$$= \min_{u_k \in \mathbb{U}} \mathbb{E}_{w_k} \{ g_k(x_k, \mu_k(x_k)) + J_{k+1}(f_k(x_k, \mu_k(x_k), w_k)) \} \quad (1.15)$$

$$= J_k(x_k), \quad (1.16)$$

where (1.9) follows from definition (1.8); (1.10) expands  $\pi^k = \{\mu_k, \pi^{k+1}\}$  and  $\sum_{i=k}^{N-1} g_i = g_k + \sum_{i=k+1}^{N-1} g_i$ ; (1.11) writes the joint minimization over  $\mu_k$  and  $\pi^{k+1}$  as equivalently first minimizing over  $\pi^{k+1}$  and then minimizing over  $\mu_k$ ; (1.12) is the key step and holds because  $g_k$  and  $w_k$  depend only on  $\mu_k$  but not on  $\pi^{k+1}$ ; (1.13) follows again from definition (1.8) with  $k$  replaced by  $k+1$ ; (1.14) results from the induction assumption; (1.15) clearly holds because any  $\mu_k(x_k)$  belongs to  $\mathbb{U}$  and any element in  $\mathbb{U}$  can be chosen by a feedback controller  $\mu_k$ ; and lastly (1.16) follows from the dynamic programming algorithm (1.7).

By induction, this shows that  $J_k^*(x_k) = J_k(x_k)$  for all  $k = 0, \dots, N$ .  $\square$

The careful reader, especially from a robotics background, may soon become disappointed when seeing the DP algorithm (1.7) because it is rather conceptual than practical. To see this, we only need to run DP for  $k = N-1$ :

$$J_{N-1}(x_{N-1}) = \min_{u_{N-1} \in \mathbb{U}} \mathbb{E}_{w_{N-1}} \{ g_{N-1}(x_{N-1}, u_{N-1}) + J_N(f_{N-1}(x_{N-1}, u_{N-1}, w_{N-1})) \}. \quad (1.17)$$

Two challenges immediately show up:

- How to perform the minimization over  $u_{N-1}$  when  $\mathbb{U}$  is a continuous constraint set? Even if we assume  $g_{N-1}$  is convex<sup>1</sup> in  $u_{N-1}$ ,  $J_N$  is convex in

---

<sup>1</sup>You may want to read Appendix B if this is your first time seeing “convex” things.

$x_N$ , and the dynamics  $f_{N-1}$  is also convex in  $u_{N-1}$  (so that the optimization (1.17) is convex), we may be able to solve the minimization *numerically* for each  $x_{N-1}$  using a convex optimization solver, but rarely will we be able to find an analytical policy  $\mu_{N-1}^*$  such that  $u_{N-1}^* = \mu_{N-1}^*(x_{N-1})$  for every  $x_{N-1}$  (i.e., the optimal policy  $\mu_{N-1}^*$  is implicit but not explicit).

- Suppose we can find an analytical optimal policy  $\mu_{N-1}^*$ , say  $\mu_{N-1}^* = Kx_{N-1}$  a linear policy, how will plugging  $\mu_{N-1}^*$  into (1.17) affect the complexity of  $J_{N-1}(x_{N-1})$ ? One can see that even if  $\mu_{N-1}^*$  is linear in  $x_{N-1}$ ,  $J_{N-1}$  may be highly nonlinear in  $x_{N-1}$  due to the composition with  $g_{N-1}$ ,  $f_{N-1}$  and  $J_N$ . If  $J_{N-1}(x_{N-1})$  becomes too complex, then clearly it becomes more challenging to perform (1.17) for the next step  $k = N - 2$ .

Due to these challenges, only in a very limited amount of cases will we be able to perform *exact dynamic programming*. For example, when the state space  $\mathbb{X}$  and control space  $\mathbb{U}$  are discrete, we can design efficient algorithms for exact DP. For another example, when the dynamics  $f_k$  is linear and the cost  $g_k$  is quadratic, we will also be able to compute  $J_k(x_k)$  in closed form (though this sounds a bit surprising!). We will study these problems in more details in Chapter 2.

For general optimal control problems with continuous state space and control space (and most problems we care about in robotics), unfortunately, we will have to resort to *approximate dynamic programming*, basically variations of the DP algorithm (1.7) where approximate value functions  $J_k(x_k)$  and/or control policies  $\mu_k(x_k)$  are used (e.g., with neural networks and machine learning).<sup>2</sup> We will introduce several popular approximation schemes in Chapter 3. We will see that, although exact DP is not possible anymore, the Bellman principle of optimality still remains one of the most important guidelines for designing approximation algorithms. Efficient algorithms for approximate dynamic programming, preferably with performance guarantees, still remain an active area of research.

---

<sup>2</sup>Another possible solution is to discretize continuous states and controls. However, when the dimension of state and control is high, discretization becomes too expensive in terms of memory and computational complexity.

## Chapter 2

# Exact Dynamic Programming

### 2.1 Linear Quadratic Regulator



## Chapter 3

# Approximate Dynamic Programming





## Chapter 4

# Stability Analysis

Optimal control formulates a control problem via the language of mathematical optimization. However, there are control problems, and sometimes even the very basic control problems, that cannot be easily stated in the optimal control formulation.

For example, suppose our goal is to *swing up a pendulum to the upright position and stabilize it there*. You may want to formalize the problem as

$$\min_{u(t) \in \mathbb{U}} \int_0^\infty \|x(t) - x_d\|^2 dt, \quad \text{subject to} \quad \dot{x} = f(x, u), x(0) = x_0, \quad (4.1)$$

where  $x_d$  is the desired upright position for the pendulum. However, does the solution of problem (4.1), if exists, guarantee the stabilization of the pendulum at the upright position? The answer is unclear without a rigorous proof.

However, after a slight change of perspective, the optimal control problem may be formulated to better match the goal. Suppose there exists a region,  $\Omega$ , in the state space such that as long as the pendulum enters  $\Omega$ , there always exists a sequence of control to bring the pendulum to the goal state  $x_d$ , then we can simply formulate a different optimal control problem

$$\min_{u(t) \in \mathbb{U}} \int_0^T \|u(t)\|^2 dt, \quad \text{subject to} \quad x(0) = x_0, x(T) \in \Omega, \dot{x} = f(x, u), \quad (4.2)$$

where now it is very clear, if a solution exists to problem (4.2), then we will definitely achieve our goal. This is because the constraint  $x(T) \in \Omega$  guarantees that we will be able to stabilize the pendulum, and the cost function of (4.2) simply encourages minimum control effort along the way.

This highlights that, sometimes the formulation of a problem may deserve more thoughts than the actual solution. Of course the formulation (4.2) may be much more difficult to solve. In fact, does the set  $\Omega$  exist, and if so, how to describe it?

This is the main focus of this chapter: to introduce tools that can help us analyze the *stability* of uncontrolled and controlled nonlinear systems. Specifically, we will introduce the notion of *stability certificates*, which are conditions that, if hold, certify the stability of the system (e.g., in the set  $\Omega$ ). Interestingly, you will see that the notion of stability certificates is intuitive and easy, but what is really challenging is to *find* and *compute* the stability certificates. We will highlight the power and also limitation of computational tools, especially those that are based on convex optimization (see Appendix B for a review of convex optimization).

## 4.1 Autonomous Systems

Let us first focus on autonomous systems, i.e., systems whose dynamics do not depend on time (and control). We introduce different concepts of stability and ways to certify them.

### 4.1.1 Concepts of Stability

Consider the autonomous system

$$\dot{x} = f(x) \quad (4.3)$$

where  $x \in \mathbb{X} \subseteq \mathbb{R}^n$  is the state and  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is the (potentially nonlinear) dynamics.

Before talking about concepts of stability, we need to define an *equilibrium point*.

**Definition 4.1** (Equilibrium Point). A state  $x^*$  is called an equilibrium point of system (4.3) if  $f(x^*) = 0$ , i.e., once the system reaches  $x^*$ , it stays at  $x^*$ .

For example, a linear system

$$\dot{x} = Ax$$

has a single equilibrium point  $x^* = 0$  when  $A$  is nonsingular, and an infinite number of equilibrium points when  $A$  is singular (those equilibrium points lie in the kernel of matrix  $A$ ).

When analyzing the behavior of a dynamical system around the equilibrium point, it is often helpful to “shift” the dynamics equation so that 0 is the equilibrium point. For example, if we are interested in the behavior of system (4.3) near the equilibrium point  $x^*$ , we can create a new variable

$$z = x - x^*,$$

so that

$$\dot{z} = \dot{x} = f(x) = f(z + x^*). \quad (4.4)$$

Clearly,  $z^* = 0$  is an equilibrium point for the shifted system (4.4).

Let us find the equilibrium points of a simple pendulum.

**Example 4.1** (Equilibrium Points of A Simple Pendulum). Consider the dynamics of an uncontrolled pendulum

$$\begin{cases} \dot{\theta} = \dot{\theta} \\ \ddot{\theta} = -\frac{1}{ml^2}(b\dot{\theta} + mgl \sin \theta) \end{cases} \quad (4.5)$$

where  $\theta$  is the angle between the pendulum and the vertical line, and  $x = [\theta, \dot{\theta}]^T$  is the state of the pendulum ( $m, g, l, b$  denote the mass, gravity constant, length, and damping constant, respectively).

To find the equilibrium points of the pendulum, we need the right hand sides of (4.5) to be equal to zero:

$$\dot{\theta} = 0, \quad -\frac{1}{ml^2}(b\dot{\theta} + mgl \sin \theta) = 0.$$

The solutions are easy to find

$$x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \pi \\ 0 \end{bmatrix},$$

corresponding to the bottomright and upright positions of the pendulum, respectively.

The pendulum dynamics has two equilibrium points, but our physics intuition tells us these two equilibrium points are dramatically different. Specifically, the bottomright equilibrium  $x^* = [0, 0]^T$  is such that if you perturb the pendulum around the equilibrium, the pendulum will go back to that equilibrium; the upright equilibrium  $x^* = [\pi, 0]^T$  is such that if you perturb the pendulum (even just a little bit) around the equilibrium, it will diverge from that equilibrium.

This physical intuition is exactly what we want to formalize as the concepts of stability.

In the following, we focus on the nonlinear autonomous system (4.3) with  $f(0) = 0$ , i.e.,  $x^* = 0$  is an equilibrium point. We now formally define the different concepts of stability.

**Definition 4.2** (Lyapunov Stability). The equilibrium point  $x = 0$  is said to be *stable in the sense of Lyapunov* if, for any  $R > 0$ , there exists  $r > 0$  such that if  $\|x(0)\| < r$ , then  $\|x(t)\| < R$  for all  $t \geq 0$ . Otherwise, the equilibrium point is unstable.

For a system that is Lyapunov stable around  $x = 0$ , the definition says that, if we want to constrain the trajectory of the system to be within the ball  $B_R = \{x \mid \|x\| < R\}$ , then we can always find a smaller ball  $B_r = \{x \mid \|x\| < r\}$  such that if the system starts within  $B_r$ , it will remain in the larger ball  $B_R$ .

On the other hand, if the system is not Lyapunov stable at  $x = 0$ , then there exists at least one ball  $B_R$ , such that no matter how close the system's initial condition is to the origin, it will eventually exit the ball  $B_R$ . The following exercise is left for you to verify the instability of the Van der Pol oscillator.

**Exercise 4.1** (Instability of the Van der Pol oscillator). Show that the Van der Pol oscillator

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_1 + (1 - x_1^2)x_2 \end{cases}$$

is unstable at the equilibrium point  $x = 0$ .

Lyapunov stability does not guarantee the system trajectory will actually converge to  $x = 0$ . Instead, asymptotic stability will ask the system trajectory to converge to  $x = 0$ .

**Definition 4.3** (Asymptotic Stability and Domain of Attraction). The equilibrium point  $x = 0$  is said to be *asymptotically stable* if (i) it is Lyapunov stable, and (ii) there exists some  $r > 0$  such that  $x(0) \in B_r$  implies  $x(t) \rightarrow 0$  as  $t \rightarrow \infty$ .

The domain of attraction (for the equilibrium  $x = 0$ ) is the largest set of points in the state space such that trajectories initiated at those points will converge to the equilibrium point. That is,

$$\Omega(x^*) = \{x \in \mathbb{X} \mid x(0) = x \implies \lim_{t \rightarrow \infty} x(t) = x^*\}.$$

The ball  $B_r$  is a domain of attraction for the equilibrium point  $x = 0$ , but not necessarily the largest domain of attraction.

You may immediately realize that in the definition of asymptotic stability, we require Lyapunov stability to hold first. Is this necessary? i.e., does there exist a system where trajectories eventually converge to zero, but is not stable in the sense of Lyapunov? You should work out the following exercise.

**Exercise 4.2** (Vinograd System). Show that for the Vinograd dynamical system (?)

$$\begin{cases} \dot{x} = \frac{x^2(y-x)+y^5}{(x^2+y^2)(1+(x^2+y^2)^2)} \\ \dot{y} = \frac{y^2(y-2x)}{(x^2+y^2)(1+(x^2+y^2)^2)} \end{cases},$$

all system trajectories converge to the equilibrium point  $(x, y) = 0$ , but the equilibrium point is not stable in the sense of Lyapunov.

(Hint: the system trajectories will behave like the following plot.)

In many cases, we want the convergence of the system trajectory towards  $x = 0$  to be fast, thus bringing in the notion of exponential stability.

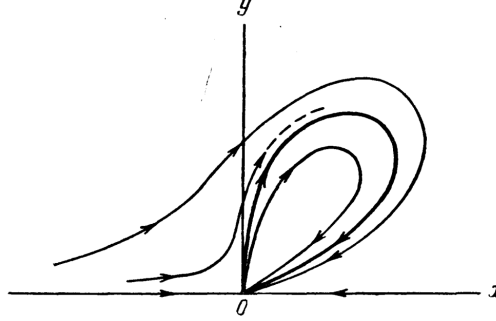


Figure 4.1: Trajectories of the Vinograd system. Copied from the original article of Vinograd.

**Definition 4.4** (Exponential Stability). An equilibrium point  $x = 0$  is said to be exponentially stable, if there exists a ball  $B_r$  such that as long as  $x(0) \in B_r$ , then

$$\|x(t)\| \leq \alpha \|x(0)\| e^{-\lambda t}, \quad \forall t,$$

for some  $\alpha > 0$  and  $\lambda > 0$  ( $\lambda$  is called the rate of exponential convergence).

Exponential stability implies asymptotic stability (and certainly also Lyapunov stability). What is nice about exponential stability is that we can quantify the distance of the system trajectory to the equilibrium point as a function of time (as long as we know the constants  $\alpha, \|x(0)\|, \lambda$ ). In many safety-critical applications, we need such performance guarantees. For example, in Chapter 5.1, we will see the application of exponential stability in observer-feedback control.

All the concepts of stability we have mentioned so far only talk about the stability of the system *locally* around the equilibrium point  $x = 0$  (via arguments like  $B_r$  and  $B_R$ ). It would be much nicer if we can guarantee stability of the system *globally*, i.e., no matter where the system starts in the state space  $\mathbb{X}$ , its trajectory will converge to  $x = 0$ .

**Definition 4.5** (Global Asymptotic and Exponential Stability). The equilibrium point  $x = 0$  is said to be globally asymptotically (exponentially) stable if asymptotic (exponential) stability holds for any initial states. That is,

$$\forall x \in \mathbb{X}, \quad x(0) = x \implies \begin{cases} \lim_{t \rightarrow \infty} x(t) = 0 & \text{global asymptotic stability} \\ \exists \alpha, \lambda > 0, \text{ s.t. } \|x(t)\| \leq \alpha \|x(0)\| e^{-\lambda t} & \text{global exponential stability} \end{cases}$$

This concludes our definitions of stability for nonlinear systems (Definition 4.2-4.5). It is worth mentioning that the concepts of stability are complicated (refined) here due to our focus on nonlinear systems. For linear systems, the

concepts of stability are simpler. Specifically, all local stability properties of linear systems are also global and asymptotic stability is equal to exponential stability. In fact, for a linear time-invariant system  $\dot{x} = Ax$ , it is either asymptotically (exponentially) stable, or marginally stable, or unstable. Moreover, we can fully characterize the stability property by inspecting the eigenvalues of  $A$  (you can find a refreshment of this in Appendix A.1).

How do we characterize the stability property of a nonlinear system? If someone gave me a nonlinear system (4.3), how can I provide a certificate to her that the system is stable or unstable (I cannot use eigenvalues anymore in this case)? Let us describe some of these certificates below.

### 4.1.2 Stability by Linearization

A natural idea is to linearize, if possible, the nonlinear system (4.3) at a given equilibrium point  $x^*$  and inspect the stability of the linearized system (for which we can compute eigenvalues). Therefore, the key question here is how does the stability and instability of the linearized system relate to the stability and instability of the original nonlinear system.

**Theorem 4.1** (Stability by Linearization). *Assume  $x = 0$  is an equilibrium point of system (4.3) and  $f$  is continuously differentiable. Let*

$$\dot{x} = Ax, \quad A = \left. \frac{\partial f}{\partial x} \right|_{x=0} \quad (4.6)$$

*be the linearized system at  $x = 0$ . The following statements are true about the stability relationship between (4.3) and (4.6).*

- *If the linearized system (4.6) is strictly stable (i.e., all eigenvalues of  $A$  have strictly negative real parts), then the original system (4.3) is asymptotically stable at  $x = 0$ .*
- *If the linearized system (4.6) is unstable (i.e., at least one eigenvalue of  $A$  has strictly positive real part), then the original system (4.3) is unstable at  $x = 0$ .*
- *If the linearized system (4.6) is marginally stable (i.e., all eigenvalues of  $A$  have nonpositive real parts, and at least one eigenvalue has zero real part), then the stability of the original system (4.3) at  $x = 0$  is indeterminate.*

Theorem 4.1 is actually quite useful when we want to quickly examine the local stability of a nonlinear system around a given equilibrium point, as we will show in the next example.

**Example 4.2** (Stability of A Simple Pendulum by Linearization). Consider the simple pendulum dynamics (4.5) in Example 4.1. Without loss of generality, let

$m = 1, l = 1, b = 0.1$ . The Jacobian of the nonlinear dynamics reads

$$A = \frac{\partial f}{\partial x} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{l} \cos \theta & -\frac{b}{ml^2} \end{bmatrix}.$$

At the bottomright equilibrium point  $\theta = 0, \dot{\theta} = 0$ , the matrix  $A$  has two eigenvalues

$$-0.0500 \pm 3.13i,$$

and hence the pendulum is asymptotically stable at the bottomright equilibrium point.

At the upright equilibrium point  $\theta = \pi, \dot{\theta} = 0$ , the matrix  $A$  has two eigenvalues

$$3.08, \quad -3.18,$$

and hence the pendulum is unstable at the upright equilibrium point.

The linearization method is easy to carry out. However, it tells us nothing about global stability or exponential stability. Moreover, when the linearized system is marginally stable, the stability of the original system is inconclusive. In the next, we will introduce a more general, and perhaps the most popular framework for analyzing the stability of nonlinear systems.

### 4.1.3 Lyapunov Analysis

**Theorem 4.2** (Lyapunov Global Stability). *For the autonomous system (4.3), suppose there exists a scalar function  $V(x)$  with continuous first order derivatives such that*

- $V(x)$  is positive definite;
- $\dot{V}(x)$  is negative definite;
- $V(x) \rightarrow \infty$  as  $\|x\| \rightarrow \infty$ ,

*then the equilibrium point  $x = 0$  is globally asymptotically stable.*

**Lemma 4.1** (Barbalat's Lemma). *Let  $f(t)$  be differentiable, if*

- $\lim_{t \rightarrow \infty} f(t)$  is finite, and
- $\dot{f}(t)$  is uniformly continuous,<sup>1</sup>

*then*

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

**Theorem 4.3** (Barbalat's Stability Certificate). *If a scalar function  $V(x, t)$  satisfies*

---

<sup>1</sup>A sufficient condition for this to hold is that  $\ddot{f}$  exists and is bounded.

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

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

*Proof.*  $V(x, t)$  is lower bounded and  $\dot{V}$  is negative semidefinite implies the limit of  $V$  as  $t \rightarrow \infty$  is finite (note that  $V(x, t) \leq V(x(0), 0)$ ). Then the theorem clearly follows from Barbalat's Lemma 4.1.  $\square$



## Chapter 5

# Output Feedback

Consider a continuous-time dynamical system

$$\begin{aligned}\dot{x} &= f(x, u) \\ y &= h(x, u)\end{aligned}\tag{5.1}$$

where  $x(t) \in \mathbb{X} \subseteq \mathbb{R}^n$  the state of the system,  $u(t) \in \mathbb{U} \subseteq \mathbb{R}^m$  the control (or input),  $y(t) \in \mathbb{Y} \subseteq \mathbb{R}^d$  the output (i.e., measurement) of the state and control, and  $f, g$  the evolution and measurement functions (which are sufficiently smooth).

### 5.1 State Observer

For the system (5.1), let us denote

- $X(x_0, t_0; t; u)$  the solution at time  $t$  with input  $u$  and initial condition  $x_0$  at time  $t_0$ ; when  $t_0 = 0$ , we write  $X(x_0; t; u)$
- $Y(x_0, t_0; t; u)$  the output at time  $t$  with input  $u$  and initial condition  $x_0$  at time  $t_0$ , i.e.,  $Y(x_0, t_0; t; u) = h(X(x_0, t_0; t; u), u(t))$ ; when  $t_0 = 0$ , we write  $y_{x_0, u}(t)$ ;
- $\mathcal{X}_0$  a subset of  $\mathbb{X}$  containing the initial conditions we consider; for any  $x_0 \in \mathcal{X}_0$ , we write  $\sigma_{\mathcal{X}}^+(x_0; u)$  the maximal time of existence of  $X(x_0, \cdot; t; u)$  in a set  $\mathcal{X}$
- $\mathcal{U}$  the set of all sufficiently many times differentiable inputs  $u : [0, +\infty) \rightarrow \mathbb{U}$ .

The problem of state observation is to produce an estimated state  $\hat{x}(t)$  of the true state  $X(x_0, t_0; t; u)$  based on knowledge about the system (5.1) and information about the history of inputs  $u_{[0, t]}$  and outputs  $y_{[0, t]}$ , so that  $\hat{x}(t)$  asymptotically converges to  $X(x_0, t_0; t; u)$ , for any initial condition  $x_0 \in \mathcal{X}_0$  and any input  $u \in \mathcal{U}$ .

There are multiple ways for solving the problem of state observation (see e.g., (?), (?)). Here we are particularly interested in the approach using a *state observer*, i.e., a dynamical system whose *internal state* evolves according to the history of inputs and outputs, from which a state estimation can be reconstructed that guarantees asymptotic convergence to the true state. We formalize this concept below.

**Definition 5.1** (State Observer). A state observer for system (5.1) is a couple  $(\mathcal{F}, \mathcal{T})$  such that

1.  $\mathcal{F} : \mathbb{R}^l \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^l$  is continuous
2.  $\mathcal{T}$  is a family of continuous functions indexed by  $u \in \mathcal{U}$  where each  $\mathcal{T}_u : \mathbb{R}^l \times [0, +\infty) \rightarrow \mathbb{R}^n$  respects the causality condition

$$\forall \tilde{u} : [0, +\infty) \rightarrow \mathbb{R}^m, \forall t \in [0, +\infty), u_{[0,t]} = \tilde{u}_{[0,t]} \Rightarrow \mathcal{F}_u(\cdot, t) = \mathcal{F}_{\tilde{u}}(\cdot, t).$$

3. For any  $u \in \mathcal{U}$ , any  $z_0 \in \mathbb{R}^l$ , and any  $x_0 \in \mathcal{X}_0$  such that  $\sigma_{\mathbb{X}}^+(x_0; u) = +\infty$ , any solution  $Z(z_0; t; u, y_{x_0, u})$ <sup>1</sup> to

$$\dot{z} = \mathcal{F}(z, u, y_{x_0, u}) \quad (5.2)$$

initialized at  $z_0$  at time 0 with input  $u$  and  $y_{x_0, u}$  exists on  $[0, +\infty)$  and satisfies

$$\lim_{t \rightarrow \infty} \|\hat{X}(x_0, z_0; t; u) - X(x_0; t; u)\| = 0, \quad (5.3)$$

with

$$\hat{X}(x_0, z_0; t; u) = \mathcal{T}_u(Z(z_0; t; u, y_{x_0, u}), t). \quad (5.4)$$

In words, (i) the state observer maintains an internal state (or latent state)  $z \in \mathbb{R}^l$  that evolves according to the latent dynamics  $\mathcal{F}$  in (5.2), where  $u$  and  $y_{x_0, u}$  are inputs; (ii) an estimated state can be reconstructed from the internal state using  $\mathcal{T}_u$  as in (5.4); and (iii) the error between the estimated state and the true state (defined by a proper distance function  $\|\cdot\|$  on  $\mathbb{X}$ ) converges to zero.

If  $\mathcal{T}_u$  is the same for any  $u \in \mathcal{U}$  and is also time independent, then we say  $\mathcal{T}$  is *stationary*.<sup>2</sup> In this case, we can simply write the observer (5.2) and (5.4) as

$$\begin{aligned} \dot{z} &= \mathcal{F}(z, u, y) \\ \hat{x} &= \mathcal{T}(z). \end{aligned} \quad (5.5)$$

<sup>1</sup>We say “any solution” because there may be several solutions to the observer (5.2) due to  $\mathcal{F}$  only being continuous. This is not a problem as long as any such solution satisfies the required convergence property.

<sup>2</sup>The time dependence of  $\mathcal{T}_u$  enables us to cover the case where the knowledge of the  $u$  and  $y_{x_0, u}$  is used to construct the estimate from the observer state. In particular, using the output sometimes can reduce the dimension of the observer state (and thus alleviate the computations), thus obtaining a reduced-order observer. For example, see (?) and (?).

If  $\hat{x}$  can be read off directly from  $z$ , then we say the observer (5.5) is *in the given coordinates*. A special case of this is when  $\hat{x} = z$ , i.e., the internal state of the observer is the same as the system state.

### 5.1.1 General Design Strategy

**Theorem 5.1** (Meta Observer). *Let  $F : \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^p$ ,  $H : \mathbb{R}^p \times \mathbb{R}^m \rightarrow \mathbb{R}^d$  and  $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^p$  be continuous functions such that*

$$\dot{\hat{\xi}} = \mathcal{F}(\hat{\xi}, u, \tilde{y}) \quad (5.6)$$

*is an observer for*

$$\dot{\xi} = F(\xi, u, H(\xi, u)), \quad \tilde{y} = H(\xi, u), \quad (5.7)$$

*i.e., for any  $\xi_0, \hat{\xi}_0 \in \mathbb{R}^p$  and any  $u \in \mathcal{U}$ , the solution of the observer (5.6), denoted by  $\hat{\Xi}(\hat{\xi}_0; t; u; \tilde{y}_{\xi_0, u})$ , and the solution of the true system (5.7), denoted by  $\Xi(\xi_0; t; u)$ , satisfy*

$$\lim_{t \rightarrow \infty} \|\hat{\Xi}(\hat{\xi}_0; t; u; \tilde{y}_{\xi_0, u}) - \Xi(\xi_0; t; u)\| = 0. \quad (5.8)$$

*Note that the observer (5.6) is stationary and in the given coordinates for system (5.7). Indeed the internal state of the observer is the same as the system state.*

*Now suppose for any  $u \in \mathcal{U}$ , there exists a continuous function (i.e., coordinate transformation)  $T_u : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^p$  and a subset  $\mathcal{X}$  of  $\mathbb{X}$  such that*

1. *For any  $x_0 \in \mathcal{X}_0$  such that  $\sigma_{\mathbb{X}}^+(x_0; u) = +\infty$ ,  $X(x_0; \cdot; u)$  remains in  $\mathcal{X}$*
2. *There exists a concave  $\mathcal{K}^3$  function  $\rho$  and a positive number  $\bar{t}$  such that*

$$\|x_a - x_b\| \leq \rho(|T_u(x_a, t) - T_u(x_b, t)|), \quad \forall x_a, x_b \in \mathcal{X}, t \geq \bar{t},$$

*i.e.,  $x \mapsto T_u(x, t)$  becomes injective on  $\mathcal{X}$ ,<sup>4</sup> uniformly in time and space, after a certain time  $\bar{t}$ .*

3.  *$T_u$  transforms the system (5.1) into the system (5.7), i.e., for all  $x \in \mathcal{X}$  and all  $t \geq 0$ , we have*

$$L_{(f,1)}T_u(x, t) = F(T_u(x, t), u, h(x, u)), \quad h(x, u) = H(T_u(x, t), u), \quad (5.9)$$

*where  $L_{(f,1)}T_u(x, t)$  is the Lie derivative of  $T_u$  along the vector field  $(f, 1)$*

$$L_{(f,1)}T_u(x, t) = \lim_{\tau \rightarrow 0} \frac{T_u(X(x, t; t + \tau; u), t + \tau) - T_u(x, t)}{\tau}.$$

<sup>3</sup>A function  $\rho : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a  $\mathcal{K}$  function if  $\rho(0) = 0$ ,  $\rho$  is continuous, and  $\rho$  is increasing.

<sup>4</sup>An injective function is a function  $f$  that maps distinct elements of its domain to distinct elements. That is,  $f(x_a) = f(x_b)$  implies  $x_a = x_b$ , or equivalently,  $x_a \neq x_b$  implies  $f(x_a) \neq f(x_b)$ .

4.  $T_u$  respects the causality condition

$$\forall \tilde{u} : [0, +\infty) \rightarrow \mathbb{R}^m, \forall t \in [0, +\infty), u_{[0,t]} = \tilde{u}_{[0,t]} \Rightarrow T_u(\cdot, t) = T_{\tilde{u}}(\cdot, t).$$

Then, for any  $u \in \mathcal{U}$ , there exists a function  $\mathcal{T}_u : \mathbb{R}^p \times [0, +\infty) \rightarrow \mathcal{X}$  (satisfying the causality condition) such that for any  $t \geq \bar{t}$ ,  $\xi \mapsto \mathcal{T}_u(\xi, t)$  is uniformly continuous on  $\mathbb{R}^p$  and satisfies

$$\mathcal{T}_u(T_u(x, t), t) = x, \forall x \in \mathcal{X}.$$

Moreover, denoting  $\mathcal{T}$  the family of functions  $\mathcal{T}_u$  for  $u \in \mathcal{U}$ , the couple  $(\mathcal{F}, \mathcal{T})$  is an observer for the system (5.1) initialized in  $\mathcal{X}_0$ .

*Proof.* See Theorem 1.1 in (?). □

A simpler version of Theorem 5.1 where the coordinate transformation  $T_u$  is stationary and fixed for all  $u$  is stated below as a corollary.

**Corollary 5.1** (Meta Observer with Fixed Transformation). *Let  $F : \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^p$ ,  $H : \mathbb{R}^p \times \mathbb{R}^m \rightarrow \mathbb{R}^d$  and  $\mathcal{F} : \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^p$  be continuous functions such that (5.6) is an observer for (5.7).*

*Suppose there exists a continuous coordinate transformation  $T : \mathbb{R}^p \rightarrow \mathbb{R}^n$  and a compact subset  $\Omega$  of  $\mathbb{R}^n$  such that*

1. *For any  $x_0 \in \mathcal{X}_0$  such that  $\sigma_{\mathbb{X}}^+(x_0; u) = +\infty$ ,  $X(x_0; \cdot; u)$  remains in  $\Omega$*
2.  *$x \mapsto T(x)$  is injective on  $\Omega$*
3.  *$T$  transforms the system (5.1) into system (5.7)*

$$L_f T(x) = F(T(x), u, h(x, u)), \quad h(x, u) = H(T(x), u),$$

*where  $L_f T(x)$  is the Lie derivative of  $T(x)$  along  $f$*

$$L_f T(x) = \lim_{\tau \rightarrow 0} \frac{T(X(x, t; t + \tau; u)) - T(x)}{\tau}.$$

*Then, there exists a uniformly continuous function  $\mathcal{T} : \mathbb{R}^p \rightarrow \mathbb{R}^n$  such that*

$$\mathcal{T}(T(x)) = x, \quad \forall x \in \Omega,$$

*and  $(\mathcal{F}, \mathcal{T})$  is an observer for system (5.1) initialized in  $\mathcal{X}_0$ .*

Theorem 5.1 and Corollary 5.1 suggest the following general observer design strategy:

1. Find an injective coordinate transformation  $T_u$  (that may be time-varying and also dependent on  $u$ ) that transforms the original system (5.1) with coordinate  $x$  into a new system (5.7) with coordinate  $\xi$

2. Design an observer (5.6),  $\hat{\xi}$ , for the new system
3. Compute a left inverse,  $\mathcal{T}_u$ , of the transformation  $T_u$  to recover a state estimation  $\hat{x}$  of the original system.

The transformed systems (5.7) are typically referred to as *normal forms*, or in my opinion, *templates*.

Of course, the general design strategy is rather conceptual, and in order for it to be practical, we have to answer three questions.

- What templates do we have, what are their associated observers, and what are the conditions for the observers to be asymptotically converging?
- What kinds of (nonlinear) systems can be transformed into the templates, and how to perform the transformation?
- How to invert the coordinate transformation? Is it analytical or does it require numerical approximation?

In the following sections, we will study several representative normal forms and answer the above questions. Before presenting the results, let us first introduce several notions of observability.

**Definition 5.2** (Observability). Consider an open subset  $\mathcal{L}$  of the state space  $\mathbb{X} \subseteq \mathbb{R}^n$  of system (5.1). The system (5.1) is said to be

- **Distinguishable** on  $\mathcal{L}$  for some input  $u(t)$ , if for all  $(x_a, x_b) \in \mathcal{L} \times \mathcal{L}$ ,

$$y_{x_a, u}(t) = y_{x_b, u}(t), \forall t \in [0, \min \{\sigma_{\mathbb{X}}^+(x_a; u), \sigma_{\mathbb{X}}^+(x_b; u)\}] \implies x_a = x_b$$

- **Instantaneously distinguishable** on  $\mathcal{L}$  for some input  $u(t)$ , if for all  $(x_a, x_b) \in \mathcal{L} \times \mathcal{L}$ , and for all  $\bar{t} \in (0, \min \{\sigma_{\mathbb{X}}^+(x_a; u), \sigma_{\mathbb{X}}^+(x_b; u)\})$ ,

$$y_{x_a, u}(t) = y_{x_b, u}(t), \forall t \in [0, \bar{t}) \implies x_a = x_b$$

- **Uniformly observable** on  $\mathcal{L}$  if it is distinguishable on  $\mathcal{L}$  for any input  $u(t)$  (not only for  $u \in \mathcal{U}$ )
- **Uniformly instantaneously observable** on  $\mathcal{L}$  if it is instantaneously observable on  $\mathcal{L}$  for any input  $u(t)$  (not only for  $u \in \mathcal{U}$ ).

Moreover, let  $\mathcal{X}$  be a subset of  $\mathbb{X}$  such that  $\text{cl}(\mathcal{X})$ , i.e., the closure of  $\mathcal{X}$ , is contained in  $\mathcal{L}$ . Then the system (5.1) is said to be

- **Backward  $\mathcal{L}$ -distinguishable on  $\mathcal{X}$**  for some input  $u(t)$ , if for any  $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$  such that  $x_a \neq x_b$ , there exists  $t \in (\max \{\sigma_{\mathcal{L}}^-(x_a; u), \sigma_{\mathcal{L}}^-(x_b; u)\}, 0]$  such that  $y_{x_a, u}(t) \neq y_{x_b, u}(t)$ , or in words similar to the definition of distinguishable on  $\mathcal{L}$ , for all  $(x_a, x_b) \in \mathcal{X} \times \mathcal{X}$

$$y_{x_a, u}(t) = y_{x_b, u}(t), \forall t \in (\max \{\sigma_{\mathcal{L}}^-(x_a; u), \sigma_{\mathcal{L}}^-(x_b; u)\}, 0] \implies x_a = x_b.$$

### 5.1.2 Luenberger Template

Consider an instance of the normal form (5.7) as follows:

$$\dot{\xi} = A\xi + B(u, y), \quad y = C\xi, \quad (5.10)$$

where  $A, C$  are constant matrices, and  $B(u, y)$  can depend nonlinearly on  $u$  and  $y$ .

For this template, we have the well-known Luenberger observer.

**Theorem 5.2** (Luenberger Observer). *If the pair  $(A, C)$  is detectable (see Theorem A.3), then there exists a matrix  $K$  such that  $A - KC$  is Hurwitz and the system*

$$\dot{\hat{\xi}} = A\hat{\xi} + B(u, y) + K(y - C\hat{\xi}) \quad (5.11)$$

*is an observer for (5.10).*

*Proof.* Define  $e(t) = \xi(t) - \hat{\xi}(t)$ . In that case,

$$\dot{e}(t) = [A - KC]e(t) \quad (5.12)$$

Solving (5.12), we obtain

$$e(t) = \exp[(A - KC)t]e(0) \quad (5.13)$$

Then, if the real components of the eigenvalues of  $A - KC$  are strictly negative (i.e.,  $A - KC$  is Hurwitz), then  $e(t) \rightarrow 0$  as  $t \rightarrow \infty$ , independent of the initial error  $e(0) = \xi(0) - \hat{\xi}(0)$ . From Theorem A.3, we know that  $(A, C)$  being detectable implies the existence of  $K$  such that  $A - KC$  is Hurwitz.

If one is further interested in estimating the convergence rate of the Luenberger observer, then one can use the result from Corollary A.1. Particularly, one can solve the Lyapunov equation

$$(A - KC)^T P + P(A - KC) = -I$$

to obtain  $P$ . Then the convergence rate of  $\|e\|$  towards zero is  $\frac{0.5}{\lambda_{\max}(P)}$ .  $\square$

The Luenberger observer is an elegant result in observer design (and even in control theory) that has far-reaching impact. In my opinion, the essence of observer design is twofold: (i) to simulate the dynamics when the state estimation is correct, and (ii) to correct the state estimation from observation when it is off. These two pieces of ideas are evident in (5.11): the observer is a copy of the original dynamics ( $A\hat{\xi} + B(u, y)$ ) plus a feedback correction from the difference between the “imagined” observation  $C\hat{\xi}$  and the true observation  $y$ .

You may think the Luenberger template is restricting because it requires the system to be linear (up to the only nonlinearity in  $B(u, y)$ ). However, it turns out the Luenberger template is already quite useful, as I will show in the following pendulum example.

**Example 5.1** (Luenberger Observer for A Simple Pendulum). Consider a simple pendulum dynamics model

$$x = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix}, \quad \dot{x} = \begin{bmatrix} \dot{\theta} \\ -\frac{1}{ml^2}(b\dot{\theta} + mgl \sin \theta) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{ml^2} \end{bmatrix} u, \quad y = \theta, \quad (5.14)$$

where  $\theta$  the angular position of the pendulum from the vertical line,  $m > 0$  the mass of the pendulum,  $l > 0$  the length,  $g$  the gravitational constant,  $b > 0$  the damping coefficient, and  $u$  the control input (torque).

We assume we can only observe the angular position of the pendulum in (5.14), e.g., using a camera, but not the angular velocity. Our goal is to construct an observer that can provide a full state estimation.

We first note that the pendulum dynamics (5.14) can actually be written in the (linear) Luenberger template (5.10) as<sup>5</sup>

$$\begin{aligned} \dot{x} &= \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & -\frac{b}{ml^2} \end{bmatrix}}_{=:A} x + \underbrace{\begin{bmatrix} 0 \\ \frac{u - mgl \sin \theta}{ml^2} \end{bmatrix}}_{=:B(u,y)} \\ y &= \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{=:C} x \end{aligned} \quad (5.15)$$

In order for us to use the Luenberger observer, we need to check if the pair  $(A, C)$  is detectable. We can easily find the eigenvalues and eigenvectors of  $A$ :

$$A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0, \quad A \begin{bmatrix} -\frac{ml^2}{b} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{b}{ml^2} \end{bmatrix} = -\frac{b}{ml^2} \begin{bmatrix} -\frac{ml^2}{b} \\ 1 \end{bmatrix}.$$

The first eigenvalue has real part equal to 0. However,

$$C \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \neq 0.$$

According to Theorem A.3, we conclude  $(A, C)$  is detectable. In fact, the pair  $(A, C)$  is more than just detectable, it is indeed observable (according to Theorem A.2). Therefore, the poles of  $A - KC$  can be arbitrarily placed.

**Finding  $K$ .** Now we need to find  $K$ . An easy choice of  $K$  is

$$K = \begin{bmatrix} k \\ 0 \end{bmatrix}, \quad A - KC = \begin{bmatrix} -k & 1 \\ 0 & -\frac{b}{ml^2} \end{bmatrix}.$$

---

<sup>5</sup>I have to say I was a bit surprised when I arrived at this formulation.

With  $k > 0$ , we know  $A - KC$  is guaranteed to be Hurwitz (the two eigenvalues of  $A - KC$  are  $-k$  and  $-b/ml^2$ ), and we have obtained an observer!

We can also estimate the convergence rate of this observer. Let us use  $m = 1, g = 9.8, l = 1, b = 0.1$  as parameters of the pendulum dynamics. According to Theorem 5.2, we solve the Lyapunov equation

$$(A - KC)^T P + P(A - KC) = -I$$

and  $\gamma = \frac{0.5}{\lambda_{\max}(P)}$  will be our best estimate of the convergence rate (of  $\|e\| = \|\hat{x} - x\|$  towards zero).

Table 5.1 below shows the convergence rates computed for different values of  $k$ . We can see that as  $k$  is increased, the convergence rate estimation is also increased. However, it appears that 0.1 is the best convergence rate we can achieve, regardless of how large  $k$  is.

Table 5.1: Convergence rate estimation of the Luenberger observer for a simple pendulum.

$k$	0.1	1	10	100	1000	10000
$\gamma$	0.0019	0.0523	0.0990	0.1000	0.1000	0.1000

**Optimal  $K$ .** Is it true that 0.1 is the best convergence rate, or in other words, what is the best  $K$  that maximizes the convergence rate  $\gamma$ ?

A natural way (and my favorite way) to answer this question is to formulate an optimization problem.

$$\begin{aligned} \min_{P, K} \quad & \lambda_{\max}(P) \\ \text{subject to} \quad & (A - KC)^T P + P(A - KC) = -I \\ & P \succeq 0 \end{aligned} \tag{5.16}$$

The above formulation seeks the best possible  $K$  that minimizes  $\lambda_{\max}(P)$  which, according to  $\gamma = 0.5/\lambda_{\max}(P)$ , also maximizes  $\gamma$ .

However, problem (5.16) is not a convex formulation due to the bilinear term  $PK$ . Nevertheless, via a simple change of variable  $H = PK$ , we arrive at the following convex formulation

$$\begin{aligned} \min_{P, H} \quad & \lambda_{\max}(P) \\ \text{subject to} \quad & A^T P - C^T H^T + PA - HC = -I \\ & P \succeq 0 \end{aligned} \tag{5.17}$$

Problem (5.17) is a semidefinite programming problem (SDP), that can be modeled and solved by off-the-shelf tools. We can recover  $K = P^{-1}H$  from (5.17) after it is solved.



Interestingly, solving problem (5.17) verifies that the minimum  $\lambda_{\max}(P)$  is 5 and the maximum converge rate is 0.1. An optimal solution of (5.17) is

$$P^* = \begin{bmatrix} 2.4923 & 0 \\ 0 & 5 \end{bmatrix}, \quad K^* = \begin{bmatrix} 0.2006 \\ 0.4985 \end{bmatrix}.$$

You should check out the Matlab code of this example here.

### 5.1.3 State-affine Template

Consider an instance of the normal form (5.7) where the dynamics is linear in  $\xi$ , but the coefficients are time-varying and dependent on the input and output

$$\dot{\xi} = A(u, y)\xi + B(u, y), \quad y = C(u)\xi. \quad (5.18)$$

The difference between the state-affine template (5.18) and the Luenberger template (5.10) is that the linear matrices  $A, C$  are allowed to depend nonlinearly on the input  $(u, y)$ .

Kalman and Bucy originally proposed an observer for linear time-varying systems (?). The result is later extened by (?) and (?) to deal with coefficient matrices dependent on the control. The following theorem is a direct extension of the result from (?) and (?) by considering  $(u, y)$  as an augmented control input.

Before presenting the theorem, we need to introduce the following terminology.

**Definition 5.3** (Linear Time-Varying System). For a linear time-varying system of the form

$$\dot{\chi} = A(\nu)\chi, \quad y = C(\nu)\chi, \quad (5.19)$$

with input  $\nu$  and output  $y$ , we define

- the *transition matrix*  $\Psi_\nu$  as the unique solution to

$$\Psi_\nu(t, t) = I, \quad \frac{\partial \Psi_\nu}{\partial \tau}(\tau, t) = A(\nu(\tau))\Psi_\nu(\tau, t).$$

Note that the transition matrix is used to express the solution to (5.19) because it satisfies

$$\chi(\chi_0, t_0; t; \nu) = \Psi_\nu(t, t_0)\chi_0.$$

- the *observability grammian* as

$$\Gamma_\nu(t_0, t_1) = \int_{t_0}^{t_1} \Psi_\nu(\tau, t_0)^T C(\nu(\tau))^T C(\nu(\tau)) \Psi_\nu(\tau, t_0) d\tau.$$

- the *backward observability grammian* as

$$\Gamma_\nu^b(t_0, t_1) = \int_{t_0}^{t_1} \Psi_\nu(\tau, t_1)^T C(\nu(\tau))^T C(\nu(\tau)) \Psi_\nu(\tau, t_1) d\tau.$$

We now introduce the Kalman-Bucy Observer for the state-affine template (5.18).

**Theorem 5.3** (Kalman-Bucy Observer). *Let  $y_{\xi_0, u}(t) = C(u(t))\Xi(\xi_0; t; u)$  be the output of system (5.18) at time  $t$  with initialization  $\xi_0$  and control  $u$ . Suppose the control  $u$  satisfies*

- *For any  $\xi_0$ ,  $t \mapsto A(u(t), y_{\xi_0, u}(t))$  is bounded by  $A_{\max}$*
- *For any  $\xi_0$ , the augmented input  $\nu = (u, y_{\xi_0, u})$  is regularly persistent for the dynamics*

$$\dot{\chi} = A(\nu)\chi, \quad y = C(\nu)\chi \quad (5.20)$$

*uniformly with respect to  $\xi_0$ . That is, there exist strictly positive numbers  $t_0, \bar{t}$ , and  $\alpha$  such that for any  $\xi_0$  and any time  $t \geq t_0 \geq \bar{t}$ ,*

$$\Gamma_v^b(t - \bar{t}, t) \succeq \alpha I,$$

*where  $\Gamma_v^b$  is the backward observability grammian associated with system (5.20).*

*Then, given any positive definite matrix  $P_0$ , there exist  $\alpha_1, \alpha_2 > 0$  such that for any  $\lambda \geq 2A_{\max}$  and any  $\xi_0 \in \mathbb{R}^p$ , the matrix differential equation*

$$\dot{P} = -\lambda P - A(u, y)^T P - P A(u, y) + C(u)^T C(u) \quad (5.21)$$

*initialized at  $P(0) = P_0$  admits a unique solution satisfying  $P(t) = P(t)^T$  and*

$$\alpha_2 I \succeq P(t) \succeq \alpha_1 I.$$

*Moreover, the system*

$$\dot{\hat{\xi}} = A(u, y)\hat{\xi} + B(u, y) + K(y - C(u)\hat{\xi}) \quad (5.22)$$

*with a time-varying gain matrix*

$$K = P^{-1}C(u)^T \quad (5.23)$$

*is an observer for the state-affine system (5.18).*

Let us work out an example of the Kalman-Bucy Observer for nonlinear systems.

**Example 5.2** (Kalman-Bucy Observer for A Simple Pendulum). Let us reconsider the pendulum dynamics (5.14) but this time try to design a Kalman-Bucy observer.

We first write the pendulum dynamics in a new coordinate system so that it is in the state-affine normal form (5.18). We choose  $\xi = [\mathfrak{s}, \mathfrak{c}, \dot{\theta}]^T$  with  $\mathfrak{s} = \sin \theta$  and

$\mathbf{c} = \cos \theta$ . Clearly, we will be able to observe  $y = [\mathbf{s}, \mathbf{c}]^T$  in this new coordinate. The state-affine normal form of the pendulum dynamics reads

$$\begin{aligned} \dot{\xi} &= \begin{bmatrix} \mathbf{c}\dot{\theta} \\ -\mathbf{s}\dot{\theta} \\ -\frac{1}{ml^2}(b\dot{\theta} + mgl\mathbf{s}) + \frac{1}{ml^2}u \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & \mathbf{c} \\ 0 & 0 & -\mathbf{s} \\ 0 & 0 & -\frac{b}{ml^2} \end{bmatrix}}_{=:A(u,y)} \xi + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \frac{u-mgl\mathbf{s}}{ml^2} \end{bmatrix}}_{=:B(u,y)} \\ y &= \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{=:C(u)} \xi \end{aligned} \quad (5.24)$$

Note that  $C(u)$  is in fact time-invariant, and  $B(u, y)$  only depends on  $u$ ; but we adopt the same notation as the general state-affine template (5.18).

In order to use the Kalman-Bucy observer in Theorem 5.3, we need to verify the boundedness of  $A(u, y)$ , and the regular persistence of (5.20).

**Boundedness of  $A(u, y)$ .** We can easily show the boundedness of  $A(u, y)$  by writing

$$\|A(u, y)\xi\| = \|\xi_3(\mathbf{c} - \mathbf{s} - b/ml^2)\| \leq |\xi_3|\sqrt{3}\sqrt{\mathbf{c}^2 + \mathbf{s}^2 + b^2/m^2l^4} \leq \|\xi\|\sqrt{3 + 3b^2/m^2l^4}.$$

Therefore, we can take  $A_{\max} = \sqrt{3 + 3b^2/m^2l^4}$ .

**Regular persistence.** We write the backward observability grammian of system (5.20)

$$\Gamma_{\nu}^b(t - \bar{t}, t) = \int_{t-\bar{t}}^t \Psi_{\nu}(\tau, t)^T C^T C \Psi_{\nu}(\tau, t) d\tau = \int_{t-\bar{t}}^t \Psi_{\nu}(\tau, t)^T \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{=: \tilde{C}} \Psi_{\nu}(\tau, t) d\tau.$$

$\Gamma_{\nu}^b(t - \bar{t}, t) \succeq \alpha I$  if and only if

$$w^T \Gamma_{\nu}^b(t - \bar{t}, t) w \geq \alpha, \quad \forall w \in \mathbb{R}^3, \|w\| = 1.$$

With this, we develop  $w^T \Gamma_{\nu}^b(t - \bar{t}, t) w$

$$\begin{aligned} w^T \Gamma_{\nu}^b(t - \bar{t}, t) w &= \int_{t-\bar{t}}^t s^T \tilde{C} s d\tau, \\ &= \int_{t-\bar{t}}^t (s_1^2 + s_2^2) d\tau, \quad s = \Psi_{\nu}(\tau, t) w \end{aligned} \quad (5.25)$$

and observe that  $s = \Psi_{\nu}(\tau, t) w$  is equivalent to

$$w = (\Psi_{\nu}(\tau, t))^{-1} s = \Psi_{\nu}(t, \tau) s,$$

that is,  $w$  is the solution of  $\dot{\xi} = A(\nu)\xi$  at time  $t$  with initial condition  $s$  at time  $\tau \leq t$ . Equivalently, this is saying  $s$  is the initial condition of  $\dot{\xi} = A(\nu)\xi$  at time

$\tau \leq t$  such that its solution at time  $t$  is  $w$ . Note that from (5.25) it is clearly that  $\int_{t-\bar{t}}^t (s_1^2 + s_2^2) d\tau \geq 0$ , and  $\int_{t-\bar{t}}^t (s_1^2 + s_2^2) d\tau = 0$  if and only if  $s_1^2 + s_2^2 = 0$ , or equivalently  $s_1 = s_2 = 0$  for any  $\tau \in [t - \bar{t}, t]$ .

We then take a closer look at the system  $\dot{\xi} = A(\nu)\xi$ :

$$\begin{aligned}\dot{\xi}_1 &= \mathfrak{c}\xi_3 \\ \dot{\xi}_2 &= -\mathfrak{s}\xi_3 \\ \dot{\xi}_3 &= -\frac{b}{ml^2}\xi_3.\end{aligned}\tag{5.26}$$

If the solution of  $\xi_3$  at time  $t$  is  $w_3$ , then

$$\xi_3(\tau) = w_3 e^{\frac{b}{ml^2}(t-\tau)}, \quad \tau \leq t.$$

We can now claim it is impossible that  $s_1 = s_2 = 0$  at any time  $\tau \in [t - \bar{t}, t]$ .

We can show this by contradiction. First of all,  $s_1 = s_2 = 0$  at  $\tau = t$  implies  $w_1 = w_2 = 0$  and hence  $w_3 = \pm 1$ . This implies  $\xi_3 \neq 0$  for any  $\tau \in [t - \bar{t}, t]$ . Then,  $s_1 = 0, \forall \tau \in [t - \bar{t}, t]$  implies  $\dot{\xi}_1 = 0$  which, due to  $\xi_3 \neq 0$ , implies  $\mathfrak{c} = 0$  for all  $\tau$ . Similarly,  $s_2 = 0, \forall \tau \in [t - \bar{t}, t]$  implies  $\dot{\xi}_2 = 0$  and  $\mathfrak{s} = 0$ . This creates a contradiction because  $\mathfrak{c}^2 + \mathfrak{s}^2 = 1$  and  $\mathfrak{c}, \mathfrak{s}$  cannot be simultaneously zero.

The above reasoning proves that the backward observability Grammian is positive definite, which is, however, still insufficient for the Kalman-Bucy observer. We need a stronger uniformly positive definite condition on  $\Gamma_\nu^b$ , i.e., to find  $t_0, \bar{t}$  and  $\alpha > 0$  so that  $\Gamma_\nu^b(t - \bar{t}, t) \succeq \alpha I$  for all  $t \geq t_0$ .

If the control  $u$  is unbounded, then sadly, one can show that the uniform positive definite condition fails to hold, as left by you to show in the following exercise.

**Exercise 5.1** (Counterexample for Kalman-Bucy Observer). Show that, if the control  $u$  is unbounded, then for any  $\alpha > 0, t_0 \geq \bar{t} > 0$ , there exists  $t \geq t_0$  such that  $\Gamma_\nu^b(t - \bar{t}, t) \prec \alpha I$ . (Hint: consider a controller that spins the pendulum faster and faster such that in time  $\bar{t}$  it has rotated  $2k\pi$ , in this case the angular velocity becomes unobservable because we are not sure how many rounds the pendulum has rotated.)

Fortunately, if the control  $u$  is bounded, then we can prove the uniform positive definite condition holds for  $\Gamma_\nu^b(t - \bar{t}, t)$ . The following proof is given by Weiyu Li.

Without loss of generality, let  $\frac{b}{ml^2} = 1$ . Assume  $u$  is bounded such that the third entry of  $B(u, y)$  in (5.24) is bounded by  $\beta > 0$

$$\left| \frac{u - mgl\mathfrak{s}}{ml^2} \right| \leq \beta.$$

Assuming the initial velocity of the pendulum is  $\dot{\theta}(0) = \dot{\theta}_0$ , we know  $\dot{\theta}(t)$  is bounded by

$$\dot{\theta}(t) \in [c_1(1 - \beta)e^{-t} - \beta, c_2(1 - \beta)e^{-t} + \beta],$$

where  $c_1, c_2$  are constants chosen to satisfy the initial condition. Clearly, for all  $t > 0$ , we see  $\dot{\theta}(t)$  is bounded, and hence we know  $\dot{\mathbf{c}}$  and  $\dot{\mathbf{s}}$  are bounded (due to  $\mathbf{c}$  and  $\mathbf{s}$  are bounded). Intuitively, what we have just shown says that when the control is bounded, the measurements  $\mathbf{c}$  and  $\mathbf{s}$  will have bounded time derivatives. (This will help us analyze the auxiliary system (5.26).)

Now back to checking regular persistence of the auxiliary system (5.26). We will discuss two cases: (1)  $w_3^2 > 1 - \delta$ , and (2)  $w_3^2 \leq 1 - \delta$ , for some constant  $\delta < 0.5$  determined later.

1.  $w_3^2 > 1 - \delta > 0.5$ . In this case we have  $w_1^2 + w_2^2 = 1 - w_3^2 < \delta$ , and hence  $w_1^2 < \delta$ ,  $w_2^2 < \delta$ . On the other hand, from (5.26) we have

$$\dot{\xi}_1^2(\tau) + \dot{\xi}_2^2(\tau) = \xi_3^2 = w_3^2 e^{2(t-\tau)} > w_3^2 > 1 - \delta, \quad \forall \tau < t.$$

Without loss of generality assume  $\dot{\xi}_1(t)^2 > (1 - \delta)/2$ . As  $\dot{\xi}_1 = \mathbf{c}\xi_3$  and both  $\mathbf{c}$  and  $\xi_3$  have bounded derivatives, we know  $\xi_1$  will not change sign for some duration  $T$  that is independent from the choice of  $\delta$  (because the time derivatives of  $\mathbf{c}$  and  $\xi_3$  do not depend on  $\delta$ ). That is  $|\dot{\xi}_1| > \sqrt{(1 - \delta)/2} > 1/2$  for  $\tau \in [t - T, t]$ . Consequently,

$$|\xi_1(t - \tau)| > \frac{1}{2}\tau - |w_1| > \frac{1}{2}\tau - \sqrt{\delta}, \quad \tau \in [0, T].$$

Choosing  $\delta$  small enough, we have  $|\xi_1(t - \tau)| > 0.25\tau$  for  $\tau \in [0.5T, T]$ . Then we have

$$\Gamma_\nu^b(t - T, t) \succ [(0.25 \times 0.5T)^2 \times 0.5T]I.$$

2.  $w_3^2 \leq 1 - \delta$ . In this case  $w_1^2 + w_2^2 = 1 - w_3^2 \geq \delta$ , and at least one of  $w_1$  and  $w_2$  has absolute value larger than  $\sqrt{\delta}/2$ . Because the derivatives of  $\xi_1$  and  $\xi_2$  are both bounded, we know  $\xi_1$  and  $\xi_2$  will remain large for some constant time. Thus there is a uniform lower bound.

The intuition of the above proof is simple: when  $\xi_1$  and/or  $\xi_2$  already have large absolute value (case 2), we can find a time window such that  $\xi_1$  and/or  $\xi_2$  remain large in that time window; when  $\xi_1$  and/or  $\xi_2$  are small (case 1), using the observation that their time derivatives are large (because  $w_3$  is large), together with the fact that these derivatives remain large (because the derivative of these derivatives are bounded), we can also find a time window that  $\xi_1$  and/or  $\xi_2$  are large (back in time). Therefore, the backward observability Grammian is uniformly positive definite.

#### 5.1.4 Kazantzis-Kravaris-Luenberger (KKL) Template

In Luenberger's original paper about observer design for linear systems (?), the goal was to transform a linear system

$$\dot{x} = Fx, \quad y = Cx$$

into a Hurwitz form

$$\dot{\xi} = A\xi + By \quad (5.27)$$

with  $A$  a Hurwitz (stable) matrix. If such a transformation is available, then the following system

$$\dot{\hat{\xi}} = A\hat{\xi} + By,$$

which is nothing but a copy of the dynamics (5.27), is in fact an observer. This is because the error  $e = \hat{\xi} - \xi$  evolves as

$$\dot{e} = Ae,$$

which implies that  $e$  tends to zero regardless of the initial error  $e(0)$ . Luenberger proved that when  $(F, C)$  is observable, a stationary transformation  $\xi = Tx$  with  $p = n$ , i.e.,  $T \in \mathbb{R}^{n \times n}$ , always exists and is unique, for any matrix  $A$  that is Hurwitz and  $(A, B)$  that is controllable. This is based on the fact that

$$(AT + BC)x = A\xi + By = \dot{\xi} = T\dot{x} = TFx, \forall x \quad (5.28)$$

$$\iff AT + BC = TF, \quad (5.29)$$

known as the Sylvester equation, admits a unique and invertible solution  $T$ .

A natural extension of Luenberger's original idea is to find a transformation that converts the nonlinear system (5.1) into the following form

$$\dot{\xi} = A\xi + B(u, y), \quad y = H(\xi, u), \quad (5.30)$$

with  $A$  a Hurwitz matrix (but  $H$  can be nonlinear, as opposed to the Luenberger template in Theorem 5.2). If such a transformation can be found, then we can design a similar observer that copies the dynamics (5.30)

$$\dot{\hat{\xi}} = A\hat{\xi} + B(u, y). \quad (5.31)$$

We refer to such a nonlinear Luenberger template the Kazantzis-Kravaris-Luenberger (KKL) template, due to the seminal work (?).

The KKL template, once found, is nice in the sense that (i) the observer (5.31) is a simple copy of the dynamics and also very easy to implement (as opposed to the Kalman-Bucy observer); and (ii) checking if the matrix  $A$  is Hurwitz is easy, at least when  $A$  has reasonable size, (e.g., compared to checking the regular persistence condition in the state-affine template in Theorem 5.3).

However, the KKL template is difficult to realize in the sense that (i) what kind of nonlinear systems can be converted to (5.30), and (ii) for those systems, how do we find the coordinate transformation?

Recent works have leveraged deep learning to learn the coordinate transformation, for example in (?), (?), (?). Before hammering the problem with deep learning, let us look at the fundamentals of the KKL observer.

### 5.1.4.1 Autonomous Systems

Consider the autonomous version of system (5.1) without control

$$\dot{x} = f(x), \quad y = h(x), \quad (5.32)$$

where  $x \in \mathbb{X} \subseteq \mathbb{R}^n, y \in \mathbb{Y} \subseteq \mathbb{R}^d$ .

The following result, established by (?), states that the KKL observer exists under mild conditions.

**Theorem 5.4** (KKL Observer for Autonomous Systems). *Assum  $\mathcal{X}$  and  $\mathcal{L}$  are open bounded sets in  $\mathbb{X}$  (the state space) such that  $\text{cl}(\mathcal{X})$  is contained in  $\mathcal{L}$  and the system (5.32) is backward  $\mathcal{L}$ -distinguishable on  $\mathcal{X}$  (cf. Definition 5.2). Then there exists a strictly positive number  $\gamma$  and a set  $\mathcal{S}$  of zero Lebesgue measure in  $\mathbb{C}^{n+1}$  such that denoting  $\Omega = \{\lambda \in \mathbb{C} \mid \text{Re}(\lambda) < -\gamma\}$ , for any  $(\lambda_1, \dots, \lambda_{n+1}) \in \Omega^{n+1} \setminus \mathcal{S}$ , there exists a function  $T : \mathbb{R}^n \rightarrow \mathbb{R}^{(n+1)d}$  uniformly injective on  $\mathcal{X}$  satisfying*

$$L_f T(x) = AT(x) + B(h(x))$$

with

$$A = \tilde{A} \otimes I_d, \quad B(y) = (\tilde{B} \otimes I_d)y \quad (5.33)$$

$$\tilde{A} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{n+1} \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (5.34)$$

Moreover, if  $\mathcal{X}$  is backward invariant, then  $T$  is unique and defined by

$$T(x) = \int_{-\infty}^0 e^{-A\tau} B(h(X(x, \tau))) d\tau. \quad (5.35)$$

*Remark.* The function  $T$  in Theorem 5.4 takes complex numbers. To simulate the observer

$$\dot{\hat{\xi}} = A\hat{\xi} + B(y),$$

one needs to implement in real numbers, for each  $\lambda_i$  and  $j \in [d]$

$$\dot{\hat{\xi}}_{\lambda_i, j} = \begin{bmatrix} -\text{Re}(\lambda_i) & -\text{Im}(\lambda_i) \\ \text{Im}(\lambda_i) & -\text{Re}(\lambda_i) \end{bmatrix} \hat{\xi}_{\lambda_i, j} + \begin{bmatrix} y_j \\ 0 \end{bmatrix}.$$

Therefore, the dimension of the observer is  $2 \times d(n+1)$ .

Theorem 5.4 states that as long as the system (5.32) is backward distinguishable, then there exists a stationary transformation  $T$  that can transform the system to a new coordinate system  $\xi$  such that the dynamics in  $\xi$  is Hurwitz. A closer look at the structure of  $A$  and  $B$  reveals that the coordinate transformation needs to satisfy  $n+1$  differential equations of the form

$$\frac{\partial T_\lambda}{\partial x}(x) \dot{x} = \lambda T_\lambda(x) + y$$

where each  $T_\lambda$  transforms the state  $x$  into a new coordinate having the same dimension of  $y$ . Clearly, if  $T = (T_\lambda)$ , i.e., there is a single  $\lambda$ , then  $T$  is not uniformly injective (as the dimension of  $\xi$  is  $d < n$ ). Consequently, by choosing

$$T = (T_{\lambda_1}, \dots, T_{\lambda_{n+1}}),$$

the uniform injectivity of  $T$  is ensured.

However, the difficulty lies in the computation of  $T$  (and  $T_\lambda$ ), let alone its inverse (that recovers  $x$  from  $\xi$ ). Even though  $\mathcal{X}$  is backward invariant, the formulation (5.35) is difficult to compute. I tried very hard to find a coordinate transformation  $T$  that can convert the non-controlled pendulum dynamics into the KKL form but did not succeed. **You should let me know if you were able to find one!** Nevertheless, the following example shows you the flavor of how such a transformation may look like for a different system.

**Example 5.3** (KKL Observer for an Oscillator with Unknown Frequency). Consider a harmonic oscillator with unknown frequency

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -x_1 x_3 & , \quad y = x_1 \\ \dot{x}_3 = 0 \end{cases}$$

Consider the coordinate transformation

$$T_{\lambda_i}(x) = \frac{\lambda_i x_1 - x_2}{\lambda_i^2 + x_3}, \quad \lambda_i > 0, i = 1, \dots, p.$$

We have

$$\frac{\partial T_{\lambda_i}(x)}{\partial x} \dot{x} = \left\langle \begin{bmatrix} \frac{\lambda_i}{\lambda_i^2 + x_3} \\ -1 \\ \frac{\lambda_i^2}{\lambda_i^2 + x_3} \\ \frac{x_2 - \lambda_i x_1}{(\lambda_i^2 + x_3)^2} \end{bmatrix}, \begin{bmatrix} x_2 \\ -x_1 x_3 \\ 0 \end{bmatrix} \right\rangle = \frac{\lambda_i x_2 + x_1 x_3}{\lambda_i^2 + x_3} \quad (5.36)$$

$$-\lambda_i T_{\lambda_i}(x) + y = \frac{-\lambda_i^2 x_1 + \lambda_i x_2 + x_1 \lambda_i^2 + x_1 x_3}{\lambda_i^2 + x_3} = \frac{\lambda_i x_2 + x_1 x_3}{\lambda_i^2 + x_3} \quad (5.37)$$

Therefore, with

$$\xi = T(x) = [T_{\lambda_1}(x), T_{\lambda_2}(x), \dots, T_{\lambda_p}(x)]^T,$$

we have

$$\dot{\xi} = \underbrace{\begin{bmatrix} -\lambda_1 & & \\ & \ddots & \\ & & -\lambda_p \end{bmatrix}}_A \xi + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} y$$

with  $A$  clearly Hurwitz.



With some extra arguments (cf. Section 8.1.1 in (?)), one can see that the transformation  $T$  is injective with  $p \geq 4$  distinct  $\lambda_i$ 's. Therefore, this is a valid KKL observer.

The final issue that one needs to think about is, since the observer is estimating  $\hat{\xi}$ , how to recover  $\hat{x}$ ? In this example, there is actually no analytical formula for recovering  $\hat{x}$  from  $\hat{\xi}$ . In this case, one approach is to solve the following optimization problem

$$\hat{x} = \arg \min_x \|\hat{\xi} - T(x)\|^2,$$

which may be quite expensive.

A more general treatment is given in Section 8.2.2 in (?).

#### 5.1.4.2 Controlled Systems

### 5.1.5 Triangular Template

### 5.1.6 Design with Convex Optimization

Consider a nonlinear system

$$\dot{x} = f(x) + \psi(u, y), \quad y = Cx \quad (5.38)$$

where  $x \in \mathbb{X} \subseteq \mathbb{R}^n$ ,  $y \in \mathbb{R}^d$ ,  $C$  a constant matrix, and  $\psi(u, y)$  a nonlinear function. We assume that  $f(x)$  is a polynomial vector map (i.e., each entry of  $f$  is a polynomial function in  $x$ ). Certainly the formulation in (5.38) is not as general as (5.1), but it is general enough to include many examples in robotics.

Recall that I said the essence of observer design is to (i) simulate the dynamics when the state estimation is correct, and (ii) to correct the state estimation from observation when it is off. Therefore, we wish to design an observer for (5.38) in the following form

$$\dot{\hat{x}} = \underbrace{f(\hat{x}) + \psi(u, y)}_{\text{dynamics simulation}} + \underbrace{K(y - \hat{y}, y)(C\hat{x} - y)}_{\text{feedback correction}}, \quad (5.39)$$

where, compared to the Luenberger observer (5.11), we allow the gain matrix  $K$  to be nonlinear functions of the true observation  $y$  and the estimated observation  $\hat{y}$ .

With the observer (5.39), the dynamics on the estimation error  $e = \hat{x} - x$  becomes

$$\dot{e} = f(x + e) - f(x) + K(Ce, Cx)Ce.$$

If we can find a Lyapunov-like function  $V(e)$  so that  $V(e)$  is positive definite and  $\dot{V}(e)$  is negative definite, then Lyapunov stability theorem 4.2 tells us that  $e = 0$  is asymptotically stable. Because we do not know the gain matrix  $K$  either, we

need to jointly search for  $V$  and  $K$  (that are polynomials). Mathematically, this is

$$\begin{aligned}
& \text{find } V, K \\
& \text{subject to } V(0) = 0, \quad V(e) > 0, \forall e \neq 0 \\
& \quad \dot{V}(e) = \frac{\partial V}{\partial e} (f(x+e) - f(x) + K(Ce, Cx)Ce) < 0, \forall e \neq 0, \forall x \in \mathbb{X} \\
& \quad V(e) \geq \epsilon \|e\|^2, \forall e
\end{aligned} \tag{5.40}$$

where the last constraint is added to make sure  $V(e)$  is radially unbounded. Furthermore, if we replace the second constraint by  $\dot{V}(e) \leq -\lambda V(e)$ , then we can guarantee  $V(e)$  converges to zero exponentially.

Problem (5.40), however, is not a convex optimization problem, due to the term  $\frac{\partial V}{\partial e} K$  being bilinear in the coefficients of  $V$  and  $K$ . Nevertheless, as shown in (?), we can use a reparameterization trick to formulate a stronger version of (5.40) as follows.

$$\begin{aligned}
& \text{find } V, Q(Ce), M(Ce, Cx) \\
& \text{subject to } V(0) = 0, \quad V(e) > 0, \forall e \neq 0 \\
& \quad \frac{\partial V}{\partial e} = e^T Q(Ce), \quad Q(Ce) \succ 0 \\
& \quad e^T Q(Ce) (f(x+e) - f(x)) + e^T M(Ce, Cx)Ce < 0, \forall e \neq 0, \forall x \in \mathbb{X} \\
& \quad V(e) \geq \epsilon \|e\|^2, \forall e
\end{aligned} \tag{5.41}$$

Clearly, if we can solve problem (5.41), then

$$K = Q(Ce)^{-1} M(Ce, Cx)$$

is the right gain matrix for the formulation (5.40).

Let us bring this idea to action in our pendulum example.

**Example 5.4** (Pendulum Observer with Convex Optimization). With  $x = [\mathfrak{s}, \mathfrak{c}, \dot{\theta}]^T$  ( $\mathfrak{s} = \sin \theta$ ,  $\mathfrak{c} = \cos \theta$ ), we can write the pendulum dynamics as

$$\dot{x} = \underbrace{\begin{bmatrix} \mathfrak{c}\dot{\theta} \\ -\mathfrak{s}\dot{\theta} \\ -\frac{b}{ml^2}\dot{\theta} \end{bmatrix}}_{=:f(x)} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \frac{u-mgls}{ml^2} \end{bmatrix}}_{=: \psi(u,y)}, \quad y = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{=:C} x$$

Clearly  $f(x)$  is a polynomial. Solving the convex optimization problem (5.41), we obtain a solution

$$V(e) = 0.5954e_1^2 + 0.5954e_2^2 + 0.9431e_3^2$$

$$Q(Ce) = \begin{bmatrix} 0.4603e_2^2 + 1.1909 & -0.4603e_1e_2 & 0 \\ -0.4603e_1e_2 & 0.4603e_1^2 + 1.1909 & 0 \\ 0 & 0 & 1.8863 \end{bmatrix}$$

$$M(Ce, Cx) = \begin{bmatrix} -2.0878e_1^2 - 0.8667e_2^2 - 0.4588(y_1^2 + y_2^2) - 0.4885 & -0.8667e_1e_2 \\ -0.8667e_1e_2 & -0.8667e_1^2 - 2.0878e_2^2 - 0.4588(y_1^2 + y_2^2) - 0.4885 \\ -1.1909y_2 & 1.1909y_1 \end{bmatrix}$$

Simulating this observer, we verify that the observer is in fact exponentially converging, as shown in Fig. 5.1.

The Matlab code for formulating and solving the convex optimization (5.41) can be found [here](#). The code for simulating the observer can be found [here](#).

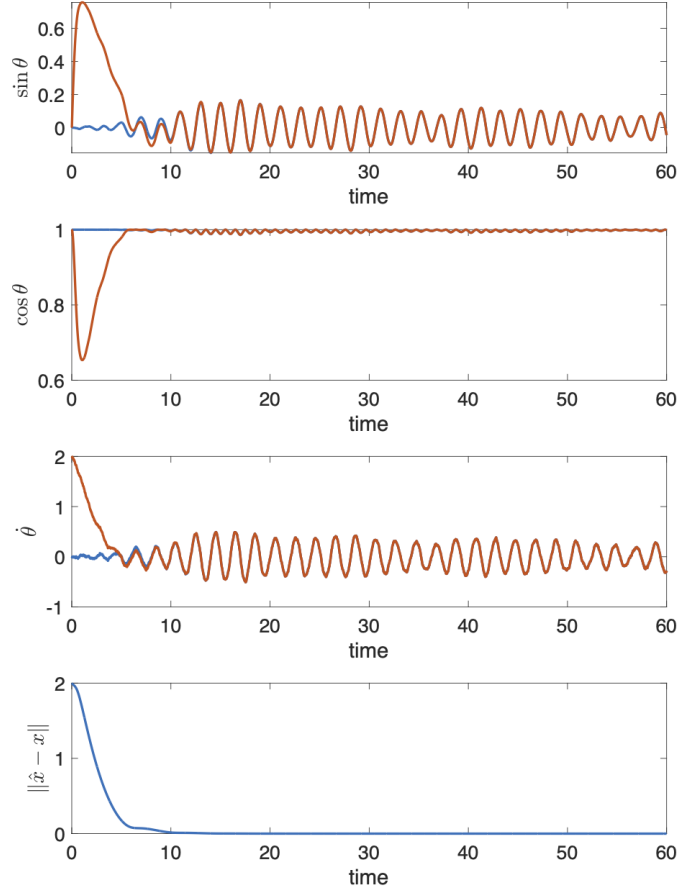


Figure 5.1: Simulation of the pendulum observer design from convex optimization

## 5.2 Observer Feedback

Now that we have good ways to design a state observer, we will see how we can use the observer for feedback control.

**Example 5.5** (Pendulum Stabilization with A Luenberger Observer). In Example 5.1, we have written the dynamics of a pendulum, and the dynamics of a Luenberger observer as

$$\dot{x} = Ax + B(u, y) \quad (5.42)$$

$$\dot{\hat{x}} = A\hat{x} + B(u, y) + KC(x - \hat{x}) \quad (5.43)$$

We wish to understand (so we can optimize) the behavior of this system under certain control input  $u$ . To do so, let us denote  $e = \hat{x} - x$ , and write the above dynamics as

$$\dot{x} = Ax + B(u, Cx) \quad (5.44)$$

$$\dot{e} = (A - KC)e \quad (5.45)$$

Denoting  $z = [x, e]^T$ , we have the augmented dynamics

$$\dot{z} = \underbrace{\begin{bmatrix} A & 0 \\ 0 & A - KC \end{bmatrix}}_{=:F} z + \underbrace{\begin{bmatrix} B(u, Dz) \\ 0 \end{bmatrix}}_{=:G(z, u)}$$

We want to stabilize the system at  $z_0 = [\pi, 0, 0, 0]^T$  (the upright position) subject to control bounds  $u \in \mathbb{U} = [-u_{\max}, u_{\max}]$ .

We need to find a control Lyapunov function (CLF),  $V(z)$ , that satisfies the following constraints:

$$\begin{aligned} V(z_0) &= 0 \\ V(z) &> 0 \quad \forall z \in \{z : V(z) < \rho, z \neq z_0\} \\ \inf_{u \in \mathbb{U}} [L_F V(z) + L_G V(z)u] &\leq 0 \quad \forall z \in \mathcal{Z} \end{aligned}$$

where  $L_F V$  and  $L_G V$  are the Lie derivatives of  $V$  along  $F$  and  $G(z, u)$ , respectively.  $\mathcal{Z}$  is the set of all possible augmented states. The CLF will define the set of admissible control inputs  $U$ .

$$U = \{u : L_f V(z) + L_g V(z)u \leq 0\}$$

To find the smallest-magnitude control input such that  $u \in K$ , we may use a quadratic program:

$$\begin{aligned} \min_{u \in \mathcal{U}} & \|u\|^2 \\ \text{s.t.} \quad & L_f V(z) + L_g V(z)u \leq -cV(z) \end{aligned}$$

where  $c$  is some positive constant. The challenge now is in choosing a suitable  $V(z)$ .

## Chapter 6

# Adaptive Control

### 6.1 Model-Reference Adaptive Control

Basic flow for designing an adaptive controller

1. Design a control law with variable parameters
2. Design an adaptation law for adjusting the control parameters
3. Analyze the convergence of the closed-loop system

The control law design at the first step typically requires the designer to know what a good controller is if the true parameters were actually known, e.g., from feedback linearization (Appendix D), sliding control (Appendix E) etc.

The design of the adaptation law typically comes from analyzing the dynamics of the tracking error, which as we will see often appears in the form of Lemma 6.1.

The convergence of the closed-loop system is usually analyzed with the help of a Lyapunov-like function introduced in Chapter 4.

**Lemma 6.1** (Basic Lemma). *Let two signals  $e(t)$  and  $\phi(t)$  be related by*

$$e(t) = H(p)[k\phi(t)^T v(t)] \quad (6.1)$$

*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*

$$\dot{\phi}(t) = -\text{sgn}(k)\gamma e(t)v(t) \quad (6.2)$$

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

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

*Proof.* Let the state-space representation of (6.1) be

$$\dot{x} = Ax + b[k\phi^T v], \quad e = c^T x. \quad (6.3)$$

Since  $H(p)$  is SPR, it follows from the Kalman-Yakubovich Lemma C.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 (6.3) with  $\phi$  chosen as in (6.2) is

$$\dot{V} = \frac{\partial V}{\partial x} \dot{x} + \frac{\partial V}{\partial \phi} \dot{\phi} \quad (6.4)$$

$$= x^T (PA + A^T P)x + 2x^T Pb(k\phi^T v) + \frac{2|k|}{\gamma} \phi^T (-\text{sgn}(k)\gamma ev) \quad (6.5)$$

$$= -x^T Qx + 2(x^T c)(k\phi^T v) - 2\phi^T (ev) \quad (6.6)$$

$$= -x^T Qx \leq 0. \quad (6.7)$$

As a result, we know  $x$  and  $\phi$  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 (6.3). Because  $\dot{V} = -2x^T Qx$  is now bounded, we know  $\dot{V}$  is uniformly continuous. Therefore, by Barbalat's stability certificate (Theorem 4.3), we know  $\dot{V}$  tends to zero as  $t$  tends to infinity, which implies  $\lim_{t \rightarrow \infty} x(t) = 0$  and hence  $\lim_{t \rightarrow \infty} e(t) = 0$ .  $\square$

### 6.1.1 First-Order Systems

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

$$\dot{x} = -ax + bu \quad (6.8)$$

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

$$\dot{x}_d = -a_d x_d + b_d r(t), \quad (6.9)$$

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

$$u = \hat{a}_r(t)r + \hat{a}_x(t)x \quad (6.10)$$

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 (6.8) with the controller (6.10) 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 (6.10) is clear: if the system parameters  $(a, b)$  were known, then choosing

$$a_r^* = \frac{b_d}{b}, \quad a_x^* = \frac{a - a_d}{b} \quad (6.11)$$

leads to the closed-loop dynamics  $\dot{x} = -a_d x + b_d r$  that is exactly what we want in (6.9).

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

$$\dot{e} = \dot{x} - \dot{x}_d \quad (6.12)$$

$$= -a_d(x - x_d) + (a_d - a + b\hat{a}_x)x + (b\hat{a}_r - b_d)r \quad (6.13)$$

$$= -a_d e + b \underbrace{(\hat{a}_x - \hat{a}_x^*)}_{=: \tilde{a}_x} x + b \underbrace{(\hat{a}_r - \hat{a}_r^*)}_{=: \tilde{a}_r} r \quad (6.14)$$

$$= -a_d e + b(\tilde{a}_x x + \tilde{a}_r r) \quad (6.15)$$

where  $\tilde{a}_x$  and  $\tilde{a}_r$  are the gain errors w.r.t. the optimal gains in (6.11) if the true parameters were known. The error dynamics (6.15) is equivalent to the following transfer function

$$e = \frac{1}{p + a_d} b(\tilde{a}_x x + \tilde{a}_r r) = \frac{1}{p + a_d} \left( b \begin{bmatrix} \tilde{a}_x \\ \tilde{a}_r \end{bmatrix}^T \begin{bmatrix} x \\ r \end{bmatrix} \right), \quad (6.16)$$

which is in the form of (6.1). Therefore, we choose the adaptation law

$$\begin{bmatrix} \dot{\tilde{a}_x} \\ \dot{\tilde{a}_r} \end{bmatrix} = -\text{sgn}(b)\gamma e \begin{bmatrix} x \\ r \end{bmatrix}. \quad (6.17)$$

**Tracking convergence.** With the control law (6.10) and the adaptation law (6.17), we can prove that the tracking error converges to zero, using Lemma 6.1. With  $\tilde{a} = [\tilde{a}_x, \tilde{a}_r]^T$ , let

$$V(e, \tilde{a}) = e^2 + \frac{|b|}{\gamma} \tilde{a}^T \tilde{a} \quad (6.18)$$

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 (6.9)) and hence  $x$  is bounded (due to  $e = x - x_d$  being bounded). Consequently, from the error dynamics (6.15) we know  $\dot{e}$  is bounded, which implies  $\dot{V} = -4a_d e \dot{e}$  is bounded and  $\dot{V}$  is uniformly continuous. By Barbalat's stability certificate 4.3, we conclude  $e(t) \rightarrow 0$  as  $t \rightarrow \infty$ .

It is always better to combine mathematical analysis with intuitive understanding. Can you explain intuitively why the adaptation law (6.17) makes sense? (Hint: think about how the control should react to a negative/positive tracking error.)

**Parameter convergence.** We have shown the control law (6.10) and the adaptation law (6.17) guarantee to track the reference trajectory. However, is it guaranteed that the gains of the controller (6.10) also converge to the optimal gains in (6.11)?

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 (6.16), 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 (6.17) 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

$$v^T \tilde{a} = 0, \quad v = \begin{bmatrix} x \\ r \end{bmatrix}, \quad (6.19)$$

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

- **Constant reference: no guaranteed convergence.** Suppose  $r(t) \equiv r_0 \neq 0$  for all  $t$ . From (6.9) we know  $x = x_d = \alpha r_0$  when  $t \rightarrow \infty$ , where  $\alpha$  is the constant DC gain of the stable filter. Therefore, the linear equation (6.19) 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

$$\int_t^{t+T} v v^T d\tau \geq \beta I, \quad (6.20)$$



then  $\tilde{a}$  is guaranteed to converge to zero. To see this, we multiply (6.19) 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 (6.20), 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.

**Exercise 6.1** (Extension to Nonlinear Systems). Design a control law and an adaptation law for the following system

$$\dot{x} = -ax - cf(x) + bu$$

with unknown true parameters  $(a, b, c)$  (assume the sign of  $b$  is known) and known nonlinearity  $f(x)$  to track a reference trajectory  $r(t)$ . Analyze the convergence of tracking error and parameter estimation error.

### 6.1.2 High-Order Systems

Consider an  $n$ -th order nonlinear system

$$q^{(n)} + \sum_{i=1}^n \alpha_i f_i(x, t) = bu \quad (6.21)$$

where  $x = [q, \dot{q}, \ddot{q}, \dots, q^{(n-1)}]^T$  is the state of the system,  $f_i$ 's are known nonlinearities,  $(\alpha_1, \dots, \alpha_n, b)$  are unknown parameters of the system (with  $\text{sgn}(b)$  known).

The goal of adaptive control is to control the system (6.21) trajectory to follow a desired trajectory  $q_d(t)$  despite not knowing the true parameters.

To facilitate the derivation of the adaptive controller, let us divide both sides of (6.21) by  $b$

$$hq^{(n)} + \sum_{i=1}^n a_i f_i(x, t) = u \quad (6.22)$$

where  $h = 1/b$  and  $a_i = \alpha_i/b$ .

**Control law.** Recall that the choice of the control law is typically inspired by the control design if the true system parameters were known. We will borrow ideas from sliding control (Appendix E).

- **Known parameters.** Let  $e = q(t) - q_d(t)$  be the tracking error, and define the following combined error

$$s = e^{(n-1)} + \lambda_{n-2}e^{(n-2)} + \cdots + \lambda_0 e = \Delta(p)e$$

where  $\Delta(p) = p^{n-1} + \lambda_{n-2}p^{(n-2)} + \cdots + \lambda_0$  is a stable polynomial with user-chosen coefficients  $\lambda_0, \dots, \lambda_{n-2}$ . The rationale for defining the combined error  $s$  is that the convergence of  $e$  to zero can be guaranteed by the convergence of  $s$  to zero (when  $\Delta(p)$  is stable). Note that  $s$  can be equivalently written as

$$s = (q^{(n-1)} - q_d^{(n-1)}) + \lambda_{n-2}e^{(n-2)} + \cdots + \lambda_0 e \quad (6.23)$$

$$= q^{(n-1)} - \underbrace{\left( q_d^{(n-1)} - \lambda_{n-2}e^{(n-2)} - \cdots - \lambda_0 e \right)}_{q_r^{(n-1)}}. \quad (6.24)$$

Now consider the control law

$$u = hq_r^{(n)} - ks + \sum_{i=1}^n a_i f_i(x, t) \quad (6.25)$$

where

$$q_r^{(n)} = q_d^{(n)} - \lambda_{n-2}e^{(n-1)} - \cdots - \lambda_0 \dot{e}$$

and  $k$  is a design constant that has the same sign as  $h$ . This choice of control, plugged into the system dynamics (6.22), leads to

$$hq^{(n)} + \sum_{i=1}^n a_i f_i(x, t) = hq_r^{(n)} - ks + \sum_{i=1}^n a_i f_i(x, t) \iff \quad (6.26)$$

$$h \left( q^{(n)} - q_r^{(n)} \right) + ks = 0 \iff \quad (6.27)$$

$$h\dot{s} + ks = 0, \quad (6.28)$$

which guarantees the exponential convergence of  $s$  to zero (note that  $h$  and  $k$  have the same sign), and hence the convergence of  $e$  to zero.

- **Unknown parameters.** Inspired by the control law with known parameters in (6.25), we design the adaptive control law as

$$u = \hat{h}q_r^{(n)} - ks + \sum_{i=1}^n \hat{a}_i f_i(x, t), \quad (6.29)$$

where the time-varying gains  $\hat{h}, \hat{a}_1, \dots, \hat{a}_n$  will be adjusted by an adaptation law.

**Adaptation law.** Inserting the adaptive control law (6.29) into the system dynamics (6.22), we obtain

$$h\dot{s} + ks = \tilde{h}q_r^{(n)} + \sum_{i=1}^n \tilde{a}_i f_i(x, t) \iff \quad (6.30)$$

$$s = \frac{1}{p + k/h} \frac{1}{h} \underbrace{\begin{pmatrix} \tilde{h} \\ \tilde{a}_1 \\ \vdots \\ \tilde{a}_n \end{pmatrix}^T \begin{pmatrix} q_r^{(n)} \\ f_1(x, t) \\ \vdots \\ f_n(x, t) \end{pmatrix}}_{=: \phi^T v} \quad (6.31)$$

where  $\tilde{h} = \hat{h} - h$  and  $\tilde{a}_i = \hat{a}_i - a_i, i = 1, \dots, n$ . Again, (6.31) is in the familiar form of (6.1), which naturally leads to the following adaptation law with  $\gamma > 0$  a chosen constant

$$\dot{\phi} = \begin{bmatrix} \dot{\tilde{h}} \\ \dot{\tilde{a}}_1 \\ \vdots \\ \dot{\tilde{a}}_n \end{bmatrix} = -\text{sgn}(h)\gamma s \begin{bmatrix} q_r^{(n)} \\ f_1(x, t) \\ \vdots \\ f_n(x, t) \end{bmatrix}. \quad (6.32)$$

**Tracking and parameter convergence.** With the following Lyapunov function

$$V(s, \phi) = |h|s^2 + \frac{1}{\gamma}\phi^T \phi, \quad \dot{V}(s, \phi) = -2|k|s^2, \quad (6.33)$$

the global convergence of  $s$  to zero can be easily shown. For parameter convergence, it is easy to see that when  $v$  satisfies the persistent excitation condition, we have that  $\phi$  converges to zero. (However, the relationship between the reference trajectory  $q_d(t)$  and the persistent excitation of  $v$  becomes nontrivial due to the nonlinearities  $f_i$ .)

### 6.1.3 Robotic Manipulator

So far our focus has been on systems with a single input ( $u \in \mathbb{R}$ ). In the following, we will show that similar techniques can be applied to adaptive control of systems with multiple inputs, particularly, trajectory control of a robotic manipulator.

Let  $q \in \mathbb{R}^n$  be the joint angles of a multi-link robotic arm, and  $\dot{q} \in \mathbb{R}^n$  be the joint velocities. The dynamics of a robotic manipulator reads

$$H(q)\ddot{q} + C(q, \dot{q})\dot{q} + g(q) = \tau, \quad (6.34)$$

where  $H(q) \in \mathbb{S}_{++}^n$  is the manipulator inertia matrix (that is positive definite),  $C(q, \dot{q})\dot{q}$  is a vector of centripetal and Coriolis torques (with  $C(q, \dot{q}) \in \mathbb{R}^{n \times n}$ ), and  $g(q)$  denotes gravitational torques.

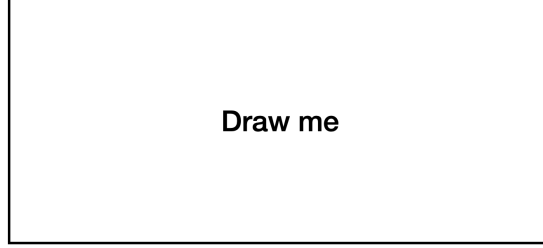


Figure 6.1: Planar two-link manipulator

**Example 6.1** (Planar Two-link Manipulator). The dynamics of a planar two-link manipulator in Fig. 6.1 is

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} -h\dot{q}_2 & -h(\dot{q}_1 + \dot{q}_2) \\ h\dot{q}_1 & 0 \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix}, \quad (6.35)$$

where

$$H_{11} = a_1 + 2a_3 \cos q_2 + 2a_4 \sin q_2 \quad (6.36)$$

$$H_{12} = H_{21} = a_2 + a_3 \cos q_2 + a_4 \sin q_2 \quad (6.37)$$

$$H_{22} = a_2 \quad (6.38)$$

$$h = a_3 \sin q_2 - a_4 \cos q_2 \quad (6.39)$$

with

$$a_1 = I_1 + m_1 l_{c1}^2 + I_e + m_e l_{ce}^2 + m_e l_1^2 \quad (6.40)$$

$$a_2 = I_e + m_e l_{ce}^2 \quad (6.41)$$

$$a_3 = m_e l_1 l_{ce} \cos \delta_e \quad (6.42)$$

$$a_4 = m_e l_1 l_{ce} \sin \delta_e. \quad (6.43)$$

As seen from the above example, the parameters  $a$  (which are nonlinear functions of the physical parameters such as mass and length) enter linearly in  $H$  and  $C$  ( $g(q)$  is ignored because the manipulator is on a horizontal plane).

The goal of the control design is to have the manipulator track a desired trajectory  $q_d(t)$ .

**Known parameters.** When the parameters are known, we follow the sliding control design framework. Let  $\tilde{q} = q(t) - q_d(t)$  be the tracking error, and define the combined error

$$s = \dot{\tilde{q}} + \Lambda \tilde{q} = \dot{q} - \underbrace{(\dot{q}_d - \Lambda \tilde{q})}_{\dot{q}_r}$$

where  $\Lambda \in \mathbb{S}_{++}^n$  is a user-chosen positive definite matrix (in general we want  $-\Lambda$  to be Hurwitz). In this case,  $s \rightarrow 0$  implies  $\tilde{q} \rightarrow 0$  as  $t \rightarrow \infty$ . Choosing the

control law (coming from feedback linearization Appendix D)

$$\tau = H\ddot{q}_r - K_D s + C\dot{q} + g(q) \quad (6.44)$$

with  $K_D \in \mathbb{S}_{++}^n$  positive definite leads to the closed-loop dynamics

$$H\dot{s} + K_D s = 0 \iff \dot{s} = -H^{-1}K_D s.$$

Because the matrix  $H^{-1}K_D$  is the product of two positive definite matrices (recall  $H$  is positive definite and so is  $H^{-1}$ ), it has strictly positive real eigenvalues.<sup>1</sup> Hence,  $-H^{-1}K_D$  is Hurwitz and  $s$  is guaranteed to converge to zero.

**Control law.** A closer look at the controller (6.44) allows us to write it in the following form

$$\tau = H\ddot{q}_r + C(s + \dot{q}_r) + g(q) - K_D s \quad (6.45)$$

$$= H\ddot{q}_r + C\dot{q}_r + g(q) + (C - K_D)s \quad (6.46)$$

$$= Y(q, \dot{q}, \ddot{q}_r)a + (C - K_D)s \quad (6.47)$$

where  $a \in \mathbb{R}^m$  contains all the parameters and  $Y \in \mathbb{R}^{n \times m}$  is the matrix that collects all the coefficients of  $a$  in  $H\ddot{q}_r + C\dot{q}_r + g(q)$ . As a result, we design the adaptive control law to be

$$\tau = Y\hat{a} - K_D s, \quad (6.48)$$

with  $\hat{a}$  the time-varying parameter that we wish to adapt. Note that here we have done something strange: the adaptive control law does not exactly follow the controller (6.44) in the known-parameter case.<sup>2</sup> We first separated  $s$  from  $\dot{q}$  and wrote  $Y a = H\ddot{q}_r + C\dot{q}_r + g$  instead of  $Y a = H\ddot{q}_r + C\dot{q} + g$ ; then we dropped the “ $C$ ” matrix in front of  $s$  in the adaptive control law. The reason for doing this will soon become clear when we analyze the tracking convergence.

**Adaptation law and tracking convergence.** Recall that the key of adaptive control is to design a control law and an adaptation law such that global convergence of the tracking error  $s$  can be guaranteed by a Lyapunov function. Looking at the previous Lyapunov functions in (6.18) and (6.33), we see that they both contain a positive definite term in the tracking error  $s$  (or  $e$  if in first-order systems) and another positive definite term in the parameter error  $\tilde{a}$ . This hints us that we may try a Lyapunov candidate function of the following form

$$V = \frac{1}{2} (s^T H s + \tilde{a}^T \Gamma^{-1} \tilde{a}), \quad (6.49)$$

<sup>1</sup>Consider two positive definite matrices  $A$  and  $B$ , let  $B = B^{1/2}B^{1/2}$ . The product  $AB$  can be written as  $AB = AB^{1/2}B^{1/2} = B^{-1/2}(B^{1/2}AB^{1/2})B^{1/2}$ . Therefore  $AB$  is similar to  $B^{1/2}AB^{1/2}$  and is positive definite.

<sup>2</sup>In fact, one can show that the controller (6.48) with known parameters, i.e.,  $\tau = Y a - K_D s$ , also guarantees the convergence of  $s$  towards zero, though it is different from the feedback linearization controller (6.44). Try proving the convergence with a Lyapunov candidate  $V = \frac{1}{2} s^T H s$ .

where  $\Gamma \in \mathbb{S}_{++}^m$  is a constant positive definite matrix, and  $\tilde{a} = \hat{a} - a$  is the parameter error.

The next step would be to derive the time derivative of  $V$ , which, as we can expect, will contain a term that involves  $\dot{H}$  and complicates our analysis. Fortunately, the following lemma will help us.

**Lemma 6.2.** *For the manipulator dynamics (6.34), there exists a way to define  $C$  such that  $\dot{H} - 2C$  is skew-symmetric.*

*Proof.* See Section 9.1, page 399-402 in (?). You should also check if this is true for the planar two-link manipulator dynamics in Example 6.1.  $\square$

With Lemma 6.2, the time derivative of  $V$  in (6.49) reads

$$\dot{V} = s^T H \dot{s} + \frac{1}{2} s^T \dot{H} s + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.50)$$

$$= s^T (H \ddot{q} - H \ddot{q}_r) + \frac{1}{2} s^T \dot{H} s + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.51)$$

$$= s^T (\tau - C \dot{q} - g - H \ddot{q}_r) + \frac{1}{2} s^T \dot{H} s + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.52)$$

$$= s^T (\tau - H \ddot{q}_r - C(s + \dot{q}_r) - g) + \frac{1}{2} s^T \dot{H} s + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.53)$$

$$= s^T (\tau - H \ddot{q}_r - C \dot{q}_r - g) + \frac{1}{2} s^T (\dot{H} - 2C) s + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.54)$$

$$= s^T (\tau - H \ddot{q}_r - C \dot{q}_r - g) + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.55)$$

$$= s^T (Y \hat{a} - K_D s - Y a) + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} \quad (6.56)$$

$$= s^T Y \tilde{a} + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}} - s^T K_D s, \quad (6.57)$$

where we used the manipulator dynamics (6.34) to rewrite  $H \ddot{q}$  in (6.52), used  $\dot{H} - 2C$  is skew-symmetric in (6.54), invoked the adaptive control law (6.48) and reused  $Y a = H \ddot{q}_r + C \dot{q}_r + g(q)$  in (6.56). The derivation above explains why the choice of the control law in (6.48) did not exactly follow its counterpart when the parameters are known: we need to use  $s^T C s$  to cancel  $\frac{1}{2} s^T \dot{H} s$  in (6.54).

We then wonder if we can design  $\dot{\tilde{a}}$  such that  $\dot{V}$  in (6.57) is negative semidefinite? This turns out to be straightforward with the adaptation law

$$\dot{\tilde{a}} = -\Gamma Y^T s, \quad (6.58)$$

to make  $s^T Y \tilde{a} + \tilde{a}^T \Gamma^{-1} \dot{\tilde{a}}$  vanish and so

$$\dot{V} = -s^T K_D s \leq 0.$$

We are not done yet. To show  $s$  converges to zero (which is implied by  $\dot{V}$  converges to zero), by Barbalat's stability certificate 4.3, it suffices to show

$$\ddot{V} = -2s^T K_D \dot{s}$$

is bounded. We already know  $s$  and  $\tilde{a}$  are bounded, due to the fact that  $V$  in (6.49) is bounded. Therefore, we only need to show  $\dot{s}$  is bounded. To do so, we plug the adaptive control law (6.48) into the manipulator dynamics (6.34) and obtain

$$H\dot{s} + (C + K_D)s = Y\tilde{a},$$

which implies the boundedness of  $\dot{s}$  (note that  $H$  is uniformly positive definite, i.e.,  $H \succeq \alpha I$  for some  $\alpha > 0$ ). This concludes the analysis of the tracking convergence  $s \rightarrow 0$  as  $t \rightarrow \infty$ .

## 6.2 Certainty-Equivalent Adaptive Control





# Appendix A

## Linear System Theory

### A.1 Stability

Check this note.

**Theorem A.1** (Lyapunov Equation). *The following is equivalent for a linear time-invariant system  $\dot{x} = Ax$*

1. *The system is globally asymptotically stable, i.e.,  $A$  is Hurwitz and  $\lim_{t \rightarrow \infty} x(t) = 0$  regardless of the initial condition;*
2. *For any positive definite matrix  $Q$ , the unique solution  $P$  to the Lyapunov equation*

$$A^T P + P A = -Q \quad (\text{A.1})$$

*is positive definite.*

*Proof.* (a):  $2 \Rightarrow 1$ . Suppose we are given two positive definite matrices  $P, Q \succ 0$  that satisfies the Lyapunov equation (A.1). Define a scalar function

$$V(x) = x^T P x.$$

It is clear that  $V > 0$  for any  $x \neq 0$  and  $V(x) = 0$  (i.e.,  $V(x)$  is positive definite). We also see  $V(x)$  is radially unbounded because:

$$V(x) \geq \lambda_{\min}(P) \|x\|^2 \Rightarrow \lim_{x \rightarrow \infty} V(x) \rightarrow \infty.$$

The time derivative of  $V$  reads

$$\dot{V} = 2x^T P \dot{x} = x^T (A^T P + P A) x = -x^T Q x.$$

Clearly,  $\dot{V} < 0$  for any  $x \neq 0$  and  $\dot{V}(0) = 0$ . According to Lyapunov's global stability theorem 4.2, we conclude the linear system  $\dot{x} = Ax$  is globally asymptotically stable at  $x = 0$ .

(b):  $1 \Rightarrow 2$ . Suppose  $A$  is Hurwitz, we want to show that, for any  $Q \succ 0$ , there exists a unique  $P \succ 0$  satisfying the Lyapunov equation (A.1). In fact, consider the matrix

$$P = \int_{t=0}^{\infty} e^{A^T t} Q e^{At} dt.$$

Because  $A$  is Hurwitz, the integral exists, and clearly  $P \succ 0$  due to  $Q \succ 0$ . To show this choice of  $P$  satisfies the Lyapunov equation, we write

$$A^T P + P A = \int_{t=0}^{\infty} \left( A^T e^{A^T t} Q e^{At} + e^{A^T t} Q e^{At} A \right) dt \quad (\text{A.2})$$

$$= \int_{t=0}^{\infty} d \left( e^{A^T t} Q e^{At} \right) \quad (\text{A.3})$$

$$= e^{A^T t} Q e^{At} \Big|_{t=\infty} - e^{A^T t} Q e^{At} \Big|_{t=0} = -Q, \quad (\text{A.4})$$

where the last equality holds because  $e^{A\infty} = 0$  (recall  $A$  is Hurwitz).

To show the uniqueness of  $P$ , we assume that there exists another matrix  $P'$  that also satisfies the Lyapunov equation. Therefore,

$$P' = e^{A^T t} P' e^{At} \Big|_{t=0} - e^{A^T t} P' e^{At} \Big|_{t=\infty} \quad (\text{A.5})$$

$$= - \int_{t=0}^{\infty} d \left( e^{A^T t} P' e^{At} \right) \quad (\text{A.6})$$

$$= - \int_{t=0}^{\infty} e^{A^T t} (A^T P' + P' A) e^{At} dt \quad (\text{A.7})$$

$$= \int_{t=0}^{\infty} e^{A^T t} Q e^{At} dt = P, \quad (\text{A.8})$$

leading to  $P' = P$ . Hence, the solution is unique.  $\square$

**Convergence rate estimation.** We now show that Theorem A.1 can allow us to quantify the convergence rate of a (stable) linear system towards zero.

For a Hurwitz linear system  $\dot{x} = Ax$ , let us pick a positive definite matrix  $Q$ . Theorem A.1 tells us we can find a unique  $P \succ 0$  satisfying the Lyapunov equation (A.1). In this case, we can upper bound the scalar function  $V = x^T P x$  as

$$V \leq \lambda_{\max}(P) \|x\|^2.$$

The time derivative of  $V$  is  $\dot{V} = -x^T Q x$ , which can be upper bounded by

$$\dot{V} \leq -\lambda_{\min}(Q) \|x\|^2 \quad (\text{A.9})$$

$$= -\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)} \underbrace{(\lambda_{\max}(P) \|x\|^2)}_{\geq V} \quad (\text{A.10})$$

$$\leq -\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)} V. \quad (\text{A.11})$$

Denoting  $\gamma(Q) = \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)}$ , the above inequality implies

$$V(0)e^{-\gamma(Q)t} \geq V(t) = x^T P x \geq \lambda_{\min}(P) \|x\|^2.$$

As a result,  $\|x\|^2$  converges to zero exponentially with a rate at least  $\gamma(Q)$ , and  $\|x\|$  converges to zero exponentially with a rate at least  $\gamma(Q)/2$ .

**Best convergence rate estimation.** I have used  $\gamma(Q)$  to make it explicit that the rate  $\gamma$  depends on the choice of  $Q$ , because  $P$  is computed from the Lyapunov equation as an implicit function of  $Q$ . Naturally, choosing different  $Q$  will lead to different  $\gamma(Q)$ . So what is the choice of  $Q$  that maximizes the convergence rate estimation?

**Corollary A.1** (Maximum Convergence Rate Estimation).  *$Q = I$  maximizes the convergence rate estimation.*

*Proof.* let us denote  $P_0$  as the solution to the Lyapunov equation with  $Q = I$

$$A^T P_0 + P_0 A = -I.$$

Let  $P$  be the solution corresponding to a different choice of  $Q$

$$A^T P + P A = -Q.$$

Without loss of generality, we can assume  $\lambda_{\min}(Q) = 1$ , because rescaling  $Q$  will rescale  $P$  by the same factor, which does not affect  $\gamma(Q)$ . Subtracting the two Lyapunov equations above we get

$$A^T (P - P_0) + (P - P_0) A = -(Q - I).$$

Since  $Q - I \succeq 0$  (due to  $\lambda_{\min}(Q) = 1$ ), we know  $P - P_0 \succeq 0$  and  $\lambda_{\max}(P) \geq \lambda_{\max}(P_0)$ . As a result,

$$\gamma(Q) = \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)} = \frac{\lambda_{\min}(I)}{\lambda_{\max}(P)} \leq \frac{\lambda_{\min}(I)}{\lambda_{\max}(P_0)} = \gamma(I),$$

and  $Q = I$  maximizes the convergence rate estimation.  $\square$

---

Consider the linear time-invariant (LTI) system

$$\dot{x} = Ax + Bu, \quad y = Cx + Du, \quad (\text{A.12})$$

where  $x \in \mathbb{R}^n$  the state,  $u \in \mathbb{R}^m$  the control, and  $A, B, C, D$  are constant matrices with proper sizes.

**Theorem A.2** (Observability). *The following are equivalent for the LTI system (A.12):*

1.  $(A, C)$  is observable;
2. The observability grammian

$$W_o(t) = \int_0^t e^{A^* \tau} C^* C e^{A \tau} d\tau$$

is positive definite for any  $t > 0$ ;

3. The observability matrix

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

has full column rank;

4. The matrix  $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$  has full column rank for all  $\lambda \in \mathbb{C}$ ;
5. For all  $\lambda$  and  $x$  such that  $Ax = \lambda x$ ,  $Cx \neq 0$ ;
6. The eigenvalues of  $A + LC$  can be freely assigned (with the restriction that complex eigenvalues are in conjugate pairs);
7.  $(A^*, C^*)$  is controllable.

**Theorem A.3** (Detectability). *The following are equivalent for the LTI system (A.12):*

1.  $(A, C)$  is detectable;
2. The matrix  $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$  has full column rank for all  $\lambda \in \mathbb{C}$  such that  $\text{Re}(\lambda) \geq 0$ ;
3. For all  $\lambda$  and  $x$  such that  $Ax = \lambda x$  and  $\text{Re}(\lambda) \geq 0$ ,  $Cx \neq 0$ ;
4. There exists a matrix  $L$  such that  $A + LC$  is Hurwitz;
5.  $(A^*, C^*)$  is stabilizable.

## Appendix B

# Convex Analysis and Optimization



## Appendix C

# The Kalman-Yakubovich Lemma

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

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

*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*

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





## Appendix D

# Feedback Linearization



## Appendix E

# Sliding Control