Linear Systems

Sara Pohland

Created: January 3, 2021
Last Modified: December 14, 2023

Contents

Part I

Overview of Linear Systems

Chapter 1

Dynamical Systems

1.1 Overview of Dynamical Systems

Our entire discussion of linear systems will revolve around the concept of dynamical systems. We will begin by defining what a dynamical system is and what it describes. We will then discuss numerous axioms, properties, and definitions related to dynamical systems, including some common categories of systems.

1.1.1 Definition of a Dynamical System

A **dynamical system** is a system in which a function describes a set of input, output, and state variables as a function of time. We often use T to denote the set of all possible times. If the functions are defined in **continuous time**, then $T = (-\infty, \infty)$ or $T = [0, \infty)$. If the functions are defined in **discrete time**, then $T = \{nT_{step},\ n \in \mathbb{Z}\}$ or $T = \{nT_{step},\ n \in \mathbb{N}\}$. A dynamical system is represented by $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$, where

$\mathcal{U}$ is the set of input functions $u : T \to U$

U is the set of all possible inputs (typically $U = \mathbb{R}^{n_i}$)

$\mathcal{X}$ is the set of state trajectories $x : T \to X$

X is the set of all possible states (typically $X = \mathbb{R}^n$)

$\mathcal{Y}$ is the set of output functions $y : T \to Y$

Y is the set of all possible outputs (typically $Y = \mathbb{R}^{n_o}$)

s is the **state transition function** $s : T \times T \times X \times \mathcal{U} \to X$

r is the **output read-out map** $r : T \times X \times U \to Y$

For initial time $t_0 \in T$, final time $t_1 \in T$, initial state $\boldsymbol{x_0} := x(t_0) \in X$, and input function $u \in \mathcal{U}$ defined over $[t_0, t_1]$, the state $\boldsymbol{x_1} = x(t_1) \in X$ is given by

$$\boldsymbol{x_1} = s(t_1, t_0, \boldsymbol{x_0}, u).$$

For some time $t \in T$, a state $x(t) \in X$ at that time, and an input $u(t) \in U$ at that time, the output $y(t) \in Y$ at that time is given by

$$y(t) = r(t, x(t), u(t)).$$

1.1.2 Response Function

The composition of the state transition function and the output read-out map is called the **response function**. Given that the system started at an initial time $t_0 \in T$ and initial state $\boldsymbol{x_0} := x(t_0)$ and assuming we applied the input u over the time interval $[t_0, t]$, the response at time $t \in T$ is

$$\begin{aligned} y(t) &= \rho(t, t_0, \boldsymbol{x_0}, u) \\ &= r(t, s(t, t_0, \boldsymbol{x_0}, u), u(t)). \end{aligned}$$

The **zero state response**, which is also referred as the **forced response**, is the response of the system when the initial condition is zero (i.e. $\boldsymbol{x_0} = \mathbf{0}_X$). The zero state response can be expressed as

$$y(t) = \rho(t, t_0, \mathbf{0}_X, u).$$

The **zero input response**, which is also referred to as the **natural response**, is the response of the system under the absence of any input (i.e. $u = 0_{\mathcal{U}}$). The zero input response can be expressed as

$$y(t) = \rho(t, t_0, \boldsymbol{x_0}, 0_{\mathcal{U}}).$$

1.2 Dynamical System Axioms

The state transition map of a valid dynamical system is required to satisfy two axioms: the state transition axiom and the semi-group axiom.

1.2.1 State Transition Axiom

The **state transition axiom** says that if $t_0, t_1 \in T$ are an initial and a final time that satisfy $t_0 \leq t_1$, $\boldsymbol{x_0} := x(t_0)$ is an initial state, and $u, \tilde{u} \in \mathcal{U}$ are two different input functions that satisfy $u(t) \equiv \tilde{u}(t)$ for all $t \in [t_0, t_1]$, then

$$s(t_1, t_0, \boldsymbol{x_0}, u) = s(t_1, t_0, \boldsymbol{x_0}, \tilde{u}).$$

This means that, given an initial state, the final state does not depend on the input prior to the initial time. The initial state summarizes all of the effects of the prior inputs. Similarly, the final state does not depend on the input after the final time, which means that the system is not anticipative/causal.

1.2.2 Semi-Group Axiom

The state transition map must also satisfy the **semi-group axiom**, which says that if $t_0, t_1, t_2 \in T$ are an initial, an intermediate, and a final time that satisfy $t_0 \leq t_1 \leq t_2$, $\boldsymbol{x_0} := x(t_0)$ is an initial state, and $u \in \mathcal{U}$ is an input function, then

$$s(t_2, t_0, \boldsymbol{x_0}, u) = s(t_2, t_1, s(t_1, t_0, \boldsymbol{x_0}, u), u).$$

This means that the final state of the system can be equivalently determined by either the initial state at the initial time under some input, or by an intermediate state at the intermediate time under the same input function.

1.3 Properties of Dynamical System

The are two important properties of dynamical systems satisfy: time-invariance and linearity. Dynamical systems are not required to satisfy either property, but systems that satisfy these properties are often more simple to work with.

1.3.1 Time-Invariance

The **shift operator** is a function $T_\tau : \mathcal{U} \to \mathcal{U}$, which is defined as $(T_\tau u)(t) = u(t-\tau)$. A dynamical system $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$ is **time-invariant** if

1. $\mathcal{U}$ is closed under T_τ for all $\tau \in T$. (Note that this is always true if $\mathcal{U}$ is composed of piecewise continuous functions.)

2. For all $t_0, t_1, \tau \in T$ such that $t_1 \geq t_0$, all $\boldsymbol{x_0} := x(t_0) \in X$, and all $u \in \mathcal{U}$,

$$\rho(t_1, t_0, \boldsymbol{x_0}, u) = \rho(t_1 + \tau, t_0 + \tau, \boldsymbol{x_0}, T_\tau u).$$

This says that a dynamical system is time-invariant if the response does not change if you start τ seconds later. Equivalently, a dynamical system is time-invariant if the evolution of the system does not depend on the initial time.

1.3.2 Linearity

A dynamical system $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$ is said to be **linear** if

1. U, X, and Y are all linear spaces over the same field, F.

2. For all $t_0, t \in T$ such that $t \geq t_0$, all initial states $\boldsymbol{x_0} := x(t_0) \in X$, all alternative initial states $\boldsymbol{\tilde{x}_0} := \tilde{x}(t_0) \in X$, and for all $u, \tilde{u} \in \mathcal{U}$,

$$\rho(t, t_0, \alpha\boldsymbol{x_0} + \tilde{\alpha}\boldsymbol{\tilde{x}_0}, \alpha u + \tilde{\alpha}\tilde{u}) = \alpha\rho(t, t_0, \boldsymbol{x_0}, u) + \tilde{\alpha}\rho(t, t_0, \boldsymbol{\tilde{x}_0}, \tilde{u}).$$

If a dynamical system is linear, then the response of the system can be separated into the zero state response and zero input response in the following way:

$$\rho(t, t_0, \boldsymbol{x_0}, u) = \rho(t, t_0, \boldsymbol{0}_X, u) + \rho(t, t_0, \boldsymbol{x_0}, 0_{\mathcal{U}}).$$

We will generally assume that the systems we work with in these notes are linear because linear systems exhibit some other very nice properties as well.

1.4 Common Types of Dynamical Systems

From our definitions of time-invariance and linearity, we can define four categories of dynamical systems: continuous linear time-varying systems, continuous linear time-invariant systems, discrete time-varying systems, and discrete time-invariant systems. Notice that all of these systems are linear. Only the second and fourth categories of systems satisfy the time-invariance property.

1.4.1 Continuous Linear Time-Varying Systems

A **continuous linear time-varying (LTV)** system can be represented as

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ y(t) = C(t)x(t) + D(t)u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$ is the state at time t, $u(t) \in \mathbb{R}^{n_i}$ is the input at time t, and $y(t) \in \mathbb{R}^{n_o}$ is the output at time t. $A(t) \in \mathbb{R}^{n \times n}$, $B(t) \in \mathbb{R}^{n \times n_i}$, $C(t) \in \mathbb{R}^{n_o \times n}$, and $D(t) \in \mathbb{R}^{n_o \times n_i}$ are matrices that describe the dynamics of the system.

1.4.2 Continuous Linear Time-Invariant Systems

A **continuous linear time-invariant (LTI)** system can be represented as

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$ is the state at time t, $u(t) \in \mathbb{R}^{n_i}$ is the input at time t, and $y(t) \in \mathbb{R}^{n_o}$ is the output at time t. $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o \times n_i}$ are constant matrices that describe the dynamics of the system.

1.4.3 Discrete Linear Time-Varying Systems

A **discrete linear time-varying (LTV)** system can be represented as

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A_k x_k} + \boldsymbol{B_k u_k} \\ \boldsymbol{y_k} = \boldsymbol{C_k x_k} + \boldsymbol{D_k u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$ is the state at time k, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$ is the input at time k, and $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$ is the output at time k. $\boldsymbol{A_k} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B_k} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C_k} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D_k} \in \mathbb{R}^{n_o \times n_i}$ are matrices that describe the dynamics of the system.

1.4.4 Discrete Linear Time-Invariant Systems

A discrete linear time-invariant (LTI) system can be represented as

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\ \boldsymbol{y_k} = \boldsymbol{C}\boldsymbol{x_k} + \boldsymbol{D}\boldsymbol{u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$ is the state at time k, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$ is the input at time k, and $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$ is the output at time k. $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$ are constant matrices that describe the dynamics of the system.

1.5 Equivalence

The last general concept we will discuss related to dynamical systems is equivalence. We can think about both equivalent states and equivalent systems.

1.5.1 Equivalent States

Let $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$ and $\tilde{\mathcal{D}} = (\mathcal{U}, \tilde{\mathcal{X}}, \mathcal{Y}, \tilde{s}, \tilde{r})$ be two dynamical systems with the same input and output spaces. The initial state $\boldsymbol{x_0} \in X$ in system $\mathcal{D}$ is **equivalent** to the initial state $\boldsymbol{\tilde{x}_0} \in \tilde{X}$ in system $\tilde{\mathcal{D}}$ if for all times $t_0, t \in T$ that satisfy $t \geq t_0$, the response of the system is the same, meaning

$$\rho(t, t_0, \boldsymbol{x_0}, u) = \rho(t, t_0, \boldsymbol{\tilde{x}_0}, u).$$

This says that two initial states are equivalent if they lead to the same response when the same input is applied over the same time interval.

1.5.2 Equivalent Systems

Let $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$ and $\tilde{\mathcal{D}} = (\tilde{\mathcal{U}}, \tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{s}, \tilde{r})$ be two dynamical systems. The systems $\mathcal{D}$ and $\tilde{\mathcal{D}}$ are **equivalent** if for all initial times $t_0 \in T$ and for all states $\boldsymbol{x} \in X$ in the dynamical system $\mathcal{D}$, there exists at least one state $\boldsymbol{\tilde{x}} \in \tilde{X}$ in the dynamical system $\tilde{\mathcal{D}}$ that is equivalent to $\boldsymbol{x}$ at time t. This means that two dynamical systems are equivalent if they have the same input-output pairs.

1.5.3 Modal Form

For any continuous LTI system, we can define an equivalent system that is said to be in modal form. Consider a continuous LTI system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, $y(t) \in \mathbb{R}^{n_o}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. Suppose $\boldsymbol{A}$ admits the diagonalization $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}$. Let's define a new trajectory function, z, such that $z(t) \coloneqq \boldsymbol{U}^{-1}x(t)$ for all times $t \in T$. We can describe the evolution of this new system as

$$\begin{aligned}\dot{z}(t) &= \boldsymbol{U}^{-1}\dot{x}(t)\\ &= \boldsymbol{U}^{-1}\big(\boldsymbol{A}x(t) + \boldsymbol{B}u(t)\big)\\ &= \boldsymbol{U}^{-1}\boldsymbol{A}x(t) + \boldsymbol{U}^{-1}\boldsymbol{B}u(t)\\ &= \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}z(t) + \boldsymbol{U}^{-1}\boldsymbol{B}u(t).\end{aligned}$$

Similarly, we can define the output of this new system as

$$y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) = \boldsymbol{C}\boldsymbol{U}z(t) + \boldsymbol{D}u(t).$$

Now we have a new linear time-invariant (LTI) system described by

$$\begin{cases}\dot{z}(t) = \tilde{\boldsymbol{A}}z(t) + \tilde{\boldsymbol{B}}u(t)\\ y(t) = \tilde{\boldsymbol{C}}z(t) + \tilde{\boldsymbol{D}}u(t)\\ z(t_0) = \boldsymbol{z_0}\end{cases},$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{\Lambda}$, $\tilde{\boldsymbol{B}} = \boldsymbol{U}^{-1}\boldsymbol{B}$, $\tilde{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{U}$, and $\tilde{\boldsymbol{D}} = \boldsymbol{D}$. This system is said to be in **modal form** because we can easily read off the modes/eigenvalues of the system in the dynamics matrix $\tilde{\boldsymbol{A}} = \boldsymbol{\Lambda}$. The original system and the system in modal form have the same input-output pairs, so they are equivalent.

Chapter 2

Differential Equations

2.1 Ordinary Differential Equations

Recall from sections 1.4.1 and 1.4.2 that the evolution of both continuous LTV and continuous LTI systems can be described by a differential equation. It is, therefore, useful to discuss some important concepts related to differential equations. **Ordinary differential equations (ODEs)** have the general form

$$\dot{x}(t) = \frac{d}{dt}x(t) = f(x(t), t),$$

where $x(t) \in \mathbb{R}^n$ is the state, $t \in \mathbb{R}_+$ is the time, $x(t_0) = \boldsymbol{x_0} \in \mathbb{R}^n$ is the initial condition, and $f : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is a function describing the way the state changes with time. The solution to an ordinary differential equation (ODE) is

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau), \tau)d\tau.$$

We can prove that an expression, $\phi(t)$ is a solution to the differential equation $\dot{x}(t) = f(x(t), t)$ with initial condition $x(t_0) = \boldsymbol{x_0}$ by showing that it satisfies

1. the differential equation – $\frac{d}{dt}\phi(t) = f(\phi(t), t)$ and
2. the initial condition – $\phi(t_0) = \boldsymbol{x_0}$.

2.2 Existence and Uniqueness

Often, we want to know if there exists a solution to a differential equation with a given initial condition. If some solution exists, we are also interested in whether this solution is unique. We can determine whether a solution exists and is unique using the fundamental theorem of differential equations. Before discussing this theorem, it is useful to first review some properties of functions: continuity, continuous differentiability, piecewise continuity, and Lipschitz continuity.

2.2.1 Continuity & Continuous Differentiability

A function $f(\boldsymbol{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is said to be **continuous** in $\boldsymbol{x}$ if for all $\boldsymbol{c} \in \mathbb{R}^n$ and for all $t \in \mathbb{R}_+$, (1) the point $f(\boldsymbol{c}, t)$ is defined and (2) f satisfies

$$\lim_{\boldsymbol{x}\to\boldsymbol{c}^-} f(\boldsymbol{x}, t) = \lim_{\boldsymbol{x}\to c^+} f(\boldsymbol{x}, t) = f(\boldsymbol{c}, t).$$

A function $f(\boldsymbol{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is said to be **continuously differentiable** in $\boldsymbol{x}$ if its gradient, $\nabla_x f(\boldsymbol{x}, t)$, exists for all values of $\boldsymbol{x} \in \mathbb{R}^n$ and is continuous.

2.2.2 Piecewise Continuity

A function $f(\boldsymbol{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is **piecewise continuous** in t for all $\boldsymbol{x}$ if

1. $f(\boldsymbol{x}, \cdot) : \mathbb{R}_+ \to \mathbb{R}^n$ is continuous except at points of discontinuity and
2. there are only finitely many points of discontinuity in any closed and bounded interval on the domain of f.

2.2.3 Lipschitz Continuity

A function $f(\boldsymbol{x}, t) : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ is **Lipschitz continuous** in $\boldsymbol{x}$ for all t if there exists a piecewise continuous function $k(t) : \mathbb{R}_+ \to \mathbb{R}_+$ that satisfies

$$||f(\boldsymbol{x}, t) - f(\boldsymbol{z}, t)|| \le k(t)||\boldsymbol{x} - \boldsymbol{z}||.$$

If the Lipschitz condition is satisfied for all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$ and all $t \in \mathbb{R}_+$, then the function f is **globally Lipschitz continuous**. If the Lipschitz condition is satisfied for all $\boldsymbol{x}, \boldsymbol{z} \in S \subset \mathbb{R}^n$ and all $t \in \mathbb{R}_+$, then the function f is **locally Lipschitz continuous** in S. In showing that a function is Lipschitz continuous, it is often helpful to use the **mean value theorem**, which says

$$||f(\boldsymbol{x}, t) - f(\boldsymbol{z}, t)|| \le \left\| D_x f(\boldsymbol{x}, t) \right\| ||\boldsymbol{x} - \boldsymbol{z}||.$$

This theorem is useful because if the induced norm of the Jacobian of f with respect to $\boldsymbol{x}$ is bounded by a piecewise continous function, $k(t)$, for all $\boldsymbol{x}$, then $f(\boldsymbol{x}, t)$ is globally Lipschitz continuous in $\boldsymbol{x}$. Similarly, if the induced norm is bounded by a piecewise continuous function for some subset of the values of $\boldsymbol{x}$, then $f(\boldsymbol{x}, t)$ is locally Lipschitz continuous. We can show that any induced norm is bounded, but it generally easiest to use the induced l_1 or l_∞ norm.

Determining Lipschitz Continuity

Lipschitz continuity is a useful property, but it is not always trivial to determine if a function is Lipschitz continuous. If we want to know whether a function is Lipschitz continuous, the following procedures can provide some guidance:

1. Check if the function is linear.

   (a) If the function is linear, it is globally Lipschitz continuous.
   (b) If it is nonlinear, it may or may not be Lipschitz continuous.

2. If the function is nonlinear, check if any of the induced norms of the Jacobian, $D_x f(\boldsymbol{x}, t)$, are bounded.
   (a) If one of the norms is bounded for all possible values of $\boldsymbol{x}$, then the function is globally Lipschitz continuous.
   (b) If one of the norms is bounded for some subset of the values of $\boldsymbol{x}$, then the function is locally Lipschitz continuous within that subset. The function may or may not be globally Lipschitz continuous.
   (c) If none of the induced norms are bounded, then the function may or may not be Lipschitz continuous.

3. If none of the induced norms of the Jacobian can determine whether the function is (globally) Lipschitz continuous, try manipulating the norms $\|f(\boldsymbol{x}, t) - f(\boldsymbol{z}, t)\|$ and $\|\boldsymbol{x} - \boldsymbol{z}\|$ directly to find a valid function, $k(t)$, that satisfies the Lipschitz condition either globally or locally.
   (a) If we find a piecewise continuous function, $k(t)$, that holds for all values of $\boldsymbol{x}$, then the function is globally Lipschitz continuous.
   (b) If we find a function, $k(t)$, that holds for some subset of $\boldsymbol{x}$, then the function is locally Lipschitz continuous within that subset. The function may or may not be globally Lipschitz continuous.
   (c) If we cannot find a piecewise continuous function, $k(t)$, then the function may or may not be Lipschitz continuous.

4. If we have not been able to show that a function is globally Lipschitz continuous, try finding a counterexample to show that the function is not globally Lipschitz continuous. To find a counterexample, assume the function is globally Lipschitz continuous. Then try to choose values for $\boldsymbol{x}$ and $\boldsymbol{z}$ in terms of $k(t)$ that make the Lipschitz inequality false, disproving the assumption that the function is globally Lipschitz continuous.
   (a) If we find a counterexample, the function is not globally Lipschitz continuous. It may or may not be locally Lipschitz continuous.
   (b) If we cannot find a counterexample, then we cannot determine whether the function is or is not globally Lipschitz continuous.

2.2.4 Fundamental Theorem of Differential Equations

Theorem: Consider the ordinary differential equation $\dot{x}(t) = f(x(t), t)$ with initial condition $x(t_0) = \boldsymbol{x_0}$. If f is (1) piecewise continuous in t and (2) Lipschitz continuous in $\boldsymbol{x}$, then the **fundamental theorem of differential equations** says that there exists a *unique* function of time $\phi : \mathbb{R}_+ \to \mathbb{R}^n$, which is continuous and differentiable everywhere, satisfying both the differential equation and initial condition. We can express these two properties of ϕ as

1. $\phi(t_0) = \boldsymbol{x_0}$
2. $\dot{\phi}(t) = f\big(\phi(t), t\big), \;\forall t \in T \backslash D$, where D is the set of all discontinuity points

Note that if $f(\boldsymbol{x}, t)$ is locally, but not globally, Lipschitz continuous, then the solution, $\phi(t)$, exists but may only exist in some bounded region.

Proof (Uniqueness): We will not entirely prove the fundamental theorem of differential equations, but we will prove the uniqueness component. If we find a solution, $\phi(t)$, to the differential equation $\dot{x}(t) = f(x(t), t)$ with initial condition $x(t_0) = \boldsymbol{x_0}$, then we can show that this solution is unique. We will do so by setting up a proof by contradiction. Suppose that there are two solutions $\phi(t)$ and $\psi(t)$, which both satisfy the differential equation such that

$$\dot{\phi}(t) = f\big(\phi(t), t\big), \;\phi(t_0) = \boldsymbol{x_0},$$

$$\dot{\psi}(t) = f\big(\psi(t), t\big), \;\psi(t_0) = \boldsymbol{x_0}.$$

Using the solution to a general ODE defined in section 2.1, we can write

$$\phi(t) = \boldsymbol{x_0} + \int_{t_0}^{t} f\big(\phi(\tau), \tau\big) d\tau,$$

$$\psi(t) = \boldsymbol{x_0} + \int_{t_0}^{t} f\big(\psi(\tau), \tau\big) d\tau.$$

Subtracting the second solution from the first, notice that

$$\phi(t) - \psi(t) = \int_{t_0}^{t} \Big(f\big(\phi(\tau), \tau\big) - f\big(\psi(\tau), \tau\big) \Big) d\tau.$$

Now if we take the norm of this difference, we can obtain an upper bound:

$$\begin{aligned} \big\|\phi(t) - \psi(t)\big\| &= \left\| \int_{t_0}^{t} \Big(f\big(\phi(\tau), \tau\big) - f\big(\psi(\tau), \tau\big) \Big) d\tau \right\| \\ &\le \int_{t_0}^{t} \Big\| f\big(\phi(\tau), \tau\big) - f\big(\psi(\tau), \tau\big) \Big\| d\tau. \end{aligned}$$

If the function f is globally Lipschitz continuous in $\boldsymbol{x}$, then

$$\big\|\phi(t) - \psi(t)\big\| \le \int_{t_0}^{t} k(\tau) \Big\| \phi(\tau) - \psi(\tau) \Big\| d\tau.$$

The **Bellman-Gronwall lemma** is useful in proving the fundamental theorem of differential equations. Let u and k be real-valued, piecewise continuous functions on $\mathbb{R}_+$, and assume that $u(t)$ and $k(t)$ are strictly positive for all $t \in \mathbb{R}_+$. Assume $c \ge 0$ and $t_0 \in \mathbb{R}_+$. The Bellman-Gronwall lemma says

$$u(t) \le c + \int_{t_0}^{t} k(\tau) u(\tau) d\tau \quad \Longrightarrow \quad u(t) \le c \exp\left(\int_{t_0}^{t} k(\tau) d\tau \right).$$

Now using the Bellman-Gronwall lemma, we find that

$$\|\phi(t) - \psi(t)\| \leq 0 \cdot \exp\left(\int_{t_0}^{t} k(\tau)d\tau\right) = 0$$

We know that norms cannot be negative, which implies that $\|\phi(t) - \psi(t)\| = 0$. From properties of norms, this then implies that $\phi(t) - \psi(t) = 0$, indicating that $\phi(t) = \psi(t)$. Now we have shown that $\phi(t)$ must be a unique solution.

2.2.5 Reverse Time Differential Equation

If the differential equation $\dot{x}(t) = f(x(t), t)$ with initial condition $x(t_0) = \boldsymbol{x_0}$ is piecewise continuous in t and Lipschitz continuous in $\boldsymbol{x}$, then there exists a unique solution for both the forward time and the reverse time differential equation. Because the differential equation has a unique solution, we know that $x(t)$ cannot diverge from a single initial condition. This is shown in figure 2.1.

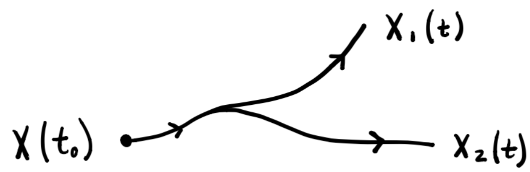

Figure 2.1: If a differential equation is piecewise continuous in time t and Lipschitz continuous in state $\boldsymbol{x}$, then it has a unique solution. This means two solutions, $x_1(t)$ and $x_2(t)$, cannot diverge from the same initial condition, $x(t_0)$.

Since the reverse time differential equation has a unique solution, $x(t)$ also cannot converge from two distinct initial conditions. This is shown in figure 2.2.

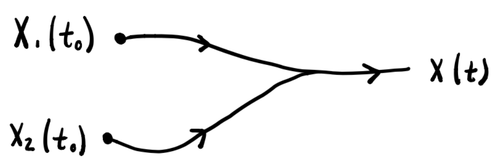

Figure 2.2: If a differential equation is piecewise continuous in time t and Lipschitz continuous in state $\boldsymbol{x}$, then the reverse time differential equation has a unique solution. This means two initial conditions, $x_1(t_0)$ and $x_2(t_0)$, cannot converge to a single solution, $x(t)$.

Chapter 3

State Transition Function

3.1 State Transition Matrix

Previously, we defined several matrices that dictated the evolution of the system for continuous LTV systems (section 1.4.1) and continuous LTI systems (1.4.2). One of these matrices is commonly referred to as the state transition matrix. We will discuss this matrix and its properties for both categories of systems.

3.1.1 Continuous LTV Systems

Consider a continuous linear time-varying (LTV) system described by

$$\begin{cases} \dot{x}(t) = A(t)x(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$ and $A(t) \in \mathbb{R}^{n \times n}$. Assume the solution to this ODE is

$$x(t) = \phi(t, t_0)\boldsymbol{x_0},$$

where $\phi(t, t_0) \in \mathbb{R}^{n \times n}$ will be referred to as the **state transition matrix**. Furthermore, we will claim that the state transition matrix satisfies the following:

1. $\dot{\phi}(t, t_0) = A(t)\phi(t, t_0)$ and
2. $\phi(t, t) = \phi(t_0, t_0) = \boldsymbol{I_n}$.

To show that $x(t) = \phi(t, t_0)\boldsymbol{x_0}$ is actually the solution, we need to show that it satisfies both the differential equation and the initial condition:

1. Differential equation

$$\begin{aligned} \dot{x}(t) &= \frac{d}{dt}x(t) = \frac{d}{dt}\phi(t, t_0)\boldsymbol{x_0} = \dot{\phi}(t, t_0)\boldsymbol{x_0} \\ &= A(t)\phi(t, t_0)\boldsymbol{x_0} = A(t)x(t) \checkmark \end{aligned}$$

2. Initial condition

$$x(t_0) = \phi(t_0, t_0)\boldsymbol{x_0} = \boldsymbol{I_n}\boldsymbol{x_0} = \boldsymbol{x_0} \checkmark$$

Now we have verified that $x(t) = \phi(t, t_0)\boldsymbol{x_0}$ is the solution to the differential equation describing the given continuous LTV system.

3.1.2 Continuous LTI Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. Let's assert that the state transition matrix for this system is $\phi(t, t_0) = e^{\boldsymbol{A}(t-t_0)}$, where $e^{\boldsymbol{A}(t-t_0)} \in \mathbb{R}^{n\times n}$ is the matrix exponential. Thus the solution to this differential equation is given by

$$x(t) = \phi(t, t_0)\boldsymbol{x_0} = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0}.$$

To verify that $x(t) = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0}$ is actually the solution to this ODE, we need to show that it satisfies both the differential equation and the initial condition:

1. Differential equation

$$\begin{aligned} \dot{x}(t) &= \frac{d}{dt}x(t) = \frac{d}{dt}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} \\ &= \boldsymbol{A}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} = \boldsymbol{A}x(t) \checkmark \end{aligned}$$

2. Initial condition

$$\begin{aligned} x(t_0) &= e^{\boldsymbol{A}(t_0-t_0)}\boldsymbol{x_0} = e^{\boldsymbol{A}(0)}\boldsymbol{x_0} \\ &= e^{\boldsymbol{0}^{n\times n}}\boldsymbol{x_0} = \boldsymbol{I_n}\boldsymbol{x_0} = \boldsymbol{x_0} \checkmark \end{aligned}$$

Now we have verified that $x(t) = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0}$ is the solution to the differential equation describing the LTI system that was given. Therefore, the state transition matrix for an LTI system is $\phi(t, t_0) = e^{\boldsymbol{A}(t-t_0)}$.

3.1.3 Properties of State Transition Matrix

For both the linear time-varying (LTV) and linear time-invariant (LTI) system, the state transition matrix must satisfy the following properties:

1. $\phi(t, t_0) = \phi(t, t_1)\phi(t_1, t_0), \ \forall t, t_0, t_1 \in \mathbb{R}_+$
2. $\big(\phi(t, t_0)\big)^{-1} = \phi(t_0, t)$
3. $\det\big(\phi(t, t_0)\big) = \exp\Big(\int_{t_0}^{t} \text{trace}(A(\tau))d\tau\Big)$

For LTI systems, the state transition matrix has additional properties described under the matrix exponential section of my linear algebra course notes.

3.1.4 Calculating State Transition Matrix

To calculate the state transition matrix for a linear time-invariant (LTI) system, we can use methods used to compute the matrix exponential described in the matrix exponential section of my linear algebra notes. To calculate the state transition matrix for a linear time-varying (LTV) system described by

$$\begin{cases} \dot{x}(t) = A(t)x(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

we can use the previously stated properties of the state transition matrix:

1. $\dot{\phi}(t, t_0) = A(t)\phi(t, t_0)$ and
2. $\phi(t, t) = \phi(t_0, t_0) = \boldsymbol{I_n}$.

One technique for identifying the state transition matrix is to first use methods from differential equations to numerically compute the solution, $x^{(i)}(t)$, of the equation $\dot{x}^{(i)}(t) = A(t)x^{(i)}(t)$ with initial condition $x^{(i)}(t_0) = e_i$ for $i = 1, \cdots, n$. Then we can use these n solutions to form the state transition matrix:

$$\phi(t, t_0) = \begin{bmatrix} x^{(1)}(t) & \cdots & x^{(n)}(t) \end{bmatrix}.$$

In general, it can be difficult to find the state transition matrix. However, consider an LTV system described by $\dot{x}(t) = A(t)x(t)$, where $A(t)$ is of the form

$$A(t) = \begin{bmatrix} \alpha(t) & \beta(t) \\ -\beta(t) & \alpha(t) \end{bmatrix}$$

for some functions of time, $\alpha(t)$ and $\beta(t)$. For this case, the state transition is

$$\phi(t, t_0) = \begin{bmatrix} \phi_{11}(t, t_0) & \phi_{12}(t, t_0) \\ \phi_{21}(t, t_0) & \phi_{22}(t, t_0) \end{bmatrix}, \text{ where}$$

$$\phi_{11}(t, t_0) = \exp\left(\int_{t_0}^{t} \alpha(\tau)d\tau\right)\cos\left(\int_{t_0}^{t} \beta(\tau)d\tau\right),$$

$$\phi_{12}(t, t_0) = \exp\left(\int_{t_0}^{t} \alpha(\tau)d\tau\right)\sin\left(\int_{t_0}^{t} \beta(\tau)d\tau\right),$$

$$\phi_{21}(t, t_0) = -\exp\left(\int_{t_0}^{t} \alpha(\tau)d\tau\right)\sin\left(\int_{t_0}^{t} \beta(\tau)d\tau\right),$$

$$\phi_{22}(t, t_0) = \exp\left(\int_{t_0}^{t} \alpha(\tau)d\tau\right)\cos\left(\int_{t_0}^{t} \beta(\tau)d\tau\right).$$

To derive this state transition matrix, we can transform the differential equation $\dot{x}^{(1)}(t) = A(t)x^{(1)}(t)$ with initial condition $x^{(1)}(t_0) = e_1$ and the differential

$\dot{x}^{(2)}(t) = A(t)x^{(2)}(t)$ with initial condition $x^{(2)}(t_0) = e_2$ to polar coordinates to solve, then transform them back to Cartesian. For more information on how to perform these computations, reference chapter 11, "Continuous-Time Linear State-Space Models" in "Lectures on Dynamic Systems and Control" by Mohammed Dahleh, Munther A. Dahleh, and George Verghese.

3.2 State Transition Function

We discussed the state transition matrix in the context of a system whose state evolves in the absence of input. We will now introduce an input to the dynamical system and discuss the state transition function. Once again, we will consider continuous LTV systems generally, then consider the special case of LTI systems.

3.2.1 Continuous LTV Systems

Consider a continuous linear time-varying (LTV) system described by

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, $A(t) \in \mathbb{R}^{n \times n}$, and $B(t) \in \mathbb{R}^{n \times n_i}$. Let's assert that the solution to this differential equation is given by the following function:

$$x(t) = \phi(t, t_0)\boldsymbol{x_0} + \int_{t_0}^{t} \phi(t, \tau)B(\tau)u(\tau)d\tau.$$

We will refer to this solution as the **state transition function**. To show that the state transition function is a valid solution to the given ODE, we need to show that it satisfies both the differential equation and the initial condition. In order to complete this proof, we will use **Leibniz integral rule**, which says

$$\frac{\partial}{\partial \boldsymbol{z}} \int_{a(\boldsymbol{z})}^{b(\boldsymbol{z})} f(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{x} = \int_{a(\boldsymbol{z})}^{b(\boldsymbol{z})} \frac{\partial}{\partial \boldsymbol{z}} f(\boldsymbol{x}, \boldsymbol{z})d\boldsymbol{x} + \frac{\partial b(\boldsymbol{z})}{\partial \boldsymbol{z}} f(b(\boldsymbol{z}), \boldsymbol{z}) - \frac{\partial a(\boldsymbol{z})}{\partial \boldsymbol{z}} f(a(\boldsymbol{z}), \boldsymbol{z}).$$

We can now show that the state transition function is actually a valid solution:

1. Differential equation

$$\begin{aligned}
\dot{x}(t) &= \frac{d}{dt}x(t) = \frac{d}{dt}\left(\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t}\phi(t,\tau)B(\tau)u(\tau)d\tau\right) \\
&= \frac{d}{dt}\phi(t,t_0)\boldsymbol{x_0} + \frac{d}{dt}\int_{t_0}^{t}\phi(t,\tau)B(\tau)u(\tau)d\tau \\
&= \frac{d}{dt}\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t}\frac{d}{dt}\phi(t,\tau)B(\tau)u(\tau)d\tau \\
&\qquad + \frac{d}{dt}(t)\Big(\phi(t,t)B(t)u(t)\Big) - \frac{d}{dt}(t_0)\Big(\phi(t,t_0)B(t_0)u(t_0)\Big) \\
&= \dot{\phi}(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t}\dot{\phi}(t,\tau)B(\tau)u(\tau)d\tau \\
&\qquad + 1\Big(\boldsymbol{I_n}B(t)u(t)\Big) - 0\Big(\phi(t,t_0)B(t_0)u(t_0)\Big) \\
&= A(t)\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t}A(t)\phi(t,\tau)B(\tau)u(\tau)d\tau + B(t)u(t) \\
&= A(t)\left(\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t}\phi(t,\tau)B(\tau)u(\tau)d\tau\right) + B(t)u(t) \\
&= A(t)x(t) + B(t)u(t) \checkmark
\end{aligned}$$

2. Initial condition

$$\begin{aligned}
x(t_0) &= \phi(t_0,t_0)\boldsymbol{x_0} + \int_{t_0}^{t_0}\phi(t_0,\tau)B(\tau)u(\tau)d\tau \\
&= \boldsymbol{I_n}\boldsymbol{x_0} + 0 = \boldsymbol{I_n}\boldsymbol{x_0} = \boldsymbol{x_0} \checkmark
\end{aligned}$$

Now we have verified that the given state transition function is a valid solution to the ODE describing a general continuous LTV system.

3.2.2 Continuous LTI Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$. Let's assert that the solution to this differential equation is given by the **state transition function**:

$$x(t) = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau.$$

To show that the state transition function is a valid solution, we need to show that it satisfies both the differential equation and the initial condition:

1. Differential equation

$$
\begin{aligned}
\dot{x}(t) &= \frac{d}{dt}x(t) = \frac{d}{dt}\left(e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau\right) \\
&= \frac{d}{dt}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \frac{d}{dt}\int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau \\
&= \frac{d}{dt}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} \frac{d}{dt}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau \\
&\qquad + \frac{d}{dt}(t)\Big(e^{\boldsymbol{A}(t-t)}\boldsymbol{B}u(t)\Big) - \frac{d}{dt}(t_0)\Big(e^{\boldsymbol{A}(t-t_0)}\boldsymbol{B}u(t_0)\Big) \\
&= \boldsymbol{A}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} \boldsymbol{A}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau \\
&\qquad + 1\Big(\boldsymbol{I_n}\boldsymbol{B}u(t)\Big) - 0\Big(e^{\boldsymbol{A}(t-t_0)}\boldsymbol{B}u(t_0)\Big) \\
&= \boldsymbol{A}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \boldsymbol{A}\int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau + \boldsymbol{B}u(t) \\
&= \boldsymbol{A}\left(e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau\right) + \boldsymbol{B}u(t) \\
&= \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \checkmark
\end{aligned}
$$

2. Initial condition

$$
\begin{aligned}
x(t_0) &= e^{\boldsymbol{A}(t_0-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t_0} e^{\boldsymbol{A}(t_0-\tau)}\boldsymbol{B}u(\tau)d\tau \\
&= \boldsymbol{I_n}\boldsymbol{x_0} + 0 = \boldsymbol{I_n}\boldsymbol{x_0} = \boldsymbol{x_0} \checkmark
\end{aligned}
$$

Now we have verified that the given state transition function is a valid solution to the ODE describing a general continuous LTI system.

3.3 Response Function

In discussing the state transition matrix and state transition function, we only considered the evolution of the state. Now we will introduce a function for the output and find an expression for the repsonse of the system. We will again consider continuous LTV systems generally, then the special case of LTI systems. We will also recall discrete time systems and derive a response for these systems.

3.3.1 Continuous LTV Systems

Consider a continuous linear time-varying (LTV) system described by

$$
\begin{cases}
\dot{x}(t) = A(t)x(t) + B(t)u(t) \\
y(t) = C(t)x(t) + D(t)u(t) \\
x(t_0) = \boldsymbol{x_0}
\end{cases},
$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $A(t) \in \mathbb{R}^{n\times n}$, $B(t) \in \mathbb{R}^{n\times n_i}$, $C(t) \in \mathbb{R}^{n_o\times n}$, and $D(t) \in \mathbb{R}^{n_o\times n_i}$. Previously, we showed that the state of a general continuous LTV system could be expressed as

$$x(t) = \phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t} \phi(t,\tau)B(\tau)u(\tau)d\tau.$$

This allows us to write the **response** of the system as

$$\begin{aligned} y(t) &= C(t)x(t) + D(t)u(t) \\ &= C(t)\left(\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t} \phi(t,\tau)B(\tau)u(\tau)d\tau\right) + D(t)u(t) \\ &= C(t)\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t} C(t)\phi(t,\tau)B(\tau)u(\tau)d\tau + D(t)u(t) \end{aligned}$$

We refer to this as the response function for continuous LTV systems.

3.3.2 Continuous LTI Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. Previously, we showed that the state of a general continuous LTI system could be expressed as

$$x(t) = e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau.$$

This allows us to write the **response** of the system as

$$\begin{aligned} y(t) &= \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ &= \boldsymbol{C}\left(e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau\right) + \boldsymbol{D}u(t) \\ &= \boldsymbol{C}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau + \boldsymbol{D}u(t) \end{aligned}$$

We refer to this as the response function for continuous LTI systems.

3.3.3 Discrete LTV Systems

So far we have focused on continuous time systems, for which we have defined a state transition function. We can also derive an expression for the response of a discrete time system. Consider a discrete LTV system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A_k x_k} + \boldsymbol{B_k u_k} \\ \boldsymbol{y_k} = \boldsymbol{C_k x_k} + \boldsymbol{D_k u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A_k} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B_k} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C_k} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D_k} \in \mathbb{R}^{n_o\times n_i}$. From the first equation, we can notice the following:

$$\begin{aligned}
\boldsymbol{x_1} &= \boldsymbol{A_0 x_0} + \boldsymbol{B_0 u_0} \\
\boldsymbol{x_2} &= \boldsymbol{A_1 x_1} + \boldsymbol{B_1 u_1} = \boldsymbol{A_1 A_0 x_0} + \boldsymbol{A_1 B_0 u_0} + \boldsymbol{B_1 u_1} \\
\boldsymbol{x_3} &= \boldsymbol{A_2 x_2} + \boldsymbol{B_2 u_2} = \boldsymbol{A_2 A_1 A_0 x_0} + \boldsymbol{A_2 A_1 B_0 u_0} + \boldsymbol{A_2 B_1 u_1} + \boldsymbol{B_2 u_2} \\
&\vdots \\
\boldsymbol{x_k} &= \prod_{i=0}^{k-1} \boldsymbol{A_i x_0} + \sum_{i=0}^{k-1} \prod_{j=i+1}^{k-1} \boldsymbol{A_j B_i u_i}
\end{aligned}$$

Combining this with the second equation, the response of the system is

$$\begin{aligned}
\boldsymbol{y_k} &= \boldsymbol{C_k x_k} + \boldsymbol{D_k u_k} \\
&= \boldsymbol{C_k} \left(\prod_{i=0}^{k-1} \boldsymbol{A_i x_0} + \sum_{i=0}^{k-1} \prod_{j=i+1}^{k-1} \boldsymbol{A_j B_i u_i} \right) + \boldsymbol{D_k u_k} \\
&= \boldsymbol{C_k} \prod_{i=0}^{k-1} \boldsymbol{A_i x_0} + \sum_{i=0}^{k-1} \boldsymbol{C_k} \prod_{j=i+1}^{k-1} \boldsymbol{A_j B_i u_i} + \boldsymbol{D_k u_k}.
\end{aligned}$$

3.3.4 Discrete LTI Systems

Because discrete LTI systems are simply a special case of discrete LTV systems, we can go through the same process to derive the response function for these systems. Consider a discrete LTI system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A x_k} + \boldsymbol{B u_k} \\ \boldsymbol{y_k} = \boldsymbol{C x_k} + \boldsymbol{D u_k} \end{cases}$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. From the first equation, we can notice the following:

$$\begin{aligned}
\boldsymbol{x_1} &= \boldsymbol{A x_0} + \boldsymbol{B u_0} \\
\boldsymbol{x_2} &= \boldsymbol{A x_1} + \boldsymbol{B u_1} = \boldsymbol{A}^2 \boldsymbol{x_0} + \boldsymbol{A B u_0} + \boldsymbol{B u_1} \\
\boldsymbol{x_3} &= \boldsymbol{A x_2} + \boldsymbol{B u_2} = \boldsymbol{A}^3 \boldsymbol{x_0} + \boldsymbol{A}^2 \boldsymbol{B u_0} + \boldsymbol{A B u_1} + \boldsymbol{B u_2} \\
&\vdots \\
\boldsymbol{x_k} &= \boldsymbol{A}^k \boldsymbol{x_0} + \sum_{i=0}^{k-1} \boldsymbol{A}^{k-1-i} \boldsymbol{B u_i}
\end{aligned}$$

Combining this with the second equation, the response of the system is

$$\begin{aligned}\boldsymbol{y_k} &= \boldsymbol{Cx_k} + \boldsymbol{Du_k} \\ &= \boldsymbol{C}\left(\boldsymbol{A}^k\boldsymbol{x_0} + \sum_{i=0}^{k-1}\boldsymbol{A}^{k-1-i}\boldsymbol{Bu_i}\right) + \boldsymbol{Du_k} \\ &= \boldsymbol{CA}^k\boldsymbol{x_0} + \sum_{i=0}^{k-1}\boldsymbol{CA}^{k-1-i}\boldsymbol{Bu_i} + \boldsymbol{Du_k}\end{aligned}$$

3.4 Linearization

Our discussion of the state transition function and response function has been limited to linear systems. Most real-world systems, however, are nonlinear. In some cases, we can obtain a linear system to approximate the evolution and response of a nonlinear system. This process of obtaining a linear approximation is referred to as **linearization**. Consider a nonlinear system described by

$$\begin{cases}\dot{x}(t) = f\big(x(t), u(t), t\big) \\ x(t_0) = \boldsymbol{x_0}\end{cases}.$$

Suppose we have some nominal input function, u^0, which results in the nominal state trajectory, x^0. Assume that this system can be described by

$$\begin{cases}\dot{x}^0(t) = f\big(x^0(t), u^0(t), t\big) \\ x^0(t_0) = \boldsymbol{x_0}\end{cases}.$$

Now let the nominal input function, u^0, be perturbed to $u^0 + \delta u$, resulting in a perturbed state trajectory, $x^0 + \delta x$. The initial condition is also perturbed from $\boldsymbol{x_0}$ to $\boldsymbol{x_0} + \boldsymbol{\delta x_0}$. Now we have a new system, which can be described by

$$\begin{cases}\frac{d}{dt}\big(x^0 + \delta x\big)(t) = f(x^0(t) + \delta x(t), u^0(t) + \delta u(t), t) \\ (x^0 + \delta x)(t_0) = \boldsymbol{x_0} + \boldsymbol{\delta x_0}\end{cases}.$$

Using the Taylor series expansion, we can express the function in the ODE as

$$\begin{aligned}f\big(x^0(t) + \delta x(t), u^0(t) + \delta u(t), t\big) = f\big(x^0(t), u^0(t), t\big) &+ \frac{\partial f}{\partial x(t)}\Big|_{\substack{x(t)=x^0(t)\\u(t)=u^0(t)}}\delta x(t) \\ &+ \frac{\partial f}{\partial u(t)}\Big|_{\substack{x(t)=x^0(t)\\u(t)=u^0(t)}}\delta u(t) + O(\delta x^2(t), \delta u^2(t)).\end{aligned}$$

We can now obtain the following equation describing the perturbation:

$$\frac{d}{dt}\delta x(t) = \frac{\partial f}{\partial x(t)}\Big|_{\substack{x(t)=x^0(t)\\u(t)=u^0(t)}}\delta x(t) + \frac{\partial f}{\partial u(t)}\Big|_{\substack{x(t)=x^0(t)\\u(t)=u^0(t)}}\delta u(t) + O(\delta x^2(t), \delta u^2(t)).$$

Assuming that higher order terms are small and can be ignored,

$$\frac{d}{dt}\delta x(t) \approx \frac{\partial f}{\partial x(t)}\Big|_{\substack{x(t)=x^0(t)\\u(t)=u^0(t)}}\delta x(t) + \frac{\partial f}{\partial u(t)}\Big|_{\substack{x(t)=x^0(t)\\u(t)=u^0(t)}}\delta u(t).$$

Now we can see that we can approximate the behavior of the nonlinear system around the nominal point, $(x^0(t), u^0(t))$, by the linear equation

$$\dot{\delta x}(t) \approx A(t)\delta x(t) + B(t)\delta u(t), \text{ where}$$

$$A(t) = D_{x(t)}f\big(x(t), u(t), t\big)|_{(x(t),u(t))=(x^0(t),u^0(t))},$$

$$B(t) = D_{u(t)}f\big(x(t), u(t), t\big)|_{(x(t),u(t))=(x^0(t),u^0(t))}.$$

We can then write the solution to this linearized system as

$$\delta x(t) \approx \phi(t, t_0)\boldsymbol{\delta x_0} + \int_{t_0}^{t} \phi(t, \tau)B(\tau)\delta u(\tau)d\tau.$$

3.5 Discretization

Most real-world systems are continuous, but simulated systems are generally discrete since computers must work in discrete time. Therefore, it is often useful to approximate continuous time systems as discrete time systems. This process of obtaining a discrete time system from a continuous time system is referred to as **discretization**. Consider continuous LTI system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases}.$$

Suppose we sample the state of the system every T seconds starting from $t = t_0$. We want to find an equation to describe the evolution of this sampled system. As shown previously, the state of the continuous time system is given by

$$x(t) = e^{\boldsymbol{A}(t-t_0)}x(t_0) + \int_{t_0}^{t} e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau.$$

Let's assume the system starts at time $t_0 = kT$ with initial condition $x(kT)$. The state of the system at time $t = (k+1)T$ is then given by

$$x((k+1)T) = e^{\boldsymbol{A}\big((k+1)T-kT\big)}x(kT) + \int_{kT}^{(k+1)T} e^{\boldsymbol{A}\big((k+1)T-\tau\big)}\boldsymbol{B}u(\tau)d\tau$$

If we assume zero-order hold input, meaning that the input is held constant between $t = kT$ and $t = (k+1)T$, then we can express the state as

$$x((k+1)T) = e^{\boldsymbol{A}\big((k+1)T-kT\big)}x(kT) + \int_{kT}^{(k+1)T} e^{\boldsymbol{A}\big((k+1)T-\tau\big)}\boldsymbol{B}u(kT)d\tau$$

To simplify the notation, we will denote $x(kT)$ by $\boldsymbol{x_k}$ and $u(kT)$ by $\boldsymbol{u_k}$. With this notation, we can now express the state of the system at $t = (k+1)T$ as

$$\begin{aligned}
\boldsymbol{x_{k+1}} &= e^{\boldsymbol{A}\left((k+1)T-kT\right)}\boldsymbol{x_k} + \int_{kT}^{(k+1)T} e^{\boldsymbol{A}\left((k+1)T-\tau\right)}\boldsymbol{B}\boldsymbol{u_k}d\tau \\
&= e^{\boldsymbol{A}T}\boldsymbol{x_k} + \left(\int_{kT}^{(k+1)T} e^{\boldsymbol{A}\left((k+1)T-\tau\right)}Bd\tau\right)\boldsymbol{u_k}.
\end{aligned}$$

Using the change of variables $s := \tau - kT$, we can express $\boldsymbol{x_{k+1}}$ as

$$\boldsymbol{x_{k+1}} = e^{\boldsymbol{A}T}\boldsymbol{x_k} + \left(\int_0^T e^{\boldsymbol{A}\left(T-s\right)}Bds\right)\boldsymbol{u_k}.$$

Now we can represent this system as the discrete time system

$$\boldsymbol{x_{k+1}} = \tilde{\boldsymbol{A}}\boldsymbol{x_k} + \tilde{\boldsymbol{B}}\boldsymbol{u_k}, \text{ where}$$

$$\tilde{\boldsymbol{A}} = e^{\boldsymbol{A}T} \quad \text{and} \quad \tilde{\boldsymbol{B}} = \int_0^T e^{\boldsymbol{A}(T-s)}Bds.$$

Now we have found a discrete time linear representation of the original continuous time linear system, which has been sampled with a period of T.

Part II

Stability Analysis

Chapter 4

BIBO Stability

4.1 Definition of BIBO Stability

We will now discuss the stability of dynamical systems. A system is considered to be **bounded-input bounded-output (BIBO) stable** if bounded inputs produce bounded outputs under zero initial condition. Written more formally, a dynamical system is BIBO stable if there exists a constant $k < \infty$ such that

$$||y||_\infty \leq k||u||_\infty, \; \forall u \in \mathcal{L}_\infty^{n_i}, \text{ where}$$

$$\mathcal{L}_\infty^{n_i} = \{u \in \mathcal{U} : ||u||_\infty < \infty\}.$$

Note that $\mathcal{U}$ is the set of input functions that map a time $t \in T$ to an input $u(t) \in \mathbb{R}^{n_i}$, as defined in section 1.1.1. As another note, $||\cdot||_\infty$ is the L_∞ norm, which is defined in the function norms section of my linear algebra notes.

In contrast, a dynamical system is not BIBO stable if there exists a bounded input that produces an unbounded output under zero initial conditions.

4.2 Impulse Response Matrix

In our discussion of BIBO stability, it is helpful to define the impulse response matrix for both continuous LTV systems and continuous LTI systems.

4.2.1 Continuous LTV Systems

Consider a continuous linear time-varying (LTV) system described by

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ y(t) = C(t)x(t) + D(t)u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $A(t) \in \mathbb{R}^{n\times n}$, $B(t) \in \mathbb{R}^{n\times n_i}$, $C(t) \in \mathbb{R}^{n_o\times n}$, and $D(t) \in \mathbb{R}^{n_o\times n_i}$. In section 3.3.1, we showed that the response of this system could be expressed as

$$y(t) = C(t)\phi(t,t_0)\boldsymbol{x_0} + \int_{t_0}^{t} C(t)\phi(t,\tau)B(\tau)u(\tau)d\tau + D(t)u(t).$$

When discussing BIBO stability, we are only interested in the relationship between the input and output, so we will focus on the zero state response, or the forced response, which is the response of the system when the initial condition is zero (i.e. $\boldsymbol{x_0} = \mathbf{0}_n$). For this LTV system, the zero state response is given by

$$y_{zs}(t) = \int_{t_0}^{t} C(t)\phi(t,\tau)B(\tau)u(\tau)d\tau + D(t)u(t).$$

We can also express the zero state response for this system as

$$y_{zs}(t) = \int_{t_0}^{t} H(t,\tau)u(\tau)d\tau, \text{ where}$$
$$H(t,\tau) = C(t)\phi(t,\tau)B(\tau) + D(t)\delta(t-\tau).$$

The matrix $H(t,\tau)$ is the **impulse response matrix** for the given LTV system. Note that the term $\delta(t-\tau)$ in this matrix is the dirac delta function.

4.2.2 Continuous LTI Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. In section 3.3.2, we showed that the response of this system could be expressed as

$$y(t) = \boldsymbol{C}e^{\boldsymbol{A}(t-t_0)}\boldsymbol{x_0} + \int_{t_0}^{t} \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau + \boldsymbol{D}u(t)$$

Again, when discussing BIBO stability, we are only interested in the relationship between the input and output, so we will look more closely at the zero state response. For this LTI system, the zero state response is given by

$$y_{zs}(t) = \int_{t_0}^{t} \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}u(\tau)d\tau + \boldsymbol{D}u(t).$$

We can also express the zero state response for this system as

$$y_{zs}(t) = \int_{t_0}^{t} H(t,\tau)u(\tau)d\tau, \text{ where}$$
$$H(t,\tau) = \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B} + \boldsymbol{D}\delta(t-\tau).$$

The matrix $H(t,\tau)$ is the **impulse response matrix** for the given LTI system. Again, the term $\delta(t-\tau)$ in the response matrix is the dirac delta function.

4.3 Transfer Function

The impulse response matrix describes the input-output relationship of a system in the time domain. For a linear time-invariant (LTI) system, we can also express this relationship in the frequency domain. For the LTI system in the previous section, we can express the impulse response matrix as

$$H(t,\tau) = \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B} + \boldsymbol{D}\delta(t-\tau).$$

Because the impulse response matrix is simply a function of $t-\tau$, we can write

$$H(t-\tau) = \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B} + \boldsymbol{D}\delta(t-\tau).$$

Defining the variable $t' := t-\tau$, we can express the impulse response matrix as

$$H(t') = \boldsymbol{C}e^{\boldsymbol{A}t'}\boldsymbol{B} + \boldsymbol{D}\delta(t')$$

Now to find a relationship between the input and output of this LTI system in the frequency domain, we need to compute the Laplace transform of the impulse response matrix. Doing so gives us the following **transfer function**:

$$H(s) = \boldsymbol{C}(s\boldsymbol{I_n} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}.$$

Notice that the transfer function is an $n_o \times n_i$ matrix. This matrix allows us to express the relationship between the input and output of an LTI system as

$$\hat{y}(s) = H(s)\hat{u}(s),$$

where $\hat{y}(s)$ represents the output function in the frequency domain and $\hat{u}(s)$ represents the input function in the frequency domain. Note that $\hat{y}(s)$ is the Laplace transform of the time-varying output function, $y(t)$, and $\hat{u}(s)$ is the Laplace transform of the time-varying input function, $u(t)$.

4.4 BIBO Stability Tests

Now that we have defined the impulse response matrix for continuous LTV systems and the transfer function for continuous LTI systems, we can use these matrices to determine whether the corresponding system is BIBO stable. In general, a continuous time linear system is BIBO stable if and only if

$$\sup_{t\in\mathbb{R}}\left\{\int_{-\infty}^{t} \|H(t,\tau)\|_\infty d\tau\right\} = k < \infty,$$

where $H(t,\tau)$ is the impulse response matrix for the given system.

4.4.1 Continuous LTV Systems

Recall that the impulse response matrix for a continuous LTV system is

$$H(t,\tau) = C(t)\phi(t,\tau)B(\tau) + D(t)\delta(t-\tau).$$

Therefore, a continuous LTV system is BIBO stable if and only if

1. $B(t)$, $C(t)$, and $D(t)$ are bounded for all values $t \in T$ and
2. $$\sup_{t\in\mathbb{R}}\left\{\int_{-\infty}^{t} \left\|C(t)\phi(t,\tau)B(\tau))\right\|_\infty d\tau\right\} = k < \infty.$$

4.4.2 Continuous LTI Systems

Recall that the impulse response matrix for a continuous LTI system is

$$H(t,\tau) = \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B} + \boldsymbol{D}\delta(t-\tau).$$

Therefore, a continuous LTI system is BIBO stable if and only if

$$\sup_{t\in\mathbb{R}}\left\{\int_{-\infty}^{t} \left\|\boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\right\|_\infty d\tau\right\} = k < \infty.$$

Let's define a new variable $s := t-\tau$. We can then express this condition as

$$\sup_{t\in\mathbb{R}}\left\{\int_{\infty}^{0} \left\|\boldsymbol{C}e^{\boldsymbol{A}s}\boldsymbol{B}\right\|_\infty (-1) ds\right\} = k < \infty.$$

Using properties of integration and noticing that the integral does not depend on t, we now obtain the following condition for BIBO stability:

$$\int_{0}^{\infty} \|\boldsymbol{C}e^{\boldsymbol{A}s}B\|_\infty ds = k < \infty.$$

As shown previously, we can also express the input-output relationship of a continuous LTI system in the frequency domain using the transfer function:

$$H(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \boldsymbol{C}(s\boldsymbol{I_n} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}.$$

We can use the transfer function to write another condition for BIBO stability. A continuous LTI system is BIBO stable if and only if all of the poles of the transfer function are in the open left half of the complex plane:

$$\text{Poles}(H(s)) \subset \mathbb{C}_{--}.$$

Example: To help see why this is true, consider the following transfer functions:

$$H_1(s) = \frac{1}{s-\alpha}, \qquad H_2(s) = \frac{1}{s^2+\omega^2}, \qquad H_3(s) = \frac{1}{s+\alpha} \qquad \text{where } \alpha > 0.$$

The first transfer function, $H_1(s)$, has one pole at $s = \alpha$, which is in the open right half of the complex plane. The second transfer function, $H_2(s)$, has two poles at $s = \pm j\omega$, which are both on the imaginary axis. The third transfer function, $H_3(s)$, has one pole at $s = -\alpha$, which is in the open left half of the complex plane. Based on our test for BIBO stability, we expect only the third transfer function to be BIBO stable. To see why this is, recall that we can express the input-output relationship for these three systems as

$$\hat{y}_1(s) = H_1(s)\hat{u}_1(s), \qquad \hat{y}_2(s) = H_2(s)\hat{u}_2(s), \qquad \hat{y}_3(s) = H_3(s)\hat{u}_3(s).$$

If we want to show that the first two systems are not BIBO stable, we need to find an example of bounded input that does not produce a bounded output. For the fist system, consider the bounded input $u_1(t) = 1$, which has the Laplace transform $\hat{u}_1(s) = \frac{1}{s}$. This input results in the output

$$\hat{y}_1(s) = \left(\frac{1}{s-\alpha}\right)\left(\frac{1}{s}\right) = \frac{1/\alpha}{s-\alpha} - \frac{1/\alpha}{s} = \frac{1}{\alpha}\left(\frac{1}{s-\alpha} - \frac{1}{s}\right).$$

Taking the inverse Laplace transform, we get the output in the time domain:

$$y_1(t) = \frac{1}{\alpha}\left(e^{\alpha t} - 1\right).$$

The output, $y_1(t)$, approaches infinity as t approaches infinity, so the bounded input $u_1(t) = 1$ produces an unbounded output. Therefore, the first system is not BIBO stable. For the second system, consider the bounded input $u_2(t) = \sin(\omega t)$, which has the Laplace transform $\frac{\omega}{s^2+\omega^2}$. This input results in the output

$$\hat{y}_2(s) = \left(\frac{1}{s^2+\omega^2}\right)\left(\frac{\omega}{s^2+\omega^2}\right) = \frac{\omega}{(s^2+\omega^2)^2}.$$

Taking the inverse Laplace transform, we get the output in the time domain:

$$y_2(t) = \frac{1}{2\omega^2}\Big(\sin(\omega t) - \omega t\cos(\omega t)\Big).$$

The output, $y_2(t)$, approaches infinity as t approaches infinity, so the bounded input $u_2(t) = sin(\omega t)$ produces an unbounded output. Therefore, the second system is also not BIBO stable. To show that the third system is BIBO stable, we need to show that the output is bounded for all bounded inputs. Taking the inverse Laplace transform of the equation $\hat{y}_3(s) = H_3(s)\hat{u}_3(s)$, we get

$$y_3(t) = e^{-\alpha t} * u_3(t) = \int_0^t e^{-\alpha\tau} u_3(t-\tau)d\tau$$

The l_∞ norm of this output is then given by

$$\begin{aligned} \|y_3(t)\|_\infty &= \left\|\int_0^t e^{-\alpha\tau}u_3(t-\tau)d\tau\right\|_\infty \le \int_0^t \left\|e^{-\alpha\tau}u_3(t-\tau)\right\|_\infty d\tau \\ &= \int_0^t \left|e^{-\alpha\tau}\right|\left\|u_3(t-\tau)\right\|_\infty d\tau = \int_0^t e^{-\alpha\tau}\left\|u_3(t)\right\|_\infty d\tau \\ &= \left(\int_0^t e^{-\alpha\tau}d\tau\right)\left\|u_3(t)\right\|_\infty = \left(1-e^{-\alpha t}\right)\left\|u_3(t)\right\|_\infty. \end{aligned}$$

Note that $\left\| u_3(t-\tau) \right\|_\infty = \left\| u_3(t) \right\|_\infty$ because the l_∞ norm is implicitly defined in terms of t and not in terms of τ. The expression $\left(1 - e^{-\alpha t}\right)$ is less than infinity for all values of t, so this system is BIBO stable, as we expected.

Chapter 5

State Space Stability

5.1 Equilibrium Points

We will also define another notion of stability, which we call state space stability. State space stability is defined in the context of the equilibrium points of the system. An **equilibrium point** of a dynamical system is a state for which the system is stationary for all future times. Once the system reaches an equilibrium point, it will remain at this point under the absence of any input. We will discuss equilibrium points for the common classes of systems defined in section 1.4

5.1.1 Continuous LTV Systems

Consider a continuous time system described by the differential equation

$$\dot{x}(t) = f(t, x(t)),$$

where $t \in \mathbb{R}$ and $x(t) \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$f(t, \tilde{\boldsymbol{x}}) = \mathbf{0}_n, \ \ \forall t \in \mathbb{R}.$$

Notice that if $\tilde{\boldsymbol{x}}$ is an equilibrium and $x(t) = \tilde{\boldsymbol{x}}$, then $\dot{x}(t) = \mathbf{0}_n$. This means that if the state of the system reaches an equilibrium point, the state will stop changing and the system will remain at this point for all future time. Consider a continuous linear time-varying system described by the equation

$$\dot{x}(t) = A(t)x(t),$$

where $t \in \mathbb{R}$ and $x(t) \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$A(t)\tilde{\boldsymbol{x}} = \mathbf{0}_n, \ \ \forall t \in \mathbb{R}.$$

Therefore, $\tilde{\boldsymbol{x}} = \mathbf{0}_n$ is always an equilibrium point for a continuous LTV system, regardless of the elements of the time-varying dynamics matrix, $A(t)$.

5.1.2 Continuous LTI Systems

Consider a continuous time system described by the differential equation

$$\dot{x}(t) = f(x(t)),$$

where $t \in \mathbb{R}$ and $x(t) \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$f(\tilde{\boldsymbol{x}}) = \mathbf{0}_n.$$

As in the time-varying case, if $\tilde{\boldsymbol{x}}$ is an equilibrium and $x(t) = \tilde{\boldsymbol{x}}$, then $\dot{x}(t) = \mathbf{0}_n$. This means that if the state of the system reaches an equilibrium point, the state will stop changing and the system will remain at this point for all future time. Consider a continuous linear time-invariant system described by the equation

$$\dot{x}(t) = \boldsymbol{A}x(t).$$

where $t \in \mathbb{R}$ and $x(t) \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$\boldsymbol{A}\tilde{\boldsymbol{x}} = \mathbf{0}_n.$$

As in the time-varying case, $\tilde{\boldsymbol{x}} = \mathbf{0}_n$ is always an equilibrium point for a continuous LTI system. If $\boldsymbol{A}$ is non-singular, then $\tilde{\boldsymbol{x}} = \mathbf{0}_n$ is the *unique* equilibrium. If $\boldsymbol{A}$ is singular, then the null space of $\boldsymbol{A}$ defines a continuum of equilibria.

5.1.3 Discrete LTV Systems

Consider a discrete time system described by the iterative equation

$$\boldsymbol{x_{k+1}} = f(k, \boldsymbol{x_k}),$$

where $k \in \mathbb{Z}$ and $\boldsymbol{x_k} \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$f(k, \tilde{\boldsymbol{x}}) = \tilde{\boldsymbol{x}}, \;\; \forall k \in \mathbb{Z}.$$

This says that the equilibrium, $\tilde{\boldsymbol{x}}$, is a fixed point of f for all discrete times. Notice that if $\tilde{\boldsymbol{x}}$ is an equilibrium and $\boldsymbol{x_k} = \tilde{\boldsymbol{x}}$, then $\boldsymbol{x_{k+1}} = \tilde{\boldsymbol{x}}$. This means that once the state of the system reaches the point $\tilde{\boldsymbol{x}}$, it will remain at this state for all future time. Consider a discrete linear time-varying system described by

$$\boldsymbol{x_{k+1}} = \boldsymbol{A_k}\boldsymbol{x_k},$$

where $k \in \mathbb{Z}$ and $\boldsymbol{x_k} \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$\boldsymbol{A_k}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}, \;\; \forall k \in \mathbb{Z}.$$

Therefore, as in the continuous case, $\tilde{\boldsymbol{x}} = \mathbf{0}_n$ is always an equilibrium point for a discrete LTV system, regardless of the time-varying elements of $\boldsymbol{A_k}$.

5.1.4 Discrete LTI Systems

Consider a discrete time system described by the iterative equation

$$\boldsymbol{x_{k+1}} = f(\boldsymbol{x_k}),$$

where $k \in \mathbb{Z}$ and $\boldsymbol{x_k} \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$f(\tilde{\boldsymbol{x}}) = \tilde{\boldsymbol{x}}.$$

This says that the equilibrium, $\tilde{\boldsymbol{x}}$, is a fixed point of f. As in the time-varying case, if $\tilde{\boldsymbol{x}}$ is an equilibrium and $\boldsymbol{x_k} = \tilde{\boldsymbol{x}}$, then $\boldsymbol{x_{k+1}} = \tilde{\boldsymbol{x}}$. This means that once the state of the system reaches the point $\tilde{\boldsymbol{x}}$, it will remain at this state for all future time. Consider a discrete linear time-varying system described by

$$\boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k},$$

where $k \in \mathbb{Z}$ and $\boldsymbol{x_k} \in \mathbb{R}^n$. The point $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ is an equilibrium if

$$\boldsymbol{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}.$$

As in the time-varying case, $\tilde{\boldsymbol{x}} = \boldsymbol{0}_n$ is always an equilibrium point for a discrete LTI system. Note that if $\boldsymbol{A} = \boldsymbol{I_n}$, then every point in $\mathbb{R}^n$ is an equilibrium point.

5.2 Definition of State Space Stability

Consider a system with the equilibrium point $\tilde{\boldsymbol{x}} = \boldsymbol{0}_n$. While state space stability is defined for individual equilibrium points, we typically say that a system is stable if its zero equilibrium point is stable. We will focus mostly on continuous linear systems whose state under the absence of input is given by $x(t) = \phi(t, t_0)\boldsymbol{x_0}$. For such a system, we have the following notions of stability:

1. **(Internal) Stability** – The system is (internally) stable if and only if for all initial times $t_0 \in \mathbb{R}$ and all initial states $\boldsymbol{x_0} := x(t_0) \in \mathbb{R}^n$, the state $x(t)$ is bounded, meaning that there exists a constant $M > 0$ such that

$$||x(t)||_2 = ||\phi(t, t_0)\boldsymbol{x_0}||_2 \leq ||\phi(t, t_0)||_2 ||\boldsymbol{x_0}||_2 \leq M, \ \forall t \geq t_0.$$

2. **Asymptotic Stability** – The system is asymptotically stable if and only if it is stable and if for all initial times $t_0 \in \mathbb{R}$ and all initial states $\boldsymbol{x_0} := x(t_0) \in \mathbb{R}^n$, the state $x(t)$ converges to the equilibrium, which means that

$$\lim_{t\to\infty} ||x(t)||_2 = \lim_{t\to\infty} ||\phi(t, t_0)\boldsymbol{x_0}||_2 = 0.$$

3. **Exponential Stability** – The system is exponentially stable if and only if the state, $x(t)$, is bounded by a decaying exponential, which places a requirement on the rate of convergence to the equilibrium. This means that there exists constants $M, \alpha > 0$ such that

$$||x(t)||_2 = ||\phi(t, t_0)\boldsymbol{x_0}||_2 \leq ||\phi(t, t_0)||_2 ||\boldsymbol{x_0}||_2 \leq M e^{-\alpha(t-t_0)} ||\boldsymbol{x_0}||_2, \ \forall t \geq t_0.$$

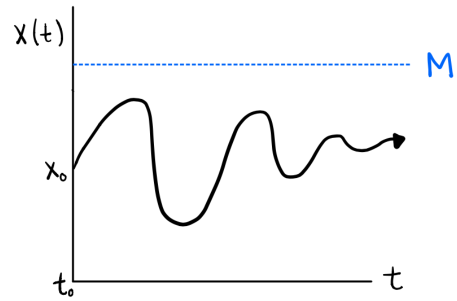

Figure 5.1: This system is (internally) stable. The state is bounded for all time but does not converge to the equilibrium.

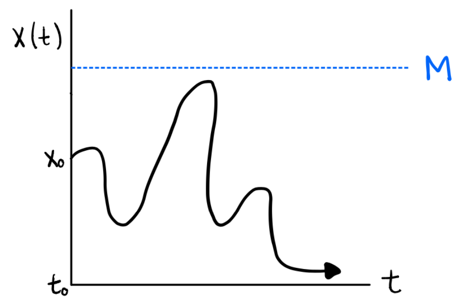

Figure 5.2: This system is asymptotically stable. The state converges to the equilibrium and is bounded for all time. There is no requirement on the rate of convergence.

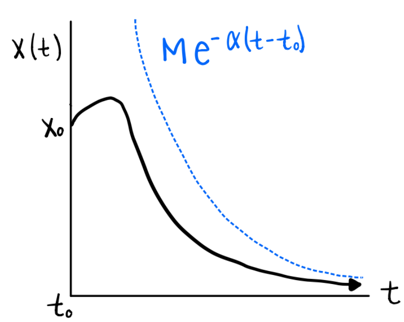

Figure 5.3: This system is exponentially stable. The state is bounded by a decaying exponential, which places a requirement on the rate of convergence.

In general, if a system is exponentially stable, it is asymptotically stable, and if it is asymptotically stable, it is (internally) stable. For a linear time-invariant (LTI) system, exponential and asymptotic stability are equivalent. Figures 5.1, 5.2, and 5.3 help demonstrate these different notions of state space stability.

5.3 State Space Stability Tests

There are various ways to check if a system is (internally) stable, asymptotically stable, or exponentially stable. The approach depends on whether the system is continuous or discrete and whether it is time-varying or time-invariant.

5.3.1 Continuous LTV Systems

Consider a continuous LTV system with the equilibrium point $\tilde{\boldsymbol{x}} = \mathbf{0}_n$ whose state under the absence of input is described by $\dot{x}(t) = A(t)x(t)$, where $x(t) \in \mathbb{R}^n$ and $A(t) \in \mathbb{R}^{n\times n}$. There are a couple cases of such systems we will consider.

Skew Symmetric System

Let's first consider the case when $A(t)$ is skew-symmetric (i.e. $A(t)^T = -A(t)$). Regardless of the specific entries of $A(t)$, this system is (internally) stable.

Proof: Consider the function $V : \mathbb{R}^n \to \mathbb{R}$ defined such that $V(x(t)) = x(t)^T x(t)$. If we compute the derivative of $V(x(t))$ with respect to t, we find

$$\begin{aligned}
\dot{V}(x(t)) &= \frac{d}{dt} x(t)^T x(t) \\
&= \dot{x}(t)^T x(t) + x(t)^T \dot{x}(t) \\
&= \big(A(t)x(t)\big)^T x(t) + x(t)^T \big(A(t)x(t)\big) \\
&= x(t)^T A(t)^T x(t) + x(t)^T A(t) x(t) \\
&= -x(t)^T A(t) x(t) + x(t)^T A(t) x(t) \\
&= 0
\end{aligned}$$

Now we have shown that derivative of $V(x(t))$ with respect to time is zero for any arbitrary state $x(t)$. If t_0 is the initial time, this implies that

$$V(x(t)) = V(x(t_0)), \;\; \forall t \geq t_0.$$

Based on how we defined V, we can equivalently write this equality as

$$||x(t)||_2^2 = ||x(t_0)||_2^2, \;\; \forall t \geq t_0.$$

Therefore, the continuous LTV system is stable if $A(t)$ is skew symmetric.

Symmetric System

For continuous time-varying systems, there is generally no connection between the eigenvalues of $A(t)$ and the state space stability. However, there are some cases when the eigenvalues can tell us about the stability of the system. In particular, we will consider the case when $A(t)$ is symmetric for all $t \in \mathbb{R}$.

If $A(t)$ is symmetric for all $t \in \mathbb{R}$, then all of its eigenvalues are real. Let $\lambda_1(t), \cdots, \lambda_n(t)$ denote the n real eigenvalues of $A(t)$. If $\lambda_i(t) \leq -\mu < 0$, for all $i = 1, \cdots, n$ and $t \in \mathbb{R}$, then the system is exponentially stable.

Proof: Consider the function $V : \mathbb{R}^n \to \mathbb{R}$ defined such that $V(x(t)) = x(t)^T x(t)$. If we compute the derivative of $V(x(t))$ with respect to t, we find

$$\begin{aligned}
\dot{V}(x(t)) &= \frac{d}{dt} x(t)^T x(t) \\
&= \dot{x}(t)^T x(t) + x(t)^T \dot{x}(t) \\
&= \big(A(t)x(t)\big)^T x(t) + x(t)^T \big(A(t)x(t)\big) \\
&= x(t)^T A(t)^T x(t) + x(t)^T A(t)x(t) \\
&= x(t)^T A(t)x(t) + x(t)^T A(t)x(t) \\
&= 2x(t)^T A(t)x(t)
\end{aligned}$$

Because $A(t)$ is symmetric, we can use the Rayleigh quotient to write

$$\lambda_{min}\big(A(t)\big)x(t)^T x(t) \leq x(t)^T A(t)x(t) \leq \lambda_{max}\big(A(t)\big)x(t)^T x(t).$$

This then gives us the following constraint on the time derivative of V:

$$\dot{V}(x(t)) \leq 2\lambda_{max}\big(A(t)\big)x(t)^T x(t).$$

We assumed that $\lambda_i(t) \leq -\mu < 0$ for all $i = 1, \cdots, n$ and $t \in \mathbb{R}$. Therefore,

$$\dot{V}(x(t)) \leq -2\mu(x(t)^T x(t)) = -2\mu V(x(t)).$$

Now we have a simple differential equation, which gives us the solution

$$V(x(t)) \leq V(x(t_0))e^{-2\mu t}.$$

From our definition of V, we can equivalently write this inequality as

$$x(t)^T x(t) \leq x(t_0)^T x(t_0)e^{-2\mu t}.$$

Taking the square root of both sides, this inequality becomes

$$\|x(t)\|_2 \leq e^{-\mu t}\|x(t_0)\|_2.$$

Because $\|x(t)\|_2$ is bounded by a decaying exponential, the given system is exponentially stable. Therefore, if $A(t)$ is a symmetric matrix and all its eigenvalues satisfy $\lambda_i(t) \leq -\mu < 0$ for all $t \in \mathbb{R}$, the system is exponentially stable.

5.3.2 Continuous LTI Systems

Consider a continuous linear time-invariant (LTI) system with the equilibrium point $\tilde{\boldsymbol{x}} = \boldsymbol{0}_n$ whose state under the absence of input is given by $\dot{x}(t) = \boldsymbol{A}x(t)$, where $x(t) \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. We can characterize the stability of this system based on the eigenvalues of the matrix $\boldsymbol{A}$ and by using Lyapunov equations.

Eigenvalue Test

Suppose that the dynamics matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has the eigenvalues $\lambda_1, \cdots, \lambda_n$. The eigenvalues of $\boldsymbol{A}$ tell us the following about the stability of the system:

1. This system is exponentially (and asymptotically) stable if and only if all of the eigenvalues of $\boldsymbol{A}$ are in the open left half of the complex plane, i.e.

$$\text{Re}\{\lambda_i\} < 0 \text{ for } i = 1, \cdots, n.$$

2. This system is (internally) stable if and only if all of the eigenvalues of $\boldsymbol{A}$ are in the closed left half plane, i.e.

$$\text{Re}\{\lambda_i\} \le 0 \text{ for } i = 1, \cdots, n,$$

and each eigenvalue on the $j\omega$-axis (i.e. λ_i whose real part is equal to zero) has a corresponding Jordan block of size one.

To see why this is true, notice that for a system described by $\dot{x}(t) = \boldsymbol{A}x(t)$, the solution is $x(t) = e^{\boldsymbol{A}t}\boldsymbol{x_0}$. If $\boldsymbol{A}$ admits the Jordan canonical form $\boldsymbol{A} = \boldsymbol{TJT}^{-1}$, then we can express this solution as $x(t) = \boldsymbol{T}e^{\boldsymbol{J}t}\boldsymbol{T}^{-1}\boldsymbol{x_0}$. Because $\boldsymbol{J}$ is a Jordan matrix and the function $e^{(\cdot)}$ is analytic, we can express $e^{\boldsymbol{J}t}$ as

$$e^{\boldsymbol{J}t} = \begin{bmatrix} e^{\boldsymbol{J_1}t} & & \\ & \ddots & \\ & & e^{\boldsymbol{J_l}t} \end{bmatrix}, \text{where } e^{\boldsymbol{J_i}t} = \begin{bmatrix} e^{\lambda_i t} & te^{\lambda_i t} & \cdots & \frac{1}{(n_i-1)!}t^{n_1-1}e^{\lambda_i t} \\ 0 & e^{\lambda_i t} & \cdots & \frac{1}{(n_i-2)!}t^{n_1-2}e^{\lambda_i t} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{\lambda_i t} \end{bmatrix}.$$

If the real part of λ_i is strictly negative, then all of the elements of $e^{\boldsymbol{J_i}t}$ will converge to zero as t goes to infinity. If the real part of λ_i is equal to zero and n_i is one, then the elements of $e^{\boldsymbol{J_i}t}$ will be bounded but will not converge to zero. If the real part of λ_i is strictly positive or if the real part of λ_i is equal to zero and n_i is greater than one, the elements of $e^{\boldsymbol{J_i}t}$ will grow without bound.

Therefore, if the real part of all the eigenvalues of $\boldsymbol{A}$ is strictly negative, then all of the elements of $e^{\boldsymbol{J}t}$ will converge to zero, thus $x(t)$ will also converge to zero. This is what we consider asymptotic stability, which is equivalent to exponential stability for LTI systems. If the real part of all the eigenvalues of $\boldsymbol{A}$ is non-positive and n_i is one for all of the eigenvalues with zero real part, then all of the elements of $e^{\boldsymbol{J}t}$ will be bounded but will not necessarily converge to zero, thus $x(t)$ will also be bounded but will not necessarily converge to zero. Therefore, we say the system is (internally) stable.

Lyapunov Equations

A Lyapunov equation is a matrix equation of the form

$$\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} = -\boldsymbol{Q}, \text{ where } \boldsymbol{A}, \boldsymbol{P}, \boldsymbol{Q} \in \mathbb{R}^{n\times n}.$$

If $\boldsymbol{A}$ is a dynamics matrix, we can use Lyapunov equations to characterize the stability of the corresponding continuous LTI system.

Theorem: The system described by $\dot{x}(t) = \boldsymbol{A}x(t)$ is exponentially stable if there exist positive definite matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ that satisfy the Lyapunov equation.

We will prove this theorem in two different ways. First, we will prove the theorem using the original definition of exponential stability. We will then prove it using the eigenvalue test discussed in the previous subsection.

Proof 1 (Definition of Stability): Given a positive definite matrix $\boldsymbol{Q} \in \mathbb{S}^n_{++}$, let $\boldsymbol{P} \in \mathbb{S}^n_{++}$ be the solution to the Lyapunov equation $\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} = -\boldsymbol{Q}$. Now consider the function $V : \mathbb{R}^n \to \mathbb{R}$, which is defined as $V\big(x(t)\big) = x(t)^T\boldsymbol{P}x(t)$. If we compute the derivative of $V\big(x(t)\big)$ with respect to t, we find

$$\begin{aligned}
\dot{V}\big(x(t)\big) &= \frac{d}{dt}\Big(x(t)^T\boldsymbol{P}x(t)\Big) \\
&= \dot{x}(t)^T\boldsymbol{P}x(t) + x(t)^T\boldsymbol{P}\dot{x}(t) \\
&= \big(\boldsymbol{A}x(t)\big)^T\boldsymbol{P}x(t) + x(t)^T\boldsymbol{P}\big(\boldsymbol{A}x(t)\big) \\
&= x(t)^T\boldsymbol{A}^T\boldsymbol{P}x(t) + x(t)^T\boldsymbol{P}\boldsymbol{A}x(t) \\
&= x(t)^T\big(\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\big)x(t) \\
&= -x(t)^T\boldsymbol{Q}x(t).
\end{aligned}$$

Because the matrices $\boldsymbol{P}$ are $\boldsymbol{Q}$ are both symmetric, we can use the Rayleigh quotient to write the following inequalities:

$$\lambda_{min}(\boldsymbol{P})||x(t)||_2^2 \leq x(t)^T\boldsymbol{P}x(t) \leq \lambda_{max}(\boldsymbol{P})||x(t)||_2^2$$

$$\lambda_{min}(\boldsymbol{Q})||x(t)||_2^2 \leq x(t)^T\boldsymbol{Q}x(t) \leq \lambda_{max}(\boldsymbol{Q})||x(t)||_2^2$$

Because we defined $V\big(x(t)\big) = x(t)^T\boldsymbol{P}x(t)$ and found that $\dot{V}\big(x(t)\big) = -x(t)^T\boldsymbol{Q}x(t)$, the Rayleigh quotient also allows us to write the following set of inequalities:

$$\lambda_{min}(\boldsymbol{P})||x(t)||_2^2 \leq V\big(x(t)\big) \leq \lambda_{max}(\boldsymbol{P})||x(t)||_2^2$$

$$\lambda_{min}(\boldsymbol{Q})||x(t)||_2^2 \leq -\dot{V}\big(x(t)\big) \leq \lambda_{max}(\boldsymbol{Q})||x(t)||_2^2$$

From these two sets of inequalities, we can pull out the following inequalities:

$$\frac{V\big(x(t)\big)}{\lambda_{max}(\boldsymbol{P})} \leq ||x(t)||_2^2 \quad \text{and} \quad ||x(t)||_2^2 \leq \frac{-\dot{V}\big(x(t)\big)}{\lambda_{min}(\boldsymbol{Q})}.$$

Combining these inequalities, we see the following relationship:

$$\frac{V\big(x(t)\big)}{\lambda_{max}(\boldsymbol{P})} \leq \frac{-\dot{V}\big(x(t)\big)}{\lambda_{min}(\boldsymbol{Q})}.$$

Rearranging, we get the following relationship between $\dot{V}\big(x(t)\big)$ and $V\big(x(t)\big)$:

$$\dot{V}\big(x(t)\big) \leq -\frac{\lambda_{min}(\boldsymbol{Q})}{\lambda_{max}(\boldsymbol{P})} V\big(x(t)\big).$$

Now we have a simple differential equation, which gives us the solution

$$V\big(x(t)\big) \leq V(\boldsymbol{x_0}) \exp\left(-\frac{\lambda_{min}(\boldsymbol{Q})}{\lambda_{max}(\boldsymbol{P})} t\right).$$

Because $V\big(x(t)\big) \geq \lambda_{min}(\boldsymbol{P})\|x(t)\|_2^2$ and $V(\boldsymbol{x_0}) \leq \lambda_{max}(\boldsymbol{P})\|\boldsymbol{x_0}\|_2^2$, we can write

$$\lambda_{min}(\boldsymbol{P})\|x(t)\|_2^2 \leq \lambda_{max}(\boldsymbol{P})\|\boldsymbol{x_0}\|_2^2 \exp\left(-\frac{\lambda_{min}(\boldsymbol{Q})}{\lambda_{max}(\boldsymbol{P})} t\right)$$

$$\|x(t)\|_2^2 \leq \frac{\lambda_{max}(\boldsymbol{P})}{\lambda_{min}(\boldsymbol{P})}\|\boldsymbol{x_0}\|_2^2 \exp\left(-\frac{\lambda_{min}(\boldsymbol{Q})}{\lambda_{max}(\boldsymbol{P})} t\right)$$

$$\|x(t)\|_2 \leq \left(\frac{\lambda_{max}(\boldsymbol{P})}{\lambda_{min}(\boldsymbol{P})}\right)^{1/2} \exp\left(-\frac{1}{2}\frac{\lambda_{min}(\boldsymbol{Q})}{\lambda_{max}(\boldsymbol{P})} t\right)\|\boldsymbol{x_0}\|_2$$

Because $\boldsymbol{Q}$ and $\boldsymbol{P}$ are positive definite, all of their eigenvalues must be strictly positive, which means $\lambda_{min}(\boldsymbol{Q}) > 0$ and $\lambda_{max}(\boldsymbol{P}) \geq \lambda_{min}(\boldsymbol{P}) > 0$. Therefore, $\|x(t)\|_2$ is bounded by a decaying exponential, which implies that the LTI system is exponentially stable. Therefore, if there exist positive definite matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ that satisfy the Lyapunov equation for the matrix $\boldsymbol{A}$ describing the dynamics of the LTI system, then the system is exponentially stable.

Proof 2 (Eigenvalue Test): Recall that the system is exponentially stable if and only if all the eigenvalues of $\boldsymbol{A}$ are in the open left half of the complex plane. Assume that $\boldsymbol{A}$ has an eigenvector $\boldsymbol{v}$ corresponding to eigenvalue λ, which implies $\boldsymbol{Av} = \lambda\boldsymbol{v}$ and $\boldsymbol{v}^T\boldsymbol{A}^T = \bar{\lambda}\boldsymbol{v}^T$. This allows us to write the following:

$$\begin{aligned}
-\boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{v} &= \boldsymbol{v}^T\big(\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\big)\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{P}\boldsymbol{A}\boldsymbol{v} \\
&= (\boldsymbol{Av})^T\boldsymbol{P}\boldsymbol{v} + \boldsymbol{v}^T P(\boldsymbol{Av}) = (\lambda\boldsymbol{v})^T\boldsymbol{P}\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{P}(\lambda\boldsymbol{v}) \\
&= \lambda\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v} + \lambda\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v} = 2\lambda\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v}
\end{aligned}$$

Because $\boldsymbol{Q}$ and $\boldsymbol{P}$ are positive definite matrices, $\boldsymbol{v}^T\boldsymbol{Q}\boldsymbol{v} > 0$ and $\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v} > 0$ for all $\boldsymbol{v} \neq \boldsymbol{0}_n$, which implies that the real part of λ is strictly negative. This holds for all eigenvalue-eigenvector pairs of $\boldsymbol{A}$, so all the eigenvalues of $\boldsymbol{A}$ have strictly negative real parts. Therefore, if there exist positive definite matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ that satisfy the Lyapunov equation, all of the eigenvalues of $\boldsymbol{A}$ are in the open left half plane, which implies that the system is exponentially stable.

5.3.3 Discrete LTI Systems

Consider a discrete linear time-invariant (LTI) system with the equilibrium point $\tilde{\boldsymbol{x}} = \mathbf{0}_n$ whose state under the absence of any input is described by $\boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k}$, where $\boldsymbol{x_k} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has eigenvalues $\lambda_1, \cdots, \lambda_n$. The eigenvalues of $\boldsymbol{A}$ tell us the following about the stability of the discrete LTI system:

1. This system is exponentially (and asymptotically) stable if and only if all of the eigenvalues of $\boldsymbol{A}$ fall within the unit circle in the complex plane, i.e.

$$|\lambda_i| < 1 \text{ for } i = 1, \cdots, n.$$

2. This system is (internally) stable if and only if all of the eigenvalues of $\boldsymbol{A}$ fall on or within the unit circle, i.e.

$$|\lambda_i| \leq 1 \text{ for } i = 1, \cdots, n,$$

and each eigenvalue on the unit circle (i.e. λ_i whose magnitude is equal to one) has a corresponding Jordan block of size one.

To see why this is true, notice that for a system described by $\boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k}$, the solution is $\boldsymbol{x_k} = \boldsymbol{A}^k\boldsymbol{x_0}$. If $\boldsymbol{A}$ admits the Jordan canonical form $\boldsymbol{A} = \boldsymbol{T}\boldsymbol{J}\boldsymbol{T}^{-1}$, then we can express this solution as $\boldsymbol{x_k} = \boldsymbol{T}\boldsymbol{J}^k\boldsymbol{T}^{-1}\boldsymbol{x_0}$. Because $\boldsymbol{J}$ is a Jordan matrix and the function $(\cdot)^k$ is analytic, we can express $\boldsymbol{J}^k$ as

$$\boldsymbol{J}^k = \begin{bmatrix} \boldsymbol{J_1}^k & & \\ & \ddots & \\ & & \boldsymbol{J_l}^k \end{bmatrix}, \text{ where}$$

$$\boldsymbol{J_i}^k = \begin{bmatrix} \lambda_i^k & k\lambda_i^{k-1} & \cdots & \frac{1}{(n_i-1)!}\prod_{j=1}^{n_i-1}(k-j+1)\lambda_i^{(k-n_i+1)} \\ 0 & \lambda_i^k & \cdots & \frac{1}{(n_i-2)!}\prod_{j=1}^{n_i-2}(k-j+1)\lambda_i^{(k-n_i+2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_i^k \end{bmatrix}.$$

If the magnitude of λ_i is strictly less than one, all the elements of $\boldsymbol{J_i}^k$ will converge to zero as k goes to infinity. If the magnitude of λ_i is equal to one and n_i is one, the elements of $\boldsymbol{J_i}^k$ will be bounded but will not converge. If the magnitude of λ_i is strictly greater than one or if the magnitude of λ_i is equal to one and n_i is greater than one, the elements of $\boldsymbol{J_i}^k$ will grow without bound.

Therefore, if the magnitude of all the eigenvalues of $\boldsymbol{A}$ is strictly less than one, then all of the elements of $\boldsymbol{J}^k$ will converge to zero, thus $\boldsymbol{x_k}$ will also converge to zero. This is what we consider asymptotic stability, which is equivalent to exponential stability for LTI systems. If the magnitude of all the eigenvalues of $\boldsymbol{A}$ is less than or equal to one and n_i is one for all of the eigenvalues with a magnitude of one, then all of the elements of $\boldsymbol{J}^k$ will be bounded but will not necessarily converge to zero, thus $\boldsymbol{x_k}$ will also be bounded but will not necessarily converge to zero. Therefore, we say the system is (internally) stable.

5.4 State Space vs. BIBO Stability

At the beginning of this chapter, we defined three different notions of state space stability. We said that a system is (internally) stable if the state remains bounded for all time in the absence of input. A system is asymptotically stable if the state converges to its equilibrium in the absence of input. Finally, a system is exponentially stable if the state converges to its equilibrium at an exponential rate in the absence of input. Recall that in Chapter 4, we said a system is considered to be bounded-input bounded-output (BIBO) stable if bounded inputs produce bounded outputs under zero initial condition. We will now consider the relationship between state space stability and BIBO stability.

5.4.1 Implications of Stability

For a linear time-varying (LTV) system, if the system is exponentially stable and the matrices $B(\cdot)$, $C(\cdot)$, and $D(\cdot)$ are all bounded, then the system is BIBO stable. For a linear time-invariant (LTI) system, if the system is exponentially stable, then it is BIBO stable. For both LTI and LTV systems, a system that is BIBO stable is not necessarily exponentially stable, or even (internally) stable.

Recall that for an LTI system, the system is BIBO stable if and only if all of the poles of the transfer function are in the open left half plane. Similarly, the system is exponentially stable if and only if all of the eigenvalues of $\boldsymbol{A}$ are in the open left half plane. All of the poles of the transfer function are eigenvalues of $\boldsymbol{A}$, but not all of the eigenvalues of $\boldsymbol{A}$ are poles of the transfer function. For this reason, exponential stability implies BIBO stability, but BIBO stability does not imply exponential stability and also cannot not imply (internal) stability.

5.4.2 Uncontrollable Modes

If a mode corresponding to an unstable eigenvalue is uncontrollable from the input, $u(t)$, then this mode does not appear in the transfer function. The unstable dynamics are effectively hidden from the input by the matrix $\boldsymbol{B}$.

Example: Consider the system described by $\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t)$, whose output is given by $y(t) = \boldsymbol{C}x(t)$. Let's define the dynamics matrices as follows:

$$\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1/10 \\ 0 \\ 1/15 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

The matrix $\boldsymbol{A}$ has three eigenvalues: $\lambda_1 = -1$, $\lambda_2 = 3$, and $\lambda_3 = -6$. This system is clearly not (internally) stable because the eigenvalue $\lambda_2 = 3$ is in the right half plane. To determine if this system is BIBO stable, we need to consider its transfer function. Recall that the transfer function is defined as

$$
\begin{aligned}
G(s) &= \boldsymbol{C}(s\boldsymbol{I}_n - \boldsymbol{A})^{-1}\boldsymbol{B} \\
&= \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{s+1} & 0 & 0 \\ 0 & \frac{1}{s-3} & 0 \\ 0 & 0 & \frac{1}{s+6} \end{bmatrix} \begin{bmatrix} 1/10 \\ 0 \\ 1/15 \end{bmatrix} \\
&= \frac{1}{10}\left(\frac{1}{s+1}\right) + \frac{1}{15}\left(\frac{1}{s+6}\right) \\
&= \frac{1}{6}\frac{s+4}{(s+1)(s+6)}.
\end{aligned}
$$

Now we can see that the poles of the transfer function are $\lambda_1 = -1$ and $\lambda_3 = -6$, which are both in the open left half plane. The unstable mode corresponding to $\lambda_2 = 3$ is said to be uncontrollable from the input because it is hidden by the matrix $\boldsymbol{B}$, and thus does not appear in the transfer function.

5.4.3 Unobservable Modes

Similarly, if a mode corresponding to an unstable eigenvalue is unobservable from the output, $y(t)$, then this mode does not appear in the transfer function. The unstable dynamics are effectively hidden from the output by the matrix $\boldsymbol{C}$.

Example: Consider the system described by $\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t)$, whose output is given by $y(t) = \boldsymbol{C}x(t)$. Let's define the dynamics matrices as follows:

$$
\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -6 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 1/10 \\ -1/6 \\ 1/15 \end{bmatrix}, \qquad \boldsymbol{C} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}.
$$

Again, $\boldsymbol{A}$ has three eigenvalues: $\lambda_1 = -1$, $\lambda_2 = 3$, and $\lambda_3 = -6$, and the system is not (internally) stable because the eigenvalue $\lambda_2 = 3$ is in the right half plane. To determine if this system is BIBO stable, consider its transfer function:

$$
\begin{aligned}
G(s) &= \boldsymbol{C}(s\boldsymbol{I}_n - \boldsymbol{A})^{-1}\boldsymbol{B} \\
&= \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{s+1} & 0 & 0 \\ 0 & \frac{1}{s-3} & 0 \\ 0 & 0 & \frac{1}{s+6} \end{bmatrix} \begin{bmatrix} 1/10 \\ -1/6 \\ 1/15 \end{bmatrix} \\
&= \frac{1}{10}\left(\frac{1}{s+1}\right) + \frac{1}{15}\left(\frac{1}{s+6}\right) \\
&= \frac{1}{6}\frac{s+4}{(s+1)(s+6)}
\end{aligned}
$$

Now we can see that the poles of the transfer function are $\lambda_1 = -1$ and $\lambda_3 = -6$, which are both in the open left half plane. The unstable mode corresponding to $\lambda_2 = 3$ is said to be unobservable from the output because it is hidden by the matrix $\boldsymbol{C}$, and thus does not appear in the transfer function.

Part III

Controllability & Observability

Chapter 6

Controllability/Observability: Continuous LTV Systems

6.1 Controllability of Continuous LTV Systems

Informally, a control system is considered controllable if its input can transform its initial state to any arbitrary state in the configuration space. We will define controllability more formally for continuous linear time-varying (LTV) systems.

6.1.1 Controllability Map

A dynamical system $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$ is (completely) controllable on the time interval $[t_0, t_1]$ if and only if for all $\boldsymbol{x_0}, \boldsymbol{x_1} \in X$, there exists an input function $u \in \mathcal{U}$ that steers $\boldsymbol{x_0} := x(t_0)$ to $\boldsymbol{x_1} = x(t_1)$. Equivalently, $\mathcal{D}$ is (completely) controllable on $[t_0, t_1]$ if and only if for all $\boldsymbol{x_0} \in X$, the state transition function $s(t_1, t_0, \boldsymbol{x_0}, u)$ is *surjective*. Recall from Section 3.2.1 that for a continuous LTV system, the state transition function is given by

$$s(t_1, t_0, \boldsymbol{x_0}, u) = \phi(t_1, t_0)\boldsymbol{x_0} + \int_{t_0}^{t_1} \phi(t_1, \tau)B(\tau)u(\tau)d\tau.$$

If we define the **controllability map** as the function $\mathcal{L}_C : \mathcal{U} \to X$ such that

$$\mathcal{L}_C(u) = \int_{t_0}^{t_1} \phi(t_1, \tau)B(\tau)u(\tau)d\tau,$$

then we can express the state transition function as

$$s(t_1, t_0, \boldsymbol{x_0}, u) = \phi(t_1, t_0)\boldsymbol{x_0} + \mathcal{L}_C(u).$$

Therefore, the system is (completely) controllable if and only if the controllability map, $\mathcal{L}_C$, is a surjective funtion. Based on our definition of surjectivity, this implies that the system is (completely) controllable if and only if $R(\mathcal{L}_C) = X$. We will assume that $X = \mathbb{R}^n$, as is common for linear systems.

6.1.2 Controllability Grammian

The adjoint of the controllability map is the map $\mathcal{L}_C^* : X \to \mathcal{U}$ that satisfies

$$\langle \boldsymbol{x}, \mathcal{L}_C(u) \rangle_X = \langle \mathcal{L}_C^*(\boldsymbol{x}), u \rangle_{\mathcal{U}},$$

where we let $X = \mathbb{R}^n$ and $\mathcal{U} = L_2([t_0, t_1], \mathbb{R}^{n_i})$. To obtain a closed-form expression for the adjoint of the controllability map, $\mathcal{L}_C^*$, we can look more closely at the expression on the left-hand side of the definition of the adjoint:

$$\begin{aligned}
\langle \boldsymbol{x}, \mathcal{L}_C(u) \rangle_X &= \boldsymbol{x}^T \mathcal{L}_C(u) \\
&= \boldsymbol{x}^T \left(\int_{t_0}^{t_1} \phi(t_1, \tau) B(\tau) u(\tau) d\tau \right) \\
&= \int_{t_0}^{t_1} \boldsymbol{x}^T \phi(t_1, \tau) B(\tau) u(\tau) d\tau \\
&= \int_{t_0}^{t_1} \Big(B^*(\tau) \phi^*(t_1, \tau) \boldsymbol{x} \Big)^* u(\tau) d\tau \\
&= \langle B^*(\cdot) \phi^*(t_1, \cdot) \boldsymbol{x}, u \rangle_{\mathcal{U}}
\end{aligned}$$

Comparing this expression to the right-hand side of the definition of the adjoint, we can conclude that the adjoint $\mathcal{L}_C^*$ is defined such that

$$\mathcal{L}_C^*(\boldsymbol{x}) = B^*(\cdot) \phi^*(t_1, \cdot) \boldsymbol{x}.$$

We define the **controllability grammian** as the function $W_C[t_0, t_1] : X \to X$, where we assume $X = \mathbb{R}^n$, such that

$$W_C[t_0, t_1] = \mathcal{L}_C \mathcal{L}_C^* = \int_{t_0}^{t_1} \phi(t_1, \tau) B(\tau) B^*(\tau) \phi^*(t_1, \tau) d\tau.$$

From the fundamental theorem of linear algebra, $R(\mathcal{L}_C) = R(\mathcal{L}_C \mathcal{L}_C^*)$. Therefore, the system is (completely) controllable on $[t_0, t_1]$ if and only if

$$R(W_C[t_0, t_1]) = R(\mathcal{L}_C \mathcal{L}_C^*) = R(\mathcal{L}_C) = X = \mathbb{R}^n.$$

Notice that the controllability grammian is always positive semidefinite. Therefore, its range is equal to $\mathbb{R}^n$ if and only if it is positive definite. We can then say that the system is (completely) controllable on $[t_0, t_1]$ if and only if the controllability grammian, $W_C[t_0, t_1]$, is positive definite.

6.1.3 Controllable Subspaces

The **controllable subspace** is the region in the state space that can be reached from some initial condition. If the system is controllable, then the controllable subspace is the entire state space (usually $\mathbb{R}^n$).

Previously, we said that for a continuous LTV system with initial state $\boldsymbol{x_0}$ at time t_0, we can express the state of the system at time t_1 as

$$x(t_1) = \phi(t_1, t_0) \boldsymbol{x_0} + \mathcal{L}_C(u),$$

where $\mathcal{L}_C$ is the controllability map defined in Section 6.1.1. This tells us that the set of states that are reachable from the initial point $\boldsymbol{x_0}$ is

$$\phi(t_1, t_0)\boldsymbol{x_0} + R(\mathcal{L}_C).$$

We call this set the reachable/controllable subspace. A state $\boldsymbol{x_1}$ is thus reachable at time t_1 from the initial state $\boldsymbol{x_0}$ at time t_0 if and only

$$\Big(\boldsymbol{x_1} - \phi(t_1, t_0)\boldsymbol{x_0}\Big) \in R(\mathcal{L}_C).$$

Note that if the system is controllable, then the range of the controllability map is the entire state space and every vector $\boldsymbol{x_1}$ is reachable.

6.1.4 Computing the Optimal Control

Suppose we want to find the best input sequence that allows us to move the state of our system from the state $\boldsymbol{x_0}$ at time t_0 to the state $\boldsymbol{x_1}$ at time t_1. As shown in the previous section, if $\boldsymbol{x_1}$ is in the controllable subspace, then there exists a solution $u \in \mathcal{U}$ to the linear equation $\boldsymbol{x_1} = \phi(t_1, t_0)\boldsymbol{x_0} + \mathcal{L}_C(u)$.

Based on our definition of controllability, if the system is controllable on the time interval $[t_0, t_1]$, then the controllability map $\mathcal{L}_C$ is surjective and $R(\mathcal{L}_C) = X = \mathbb{R}^n$. This means that if the system is controllable, then $\boldsymbol{x_1}$ is always in the controllable subspace, so there is always a solution to the linear equation. From linear matrix equations, the minimum norm solution is given by

$$u = \mathcal{L}_C^*(\mathcal{L}_C\mathcal{L}_C^*)^{-1}\Big(\boldsymbol{x_1} - \phi(t_1, t_0)\boldsymbol{x_0}\Big).$$

If the system is not controllable on the time onterval $[t_0, t_1]$, then the map $\mathcal{L}_C$ is not surjective and $R(\mathcal{L}_C) \subset X = \mathbb{R}^n$. If $\boldsymbol{x_1}$ is still in the controllable subspace, then the minimum norm solution is now given by

$$u = \mathcal{L}_C^*(\mathcal{L}_C\mathcal{L}_C^*)^{\dagger}\Big(\boldsymbol{x_1} - \phi(t_1, t_0)\boldsymbol{x_0}\Big).$$

If $\boldsymbol{x_1}$ is not in the controllable subspace, then there is no input $u \in \mathcal{U}$ that can steer the system from the initial state $\boldsymbol{x_0}$ to $\boldsymbol{x_1}$. Instead, we can choose a control that steers the system from $\boldsymbol{x_0}$ to a state $\hat{\boldsymbol{x}}_1$, which is the closest state to $\boldsymbol{x_1}$ within the controllable subspace. Now the minimum norm solution is

$$\hat{u} = \mathcal{L}_C^*(\mathcal{L}_C\mathcal{L}_C^*)^{\dagger}\Big(x(t_1) - \phi(t_1, t_0)\boldsymbol{x_0}\Big).$$

6.2 Observability of Continuous LTV Systems

Informally, a control system is considered observable if its initial state can be uniquely determined by observing its inputs and outputs. We will define observability more formally for continuous linear time-varying (LTV) systems.

6.2.1 Observability Map

A dynamical system $\mathcal{D} = (\mathcal{U}, \mathcal{X}, \mathcal{Y}, s, r)$ is (completely) observable on the time interval $[t_0, t_1]$ if and only if for all input functions $u \in \mathcal{U}$ and for all output functions $y \in \mathcal{Y}$, the initial state $\boldsymbol{x_0} := x(t_0)$ can be uniquely determined. Equivalently, $\mathcal{D}$ is (completely) observable on $[t_0, t_1]$ if and only if for all outputs $y \in \mathcal{Y}$, the response function $\rho(t_1, t_0, \boldsymbol{x_0}, u)$ is injective. Recall that for a linear time-varying (LTV) system, the response function is given by

$$\rho(t_1, t_0, \boldsymbol{x_0}, u) = C(t_1)\phi(t_1, t_0)\boldsymbol{x_0} + \int_{t_0}^{t_1} C(t_1)\phi(t_1, \tau)B(\tau)u(\tau)d\tau + D(t_1)u(t_1).$$

If we define the **observability map** as the function $\mathcal{L}_O : X \to \mathcal{Y}$ such that

$$\mathcal{L}_O(\boldsymbol{x_0}) = C(\cdot)\phi(\cdot, t_0)\boldsymbol{x_0},$$

then we can express the response function as

$$\rho(t_1, t_0, \boldsymbol{x_0}, u) = \mathcal{L}_O(\boldsymbol{x_0}) + \int_{t_0}^{t_1} C(t_1)\phi(t_1, \tau)B(\tau)u(\tau)d\tau + D(t_1)u(t_1).$$

Therefore, the system is (completely) observable if and only if the observability map, $\mathcal{L}_O$, is an injective function. Based on our definition of injectivity, this implies that the system is (completely) observable if and only if $N(\mathcal{L}_O) = \{\mathbf{0}_n\}$.

6.2.2 Observability Grammian

The adjoint of the observability map $\mathcal{L}_O$ is the map $\mathcal{L}_O^* : \mathcal{Y} \to X$ that satisfies

$$\langle y, \mathcal{L}_O(\boldsymbol{x_0})\rangle_{\mathcal{Y}} = \langle \mathcal{L}_O^*(y), \boldsymbol{x_0}\rangle_X,$$

where we let $X = \mathbb{R}^n$ and $\mathcal{Y} = L_2([t_0, t_1], \mathbb{R}^{n_o})$. To obtain a closed-form expression for the adjoint of the observability map, $\mathcal{L}_O^*$, we can look more closely at the expression on the left-hand side of the definition of the adjoint:

$$\begin{aligned}
\langle y, \mathcal{L}_O(\boldsymbol{x_0})\rangle_{\mathcal{Y}} &= \int_{t_0}^{t_1} y^*(\tau)\mathcal{L}_O(\boldsymbol{x_0})d\tau \\
&= \int_{t_0}^{t_1} y^*(\tau)\Big(C(\tau)\phi(\tau, t_0)\boldsymbol{x_0}\Big)d\tau \\
&= \int_{t_0}^{t_1} \Big(\phi^*(\tau, t_0)C^*(\tau)y(\tau)\Big)^* \boldsymbol{x_0} d\tau \\
&= \left(\int_{t_0}^{t_1} \phi^*(\tau, t_0)C^*(\tau)y(\tau)d\tau\right)^* \boldsymbol{x_0} \\
&= \left\langle \left(\int_{t_0}^{t_1} \phi^*(\tau, t_0)C^*(\tau)y(\tau)d\tau\right), \boldsymbol{x_0}\right\rangle_X
\end{aligned}$$

Comparing this expression to the right-hand side of the definition of the adjoint, we can conclude that $\mathcal{L}_O^*$ is defined such that

$$\mathcal{L}_O^*(y) = \int_{t_0}^{t_1} \phi^*(\tau, t_0)C^*(\tau)y(\tau)d\tau.$$

We define the **observability grammian** as the function $W_O[t_0, t_1] : X \to X$, where we assume $X = \mathbb{R}^n$, such that

$$W_O[t_0, t_1] = \mathcal{L}_O^*\mathcal{L}_O = \int_{t_0}^{t_1} \phi^*(\tau, t_0)C^*(\tau)C(\tau)\phi(\tau, t_0)d\tau.$$

From the fundamental theorem of linear algebra, $N(\mathcal{L}_O) = N(\mathcal{L}_O^*\mathcal{L}_O)$. Therefore, the system is (completely) observable on $[t_0, t_1]$ if and only if

$$N(W_O[t_0, t_1]) = N(\mathcal{L}_O^*\mathcal{L}_O) = N(\mathcal{L}_O) = \{\mathbf{0}_n\}.$$

By the rank-nullity theorem, this implies that the system is (completely) observable on $[t_0, t_1]$ if and only if $R(W_O[t_0, t_1]) = X = \mathbb{R}^n$. The observability grammian is always positive semidefinite, so its range is equal to $\mathbb{R}^n$ if and only if it is positive definite. Therefore, the system is (completely) observable on $[t_0, t_1]$ if and only if the observability grammian, $W_O[t_0, t_1]$, is positive definite.

6.2.3 Observable Subspaces

The **observable subspace** is the region in the state space for which initial conditions in this space can be uniquely determined by observing the inputs and outputs of the system over time. If the system is observable, then the observable subspace is the entire state space (usually $\mathbb{R}^n$).

Previously, we said that for a continuous LTV system with initial state $\boldsymbol{x_0}$ at time t_0, we can express the output of the system at time t as

$$y(t) = \mathcal{L}_O(\boldsymbol{x_0}) + f(u, t), \text{ where}$$

$$f(u, t) \coloneqq \int_{t_0}^{t} C(t)\phi(t, \tau)B(\tau)u(\tau)d\tau + D(t)u(t).$$

If $(\boldsymbol{x_0} - \bar{\boldsymbol{x}}_{\mathbf{0}}) \in N(\mathcal{L}_0)$, then $\mathcal{L}_O(\boldsymbol{x_0}) = \mathcal{L}_O(\bar{\boldsymbol{x}}_{\mathbf{0}})$, so the initial conditions $\boldsymbol{x_0}$ and $\bar{\boldsymbol{x}}_{\mathbf{0}}$ are not distinguishable for a given input $u \in \mathcal{U}$ and output $y \in \mathcal{Y}$. Furthermore, if $\boldsymbol{x_0} \in N(\mathcal{L}_0)$, then the initial condition $\boldsymbol{x_0}$ is not distinguishable from the zero vector $\mathbf{0}_n$ and thus cannot be uniquely determined for the given input and output functions. For this reason, we call $N(\mathcal{L}_0)$ the **unobservable subspace**. The **observable subspace** is then defined as $X \setminus N(\mathcal{L}_0)$.

6.2.4 Determining the Initial State

Suppose we want to determine the initial state of our system after observing a series of outputs resulting from known inputs. To do so, recall that in the previous section, we express the output of our system as

$$y(t) = \mathcal{L}_O(\boldsymbol{x_0}) + f(u, t).$$

Because the observability map is a linear function, a solution to the above equation exists if and only if $y(t) - f(u,t)$ is an element of the range space of $\mathcal{L}_O$. This solution is unique if and only if the map $\mathcal{L}_O$ is injective.

Based on our definition of observability, if the system is (completely) observable on the time interval $[t_0, t]$, then the map $\mathcal{L}_O$ is injective and $N(\mathcal{L}_O) = \{\mathbf{0}_n\}$. In the case that the observability map is injective, the initial state $\boldsymbol{x_0}$ can always be uniquely determined by computing the solution to the linear equation:

$$\boldsymbol{x_0} = (\mathcal{L}_O^* \mathcal{L}_O)^{-1} \mathcal{L}_O^* \Big(y(t) - f(u,t) \Big).$$

If the system is not (completely) observable on $[t_0, t]$, then the map $\mathcal{L}_O$ is not injective and the initial state cannot be uniquely determined. We can define the best approximate of the initial state as the following optimal solution:

$$\hat{\boldsymbol{x}}_{\mathbf{0}} = (\mathcal{L}_O^* \mathcal{L}_O)^{\dagger} \mathcal{L}_O^* \Big(y(t) - f(u,t) \Big).$$

Now suppose that the output of our system is described by

$$y(t) = \mathcal{L}_O(\boldsymbol{x_0}) + f(u,t) + z(t),$$

where $z(t)$ is some unknown error or measurement noise. In this case, we may find that $y(t) - f(u,t)$ is not in the range space of $\mathcal{L}_O$. In this case, let $\hat{y}(t) = \mathcal{L}_O(\hat{\boldsymbol{x}}_{\mathbf{0}}) + f(u,t)$. We aim to find the solution $\hat{\boldsymbol{x}}_{\mathbf{0}}$ such that $\hat{y}(t) - f(u,t)$ is as close to $y(t) - f(u,t)$ as possible. Assuming the system is (completely) observable, the least squares solution is given by

$$\hat{\boldsymbol{x}}_{\mathbf{0}} = (\mathcal{L}_O^* \mathcal{L}_O)^{-1} \mathcal{L}_O^* \Big(y(t) - f(u,t) \Big).$$

If the system is not (completely) observable, then the map $\mathcal{L}_O$ is not injective and the initial condition cannot be uniquely determined. The "best" solution is

$$\hat{\boldsymbol{x}}_{\mathbf{0}} = (\mathcal{L}_O^* \mathcal{L}_O)^{\dagger} \mathcal{L}_O^* \Big(y(t) - f(u,t) \Big).$$

6.3 Time Intervals

We have defined the controllability and observability of a system for an arbitrary time interval $[t_0, t_1]$. If we know the system is *controllable* on this time interval, we can also consider whether it is *controllable* on other time intervals. Similarly, if we know the system is *observable* on this time interval, we can also consider whether it is *observable* on other time intervals.

Proposition: If an LTV system is controllable on the time interval $[t_0, t_1]$, then it is controllable on any interval $[t_0', t_1']$, such that $t_0' \leq t_0 < t_1 \leq t_1'$.

Proof: To show that this is true, we can start by expressing the controllability grammian over the interval $[t_0', t_1']$ as

$$W_C[t_0', t_1'] = \int_{t_0'}^{t_1'} \phi(t_1', \tau) B(\tau) B^*(\tau) \phi^*(t_1', \tau) d\tau.$$

We can break this integral into three integrals over three distinct time intervals:

$$\begin{aligned} W_C[t_0', t_1'] = &\int_{t_0'}^{t_0} \phi(t_1', \tau) B(\tau) B^*(\tau) \phi^*(t_1', \tau) d\tau \\ &+ \int_{t_0}^{t_1} \phi(t_1', \tau) B(\tau) B^*(\tau) \phi^*(t_1', \tau) d\tau \\ &+ \int_{t_1}^{t_1'} \phi(t_1', \tau) B(\tau) B^*(\tau) \phi^*(t_1', \tau) d\tau. \end{aligned}$$

Recall from Section 3.1.3 that the state transition matrix satisfies the following: $\phi(t_2, t_0) = \phi(t_2, t_1)\phi(t_1, t_0)$ for all $t_0, t_1, t_2 \in \mathbb{R}_+$ such that $t_0 \leq t_1 \leq t_2$. This property allows us to express the controllability grammian as

$$\begin{aligned} W_C[t_0', t_1'] = &\int_{t_0'}^{t_0} \phi(t_1', t_0)\phi(t_0, \tau) B(\tau) B^*(\tau) \phi^*(t_0, \tau)\phi^*(t_1', t_0) d\tau \\ &+ \int_{t_0}^{t_1} \phi(t_1', t_1)\phi(t_1, \tau) B(\tau) B^*(\tau) \phi^*(t_1, \tau)\phi^*(t_1', t_1) d\tau \\ &+ \int_{t_1}^{t_1'} \phi(t_1', \tau) B(\tau) B^*(\tau) \phi^*(t_1', \tau) d\tau. \end{aligned}$$

Pulling out the state transition matrices that do not depend on the variable of integration (τ) from the integrals, we express the controllability grammian as

$$\begin{aligned} W_C[t_0', t_1'] = &\phi(t_1', t_0)\left(\int_{t_0'}^{t_0} \phi(t_0, \tau) B(\tau) B^*(\tau) \phi^*(t_0, \tau) d\tau\right)\phi^*(t_1', t_0) \\ &+ \phi(t_1', t_1)\left(\int_{t_0}^{t_1} \phi(t_1, \tau) B(\tau) B^*(\tau) \phi^*(t_1, \tau) d\tau\right)\phi^*(t_1', t_1) \\ &+ \int_{t_1}^{t_1'} \phi(t_1', \tau) B(\tau) B^*(\tau) \phi^*(t_1', \tau) d\tau. \end{aligned}$$

From the definition of the controllability grammian, we can express this as

$$W_C[t_0', t_1'] = \phi(t_1', t_0) W_C[t_0', t_0] \phi^*(t_1', t_0) + \phi(t_1', t_1) W_C[t_0, t_1] \phi^*(t_1', t_1) + W_C[t_1, t_1'].$$

We know that, in general, the controllability grammian is positive semidefinite, so $W_C[t_0', t_0]$, $W_C[t_0, t_1]$, and $W_C[t_1, t_1']$ are all PSD. Furthermore, because the system is controllable on $[t_0, t_1]$, we know that $W_C[t_0, t_1]$ is positive definite.

Another important property of the state transition matrix from Section 3.1.3 is that it is always invertible. This means that the first term of our expression for $W_C[t_0', t_1']$ is a congruence transformation of a PSD matrix, which is a PSD matrix. Similarly, the second term of our expression for $W_C[t_0', t_1']$ is a congruence transformation of a PD matrix, which is a PD matrix. The third term is simply a PSD matrix. The sum of a PD matrix and two PSD matrices is a PD matrix. This then implies that the controllability grammian, $W_C[t_0', t_1']$, is PD.

As we said previously, this implies that the system is (completely) controllable over the time interval $[t_0', t_1']$. Therefore, we have shown that if an LTV system is (completely) controllable on the time interval $[t_0, t_1]$, then it is completely controllable on any interval $[t_0', t_1']$ such that $t_0' \le t_0 < t_1 \le t_1'$.

Note that the same system that is controllable on $[t_0, t_1]$ is not necessarily controllable on any time interval $[t_0', t_1']$ if $[t_0, t_1]$ is not a subset of $[t_0', t_1']$.

Proposition: If an LTV system is observable on the time interval $[t_0, t_1]$, then it is observable on any interval $[t_0', t_1']$, such that $t_0' \le t_0 < t_1 \le t_1'$. The proof of this fact is very similar to the one given for the controllability case.

Chapter 7

Controllability/Observability: Continuous LTI Systems

7.1 Controllability of LTI Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. We will discuss methods for determining whether this system is (completely) controllable on an arbitrary time interval.

7.1.1 Controllability Grammian

Previously we said that a continuous linear system is controllable if and only if the controllability grammian is positive definite (or full rank). Recall that our definition of the controllability grammian for a continuous linear system is

$$W_C[t_0, t_1] = \mathcal{L}_C\mathcal{L}_C^* = \int_{t_0}^{t_1} \phi(t_1,\tau)B(\tau)B^*(\tau)\phi^*(t_1,\tau)d\tau.$$

For LTI systems, $B(\tau)$ is simply a matrix that is independent of time, and the state transition function is a matrix defined such that

$$\phi(t, t_0) = e^{\boldsymbol{A}(t-t_0)}.$$

We further assume that our system is in real space, so the controllability grammian for continuous LTI systems can then be expressed as

$$W_C[t_0, t_1] = \int_{t_0}^{t_1} e^{\boldsymbol{A}(t_1-\tau)}\boldsymbol{B}\boldsymbol{B}^T e^{\boldsymbol{A}^T(t_1-\tau)}d\tau.$$

While we can determine the controllability of our system using this matrix in same way we would for a time-varying system, we have two tests for determining the controllability of continuous LTI systems that do not require us to compute the controllability map or controllability grammian. I refer to these two tests as the *controllability matrix rank test* and the *PBH test for controllability*.

7.1.2 Controllability Matrix Rank Test

Let's start by defining the **controllability matrix** $\boldsymbol{Q} \in \mathbb{R}^{n \times nn_i}$ such that

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix}.$$

Theorem: The controllability matrix rank test says that the given LTI system is (completely) controllable on the interval $[t_0, t_1]$ if and only if $\text{rank}(\boldsymbol{Q}) = n$.

Proof (If): We will first show that if the controllability matrix has a rank of n, then the system is completely controllable. Let's suppose that the rank of the controllability matrix is equal to n but that the system is not controllable. If the system is not completely controllable on $[t_0, t_1]$, $W_C[t_0, t_1]$ is not positive definite, which implies that there exists a non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$ such that

$$\boldsymbol{v}^T W_C[t_0, t_1]\boldsymbol{v} = \boldsymbol{v}^T \left(\int_{t_0}^{t_1} e^{\boldsymbol{A}(t_1-\tau)} \boldsymbol{B}\boldsymbol{B}^T e^{\boldsymbol{A}^T(t_1-\tau)} d\tau \right) \boldsymbol{v} = 0.$$

We can equivalently write this equality as

$$\int_{t_0}^{t_1} \|\boldsymbol{v}^T e^{\boldsymbol{A}(t_1-\tau)} \boldsymbol{B}\|_2^2 d\tau = 0.$$

Because the vector norm is always non-negative and only equal to zero for the zero vector, this equality implies that $\boldsymbol{v}^T e^{\boldsymbol{A}(t_1-\tau)} B = \boldsymbol{0}_{n_i}$ for times $\tau \in [t_0, t_1]$. Evaluating the left-hand side of this expression at $\tau = t_1$, we see

$$\boldsymbol{v}^T e^{\boldsymbol{A}(t_1-t_1)} \boldsymbol{B} = \boldsymbol{v}^T \boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

If we differentiate the left-hand side of the previous expression with respect to τ, negate it, and evaluate it at $\tau = t_1$, then we find that

$$\boldsymbol{v}^T \boldsymbol{A} e^{\boldsymbol{A}(t_1-t_1)} \boldsymbol{B} = \boldsymbol{v}^T \boldsymbol{AB} = \boldsymbol{0}_{n_i}.$$

Continuing to differentiate and evaluate the expression at $\tau = t_1$, we find that

$$\boldsymbol{v}^T \boldsymbol{B} = \boldsymbol{v}^T \boldsymbol{AB} = \cdots = \boldsymbol{v}^T \boldsymbol{A}^{n-1} \boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

We can rewrite all of these equalities as a single matrix-vector equation:

$$\boldsymbol{v}^T \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix} = \boldsymbol{0}_{nn_i}.$$

Using our definition of the controllability matrix, we can write this as

$$\boldsymbol{v}^T \boldsymbol{Q} = \boldsymbol{0}_{nn_i}.$$

This means that the non-zero vector $\boldsymbol{v}^T$ is in the left null space of $\boldsymbol{Q}$, which implies that the non-zero vector $\boldsymbol{v}$ is in the null space of $\boldsymbol{Q}^T$. However, if the rank of $\boldsymbol{Q}$ is n, then its range space is $\mathbb{R}^n$. By the fundamental theorem of linear algebra, the range space of $\boldsymbol{Q}$ and null space of $\boldsymbol{Q}^T$ are orthogonal spaces, so the null space of $\boldsymbol{Q}^T$ must be $\{\boldsymbol{0}_n\}$. Our assumption that the system is not controllable then contradicts the fact that $\text{rank}(\boldsymbol{Q}) = n$. Therefore, if $\text{rank}(\boldsymbol{Q}) = n$, then the given LTI system is controllable on the interval $[t_0, t_1]$.

Proof (Only If): Now we will show the reverse direction: if the system is controllable on the interval $[t_0, t_1]$, then the rank of the controllability matrix is n. Suppose that the system is controllable on $[t_0, t_1]$ but $\text{rank}(\boldsymbol{Q}) < n$.

If the rank of $\boldsymbol{Q}$ is less than n, then by the fundamental theorem of linear algebra, the nullity of $\boldsymbol{Q}^T$ is greater than 0. This implies there is a non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$ in the null space of $\boldsymbol{Q}^T$, which means there is a non-zero vector $\boldsymbol{v}^T \in \mathbb{R}^n$ in the left null space of $\boldsymbol{Q}$. We can express this conclusion as

$$\boldsymbol{v}^T \boldsymbol{Q} = \boldsymbol{0}_{nn_i}$$

Plugging in our definition for the controllability matrix, we have

$$\boldsymbol{v}^T \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix} = \boldsymbol{0}_{nn_i}.$$

From this equation, we obtain the following set of n equalities:

$$\boldsymbol{v}^T \boldsymbol{B} = \boldsymbol{v}^T \boldsymbol{AB} = \cdots = \boldsymbol{v}^T \boldsymbol{A}^{n-1} \boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

The matrix exponential is an analytic function (see my linear algebra notes for more information) that can be expressed as a polynomial of the form

$$e^{\boldsymbol{A}t} = a_1 \boldsymbol{A}^{n-1} + \cdots a_{n-1}\boldsymbol{A} + a_n \boldsymbol{I_n},\ a_1, \cdots, a_n \in \mathbb{R}.$$

From this expression, we can see that $\boldsymbol{v}^T e^{\boldsymbol{A}t}\boldsymbol{B}$ must be equal to zero. For this particular non-zero vector $\boldsymbol{v}$, we can then see that

$$\boldsymbol{v}^T W_C[t_0, t_1]\boldsymbol{v} = \int_{t_0}^{t_1} \|\boldsymbol{v}^T e^{\boldsymbol{A}(t_1-\tau)}\boldsymbol{B}\|_2^2 d\tau = 0.$$

Because there exists some non-zero vector $\boldsymbol{v}$ such that $\boldsymbol{v}^T W_C[t_0, t_1]\boldsymbol{v}$ is not strictly positive, $W_C[t_0, t_1]$ cannot be positive definite, which implies that the system is not actually controllable. Therefore, if the system is completely controllable, then the rank of the controllability matrix must be equal to n.

7.1.3 PBH Test for Controllability

Another useful test is the PBH (Popov–Belevitch–Hautus) controllability test.

Theorem: The PBH test for controllability says that an LTI system is (completely) controllable on the time interval $[t_0, t_1]$ if and only if

$$\text{rank}\begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} = n \ \forall s \in \mathbb{C}.$$

This matrix can only lose rank for values of s on the spectrum of $\boldsymbol{A}$, so we only need to check values of s that are eigenvalues of $\boldsymbol{A}$. Therefore, an LTI system is (completely) controllable on any time interval if and only if

$$\operatorname{rank}\begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} = n \ \forall s \in \lambda(\boldsymbol{A}).$$

Proof: To show the validity of this controllability test, recall that we just showed that the LTI system is (completely) controllable on the interval $[t_0, t_1]$ if and only if $\operatorname{rank}(\boldsymbol{Q}) = n$. Therefore, to show that the LTI system is (completely) controllable on the interval $[t_0, t_1]$ if and only if the PBH test holds, we can show that the PBH test holds if and only if $\operatorname{rank}(\boldsymbol{Q}) = n$.

Proof (If): Let's first show that the PBH test holds if the controllability matrix has rank n by assuming that $\operatorname{rank}(\boldsymbol{Q}) = n$ but the PBH test does not hold. If the PBH test does not hold, then there exists an eigenvalue $\lambda_i \in \lambda(\boldsymbol{A})$ for which the rank of the PBH matrix is less than n. By the fundamental theorem of linear algebra, this means that for $s = \lambda_i$, there is a non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$ in the null space of the adjoint of PBH matrix and a non-zero vector $\boldsymbol{v}^T \in \mathbb{R}^n$ in the left null space of the the PBH matrix. We can express this as

$$\boldsymbol{v}^T \begin{bmatrix} \lambda_i \boldsymbol{I_n} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} = \boldsymbol{0}_{n+n_i}.$$

We can write this single equation as the following two equalities:

$$\boldsymbol{v}^T (\lambda_i \boldsymbol{I_n} - \boldsymbol{A}) = \boldsymbol{0}_n \quad \text{and} \quad \boldsymbol{v}^T \boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

From the first constraint, we can see that

$$\boldsymbol{v}^T \boldsymbol{A} = \lambda_i \boldsymbol{v}^T.$$

Right multiplying both sides of the equation above by $\boldsymbol{B}$, we can see that

$$\boldsymbol{v}^T \boldsymbol{A}\boldsymbol{B} = \lambda_i \boldsymbol{v}^T \boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

Similarly, right multiplying both sides of the equation by $\boldsymbol{A}\boldsymbol{B}$, we can see that

$$\boldsymbol{v}^T \boldsymbol{A}^2 \boldsymbol{B} = \lambda_i \boldsymbol{v}^T \boldsymbol{A}\boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

If we continue with this process, we will find that

$$\boldsymbol{v}^T \boldsymbol{B} = \boldsymbol{v}^T \boldsymbol{A}\boldsymbol{B} = \cdots = \boldsymbol{v}^T \boldsymbol{A}^{n-1} \boldsymbol{B} = \boldsymbol{0}_{n_i}.$$

We can combine all of these equalities into the single matrix-vector equation:

$$\boldsymbol{v}^T \begin{bmatrix} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix} = \boldsymbol{0}_{nn_i}.$$

Now plugging in our definition for the controllability matrix we have

$$\boldsymbol{v}^T \boldsymbol{Q} = \boldsymbol{0}_{nn_i}.$$

This implies that there is a non-zero vector $\boldsymbol{v}^T$ in the left null space of $\boldsymbol{Q}$, which means $\text{rank}(\boldsymbol{Q}) < n$. We have found a contradiction to our assumption that $\text{rank}(\boldsymbol{Q}) = n$. Therefore, if $\text{rank}(\boldsymbol{Q}) = n$, then the PBH test holds.

Proof (Only If): Now we want to show that if the PBH test holds, then the rank of the controllability matrix must be equal to n. To show this, let's assume the PBH test holds but the rank of the controllability matrix is less than n.

If $\text{rank}(\boldsymbol{Q}) < n$, then $R(\boldsymbol{Q}) = \text{span}(\boldsymbol{B}, \boldsymbol{AB}, \cdots, \boldsymbol{A}^{n-1}\boldsymbol{B})$ is a strict subset of $\mathbb{R}^n$. This then implies that there is a non-empty subspace $V \subset \mathbb{R}^n$ such that $R(\boldsymbol{Q}) \oplus V = \mathbb{R}^n$. If $\boldsymbol{y} \in R(\boldsymbol{Q})$, then it can be expressed as a linear combination of the columns of $\boldsymbol{B}, \boldsymbol{AB}, \cdots, \boldsymbol{A}^{n-1}\boldsymbol{B}$. This implies that $\boldsymbol{Ay}$ can be expressed as a linear combination of the columns of $\boldsymbol{AB}, \boldsymbol{A}^2\boldsymbol{B}, \cdots, \boldsymbol{A}^n\boldsymbol{B}$. By the Cayley Hamilton theorem (discussed in my linear algebra notes), this is equivalent to saying that $\boldsymbol{Ay}$ can be expressed as a linear combination of the columns of $\boldsymbol{B}, \boldsymbol{AB}, \cdots, \boldsymbol{A}^{n-1}\boldsymbol{B}$, which means $\boldsymbol{Ay} \in R(\boldsymbol{Q})$. Because $\boldsymbol{y} \in R(\boldsymbol{Q})$ implies that $\boldsymbol{Ay} \in R(\boldsymbol{Q})$, $R(\boldsymbol{Q})$ is an A-invariant subspace. By the second representation theorem (see linear algebra notes), there exists a representation of $\boldsymbol{A}$ of the form

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \tilde{\boldsymbol{A}}_{11} & \tilde{\boldsymbol{A}}_{12} \\ 0 & \tilde{\boldsymbol{A}}_{22} \end{bmatrix}.$$

If $\boldsymbol{y} \in R(\boldsymbol{Q})$, then it can be expressed as a linear combination of the columns of $\boldsymbol{B}, \boldsymbol{AB}, \cdots, \boldsymbol{A}^{n-1}\boldsymbol{B}$. This implies $\boldsymbol{y} \in R(\boldsymbol{B})$, which means $R(\boldsymbol{B}) \subset R(\boldsymbol{Q})$. Because $R(\boldsymbol{B}) \subset R(\boldsymbol{Q})$, there also exists a representation of $\boldsymbol{B}$ of the form

$$\tilde{\boldsymbol{B}} = \begin{bmatrix} \tilde{\boldsymbol{B}}_1 \\ 0 \end{bmatrix}.$$

Because we can represent $\boldsymbol{A}$ and $\boldsymbol{B}$ in this way, there exists an invertible matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n}$ such that $\tilde{\boldsymbol{A}} = \boldsymbol{TAT}^{-1}$ and $\tilde{\boldsymbol{B}} = \boldsymbol{TB}$. This allows us to write

$$\begin{bmatrix} s\boldsymbol{I_n} - \tilde{\boldsymbol{A}} & \tilde{\boldsymbol{B}} \end{bmatrix} = \begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{TAT}^{-1} & \boldsymbol{TB} \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{T}^{-1} & 0 \\ 0 & \boldsymbol{I_{n_i}} \end{bmatrix}.$$

Now we can see that the leftmost matrix the one used in the PBH controllability test are *similar matrices*, which implies that they have the same rank. Now let's look more closely at this matrix to determine its rank:

$$\begin{bmatrix} s\boldsymbol{I_n} - \tilde{\boldsymbol{A}} & \tilde{\boldsymbol{B}} \end{bmatrix} = \begin{bmatrix} s\boldsymbol{I} - \tilde{\boldsymbol{A}}_{11} & -\tilde{\boldsymbol{A}}_{12} & \tilde{\boldsymbol{B}}_1 \\ 0 & s\boldsymbol{I} - \tilde{\boldsymbol{A}}_{22} & 0 \end{bmatrix}$$

This matrix clearly loses rank for $s \in \lambda(\tilde{\boldsymbol{A}}_{22})$, which implies that its rank is less than n. This means that under our assumption that $\text{rank}(\boldsymbol{Q}) < n$, the PBH test does not hold. Therefore, if the PBH test holds, then $\text{rank}(\boldsymbol{Q}) = n$.

7.1.4 Stabilizable Systems

Consider the LTI system given previously. This system is said to be **stabilizable** if all of its uncontrollable modes (refer to Section 5.4.2) are stable. This is equivalent to saying that the system is **stabilizable** if all of its unstable modes are controllable. To check if the given LTI system is stabilizable, we have a modified PBH test, which says that the system is stabilizable if and only if

$$\text{rank}\begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} = n \;\; \forall s \in \lambda(\boldsymbol{A}) \cap \mathbb{C}_+.$$

7.1.5 Controllable Subspaces

Recall that the **controllable subspace** is the region in the state space that can be reached from some initial condition. If the system is controllable, then the controllable subspace is the entire state space (usually $\mathbb{R}^n$). In Section 6.1.3, we showed that the controllable subspace for a continuous LTV system is

$$\phi(t_1, t_0)\boldsymbol{x_0} + R(\mathcal{L}_C).$$

Furthermore, a state $\boldsymbol{x_1}$ is reachable at time t_1 from initial state $\boldsymbol{x_0}$ if and only

$$\left(\boldsymbol{x_1} - \phi(t_1, t_0)\boldsymbol{x_0}\right) \in R(\mathcal{L}_C).$$

For a continuous LTI system with initial state $\boldsymbol{x_0}$ at time t_0, the state transition function is given by $\phi(t_1, t_0) = e^{\boldsymbol{A}(t_1 - t_0)}$, so the controllable subspace is

$$e^{\boldsymbol{A}(t_1 - t_0)}\boldsymbol{x_0} + R(\boldsymbol{\mathcal{L}_C}).$$

Furthermore, a state $\boldsymbol{x_1}$ is reachable at time t_1 if and only

$$\left(\boldsymbol{x_1} - e^{\boldsymbol{A}(t_1 - t_0)}\boldsymbol{x_0}\right) \in R(\mathcal{L}_C).$$

The range of the controllability map is the same as the range of the controllability matrix, $\boldsymbol{Q}$, given in Section 7.1.2. Therefore, for LTI systems the reachable/controllable subspace can equivalently be expressed as

$$e^{\boldsymbol{A}(t_1 - t_0)}\boldsymbol{x_0} + R(\boldsymbol{Q}).$$

Furthermore, there exists an input $u \in \mathcal{U}$ that steers the system from the initial state $\boldsymbol{x_0}$ at time t_0 to the state $\boldsymbol{x_1}$ at time t_1 if and only if

$$\left(\boldsymbol{x_1} - e^{\boldsymbol{A}(t_1 - t_0)}\boldsymbol{x_0}\right) \in R(\boldsymbol{Q}).$$

Note that if the system is controllable, then the range of the controllability matrix is the entire state space and every vector $\boldsymbol{x_1}$ is reachable.

7.1.6 Computing the Optimal Control

Suppose we want to find the best input sequence that allows us to move the state of our system from the state $\boldsymbol{x_0}$ at time t_0 to the state $\boldsymbol{x_1}$ at time t_1. As shown in the previous section, if $\boldsymbol{x_1}$ is in the controllable subspace, there exists a solution $u \in \mathcal{U}$ to the linear equation $\boldsymbol{x_1} = e^{\boldsymbol{A}(t_1 - t_0)} + \mathcal{L}_C(u)$. To find the optimal solution, we can follow the same approach discussed in Section 6.1.4.

7.2 Observability of LTI Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o \times n_i}$. We will discuss methods for determining whether this system is (completely) observable on an arbitrary time interval.

7.2.1 Observability Grammian

Previously we said that a continuous linear system is observable if and only if the observability grammian is positive definite (or full rank). Recall that our definition of the observability grammian for a continuous linear system is

$$W_O[t_0, t_1] = \mathcal{L}_O^* \mathcal{L}_O = \int_{t_0}^{t_1} \phi^*(\tau, t_0) C^*(\tau) C(\tau) \phi(\tau, t_0) d\tau.$$

For LTI systems, $C(\tau)$ is simply a matrix that is independent of time, and the state transition function is a matrix defined such that

$$\phi(t, t_0) = e^{\boldsymbol{A}(t - t_0)}.$$

We further assume that our system is in real space, so the observability grammian for continuous LTI systems can then be expressed as

$$W_O[t_0, t_1] = \int_{t_0}^{t_1} e^{\boldsymbol{A}^T(t_1 - \tau)} \boldsymbol{C}^T \boldsymbol{C} e^{\boldsymbol{A}(t_1 - \tau)} d\tau.$$

While we can determine the observability of our system using this matrix in same way we would for a time-varying system, we have two tests for determining the observability of continuous LTI systems that do not require us to compute the observability map or observability grammian. I refer to these two tests as the *observability matrix rank test* and the *PBH test for observability*.

7.2.2 Observability Matrix Rank Test

Let's start by defining the observability matrix $\boldsymbol{O} \in \mathbb{R}^{nn_o \times n}$ such that

$$\boldsymbol{O} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{n-1} \end{bmatrix}.$$

Theorem: The observability matrix rank test says that the given LTI system is (completely) observable on the interval $[t_0, t_1]$ if and only if $\text{rank}(\boldsymbol{O}) = n$.

Proof: The proof that this statement holds is very similar to the one given for the controllability matrix rank test (Section 7.1.2). We can follow the same steps in this proof using the matrices $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ instead of $(\boldsymbol{A}, \boldsymbol{B})$.

7.2.3 PBH Test for Observability

Theorem: The PBH test for observability says that an LTI system is (completely) observable on the time interval $[t_0, t_1]$ if and only if

$$\text{rank} \begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} = n \;\; \forall s \in \mathbb{C}.$$

This matrix can only lose rank for values of s on the spectrum of $\boldsymbol{A}$, so we only need to check values of s that are eigenvalues of $\boldsymbol{A}$. Therefore, an LTI system is (completely) observable on any time interval if and only if

$$\text{rank} \begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} \\ \boldsymbol{C} \end{bmatrix} = n \;\; \forall s \in \lambda(\boldsymbol{A}).$$

Proof: To prove the validity of the PBH test for observability, we can follow the same steps in the proof given for the PBH test for controllability (Section 7.1.3) using the matrices $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ instead of $(\boldsymbol{A}, \boldsymbol{B})$.

7.2.4 Detectable Systems

Consider the LTI system given previously. This system is said to be **detectable** if all of its unobservable modes (refer to Section 5.4.3) are stable. This is equivalent to saying that the system is **detectable** if all of its unstable modes are observable. To check if the given LTI system is detectable, we have a modified PBH test, which says that the system is detectable if and only if

$$\text{rank} \begin{bmatrix} s\boldsymbol{I_n} - \boldsymbol{A} \\ C \end{bmatrix} = n \;\; \forall s \in \lambda(\boldsymbol{A}) \cap \mathbb{C}_+$$

7.2.5 Observable Subspaces

Recall that the **observable subspace** is the region in the state space for which initial conditions in this space can be uniquely determined by observing the inputs and outputs of the system over time. If the system is observable, then the observable subspace is the entire state space (usually $\mathbb{R}^n$). In Section 6.2.3, we showed that the observable subspace for a continuous LTV system is $X \setminus N(\mathcal{L}_0)$.

For an LTI system, $N(\mathcal{L}_O) = N(\boldsymbol{O})$, where $\boldsymbol{O}$ is the observability matrix defined in Section 7.2.2. Therefore, for LTI systems, we call the null space $N(\boldsymbol{O})$ the unobservable subspace and $X \setminus N(\boldsymbol{O})$ the observable subspace.

7.2.6 Determining the Initial Condition

Suppose we want to determine the initial state of our system after observing a series of outputs resulting from known inputs. To do so, we can follow the same approach discussed for LTV systems in Section 6.2.4.

7.3 Canonical Forms

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{b}u(t) \\ y(t) = \boldsymbol{c}^T x(t) + du(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, and $y(t) \in \mathbb{R}$. Assume $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{c} \in \mathbb{R}^n$, and $d \in \mathbb{R}$. Consider the state transformation $\tilde{x}(t) = \boldsymbol{T}x(t)$, where $\boldsymbol{T}$ is an invertible matrix. We can express an equivalent system (recall Section 1.5.2) as

$$\begin{cases} \dot{\tilde{x}}(t) = \tilde{\boldsymbol{A}}\tilde{x}(t) + \tilde{\boldsymbol{b}}u(t) \\ y(t) = \tilde{\boldsymbol{c}}^T \tilde{x}(t) + \tilde{d}u(t) \\ \tilde{x}(t_0) = \boldsymbol{T}\boldsymbol{x_0} \end{cases},$$

where $\tilde{\boldsymbol{A}} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}$, $\tilde{\boldsymbol{b}} = \boldsymbol{T}\boldsymbol{b}$, $\tilde{\boldsymbol{c}}^T = \boldsymbol{c}^T\boldsymbol{T}^{-1}$, and $\tilde{d} = d$. We can select the transformation matrix, $\boldsymbol{T}$, to place the system into either controllable or observable canonical form (assuming the system is controllable and observable).

Recall from Section 4.3 that the transfer function of this system is defined as

$$H(s) = \tilde{\boldsymbol{c}}^T(s\boldsymbol{I_n} - \tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{b}} + \tilde{d} = \boldsymbol{c}^T(s\boldsymbol{I_n} - \boldsymbol{A})^{-1}\boldsymbol{b} + d.$$

Suppose that the transfer function has the form

$$H(s) = \frac{b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n}.$$

for some real coefficients $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$.

7.3.1 Controllable Canonical Form

Suppose there exists an invertible transformation matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ such that

$$\tilde{\boldsymbol{A}} = \boldsymbol{TAT}^{-1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}, \quad \tilde{\boldsymbol{b}} = \boldsymbol{Tb} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

$$\tilde{\boldsymbol{c}}^T = \boldsymbol{c}^T\boldsymbol{T}^{-1} = \begin{bmatrix} b_n & b_{n-1} & b_{n-2} & \cdots b_1 \end{bmatrix}, \qquad \tilde{d} = d.$$

We say that the system defined by these matrices is in **controllable canonical form**. To find the transformation matrix, let $\boldsymbol{T}$ be defined such that

$$\boldsymbol{T}^{-1} := \begin{bmatrix} \boldsymbol{t_1} & \boldsymbol{t_2} & \cdots & \boldsymbol{t_n} \end{bmatrix}.$$

Now notice that we should choose the columns of $\boldsymbol{T}^{-1}$ in the following way:

$$\boldsymbol{t_i} = \begin{cases} \boldsymbol{b} & \text{for } i = n \\ \boldsymbol{At_{i+1}} + a_{n-i}\boldsymbol{b} & \text{for } i = n-1, \cdots, 1 \end{cases}.$$

Theorem: A continuous LTI single-input single-output (SISO) system can be expressed in controllable canonical form if and only if it is controllable.

Proof: We already showed how to find the inverse of the transformation matrix. Now we need to show that this matrix is invertible if and only if the system is controllable. To do so, notice that we can express the columns of $\boldsymbol{T}^{-1}$ as

$$\boldsymbol{t_n} = \boldsymbol{b}$$

$$\boldsymbol{t_{n-1}} = \boldsymbol{Ab} + a_1\boldsymbol{b}$$
$$\boldsymbol{t_{n-2}} = \boldsymbol{A}^2\boldsymbol{b} + a_1\boldsymbol{Ab} + a_2\boldsymbol{b}$$
$$\vdots$$
$$\boldsymbol{t_1} = \boldsymbol{A}^{n-1}\boldsymbol{b} + a_1\boldsymbol{A}^{n-2} + \cdots + a_{n-2}\boldsymbol{Ab} + a_{n-1}\boldsymbol{b}$$

Now notice we can express the inverse of the desired transformation matrix as

$$\boldsymbol{T}^{-1} = \begin{bmatrix} \boldsymbol{b} & \boldsymbol{AB} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{b} \end{bmatrix} \begin{bmatrix} a_{n-1} & a_{n-2} & \cdots & a_1 & 1 \\ a_{n-2} & \vdots & \iddots & 1 & 0 \\ \vdots & a_1 & \iddots & 0 & \vdots \\ a_1 & 1 & \iddots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

The matrix $\boldsymbol{T}^{-1}$ is invertible if and only if its determinant is not equal to zero, which is true if and only if the determinant of the two matrix factors are non-zero (i.e., both matrices have full rank). The first matrix is the controllability matrix defined in Section 7.1.2, which has full rank if any only if the system is controllable. The second matrix has n linearly independent rows/columns regardless of the coefficients $a_1, a_2, \cdots, a_n$, so it necessarily has full rank. Therefore, $\boldsymbol{T}^{-1}$ is invertible if and only if the system is controllable.

7.3.2 Observable Canonical Form

Suppose there exists an invertible transformation matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ such that

$$\tilde{\boldsymbol{A}} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & \cdots & 0 & -a_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_1 \end{bmatrix}, \qquad \tilde{\boldsymbol{b}} = \boldsymbol{T}\boldsymbol{b} = \begin{bmatrix} b_n \\ b_{n-1} \\ b_{n-2} \\ \vdots \\ b_1 \end{bmatrix},$$

$$\tilde{\boldsymbol{c}}^T = \boldsymbol{c}^T\boldsymbol{T}^{-1} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, \qquad \tilde{d} = d.$$

We say that the system defined by these matrices is in **observable canonical form**. To find the transformation matrix, let $\boldsymbol{T}$ be defined such that

$$\boldsymbol{T} := \begin{bmatrix} \boldsymbol{t_1}^T \\ \boldsymbol{t_2}^T \\ \vdots \\ \boldsymbol{t_n}^T \end{bmatrix}.$$

Now notice that we should choose the rows of $\boldsymbol{T}$ in the following way:

$$\boldsymbol{t_i}^T = \begin{cases} \boldsymbol{c}^T & \text{for } i = n \\ \boldsymbol{t_{i+1}}^T\boldsymbol{A} + a_{n-i}\boldsymbol{c}^T & \text{for } i = n-1, \cdots, 1 \end{cases}.$$

Theorem: A continuous LTI single-input single-output (SISO) system can be expressed in observable canonical form if and only if it is observable.

The proof of this theorem is very similar to the one shown in the previous section for the controllable canonical form, so I will not include it here.

7.4 Kalman Decomposition

A **Kalman decomposition** provides a mathematical way to convert a representation of any continuous LTI system to a form that indicates the controllable and observable components of the system. Once the system is expressed in this form, we can easily infer the system's reachable and observable subspaces.

Recall from Section 7.1.5 that for a continuous LTI system with initial state $\boldsymbol{x_0}$ at time t_0, the controllable subspace is $R(\boldsymbol{Q}) + e^{\boldsymbol{A}(t_1-t_0)}\boldsymbol{x_0}$, where $\boldsymbol{Q}$ is the controllability matrix defined in Section 7.1.2. If we assume that the initial state is the zero vector, then the controllable subspace is simply $R(\boldsymbol{Q})$. Let V_1 be the uncontrollable subspace defined such that $R(\boldsymbol{Q}) \oplus V_1 = \mathbb{R}^n$.

In Section 7.2.5, we said that the unobservable subspace is defined by $N(\boldsymbol{O})$, where $\boldsymbol{O}$ is the observability matrix defined in Section 7.2.2. Let V_2 be the observable subspace defined such that $N(\boldsymbol{O}) \oplus V_2 = \mathbb{R}^n$.

With these definitions, we can define the following four subspaces:

1. $\Sigma_{co} := R(\boldsymbol{Q}) \cap V_2$ – Set of states that are controllable and observable
2. $\Sigma_{c\bar{o}} := R(\boldsymbol{Q}) \cap N(\boldsymbol{O})$ – Set of states that are controllable and unobservable
3. $\Sigma_{\bar{c}o} := V_1 \cap V_2$ – Set of states that are uncontrollable and observable
4. $\Sigma_{\bar{c}\bar{o}} := V_1 \cap N(\boldsymbol{O})$ – Set of states that are uncontrollable and unobservable

Notice that we can represent the controllable and uncontrollable subspaces as

$$R(\boldsymbol{Q}) = \Sigma_{co} \oplus \Sigma_{c\bar{o}} \quad \text{and} \quad V_1 = \Sigma_{\bar{c}o} \oplus \Sigma_{\bar{c}\bar{o}}.$$

Similarly, we can represent the unobservable and observable subspaces as

$$N(\boldsymbol{O}) = \Sigma_{c\bar{o}} \oplus \Sigma_{\bar{c}\bar{o}} \quad \text{and} \quad V_2 = \Sigma_{co} \oplus \Sigma_{\bar{c}o}.$$

Recall that we chose V_1 and V_2 such that $R(\boldsymbol{Q}) \oplus V_1 = \mathbb{R}^n$ and $N(\boldsymbol{O}) \oplus V_2 = \mathbb{R}^n$. Therefore, the entire n-dimensional Euclidean space can also be expressed as

$$\mathbb{R}^n = \Sigma_{co} \oplus \Sigma_{c\bar{o}} \oplus \Sigma_{\bar{c}o} \oplus \Sigma_{\bar{c}\bar{o}}.$$

If we define T_{co} as the matrix representation of the subspace Σ_{co}, $T_{c\bar{o}}$ as the matrix representation of the subspace $\Sigma_{c\bar{o}}$, $T_{\bar{c}o}$ as the matrix representation of the subspace $\Sigma_{\bar{c}o}$, and $T_{\bar{c}\bar{o}}$ as the matrix representation of the subspace $\Sigma_{\bar{c}\bar{o}}$, then we can define the invertible matrix $\boldsymbol{T}^{-1} \in \mathbb{R}^{n\times n}$ such that

$$\boldsymbol{T}^{-1} = \begin{bmatrix} T_{co} & T_{c\bar{o}} & T_{\bar{c}o} & T_{\bar{c}\bar{o}} \end{bmatrix}.$$

Using this transformation matrix, we can define a new set of matrices:

$$\hat{\boldsymbol{A}} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1}, \quad \hat{\boldsymbol{B}} = \boldsymbol{T}\boldsymbol{B}, \quad \hat{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{T}^{-1}, \quad \hat{\boldsymbol{D}} = \boldsymbol{D}.$$

The transformed system defined by the matrices $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{B}}$, $\hat{\boldsymbol{C}}$, and $\hat{\boldsymbol{D}}$ has the form:

$$\hat{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A_{11}} & 0 & \boldsymbol{A_{13}} & 0 \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} & \boldsymbol{A_{23}} & \boldsymbol{A_{24}} \\ 0 & 0 & \boldsymbol{A_{33}} & 0 \\ 0 & 0 & \boldsymbol{A_{43}} & \boldsymbol{A_{44}} \end{bmatrix}, \qquad \hat{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{B_1} \\ \boldsymbol{B_2} \\ 0 \\ 0 \end{bmatrix}$$

$$\hat{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{C_1} & 0 & \boldsymbol{C_3} & 0 \end{bmatrix}, \qquad \hat{\boldsymbol{D}} = \boldsymbol{D}.$$

This is the Kalman decomposition of the given continuous LTI system.

Chapter 8

Controllability/Observability: Discrete LTV Systems

8.1 Controllability of Discrete LTV Systems

Consider the discrete linear time-varying (LTV) system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A_k x_k} + \boldsymbol{B_k u_k} \\ \boldsymbol{y_k} = \boldsymbol{C_k x_k} + \boldsymbol{D_k u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A_k} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B_k} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C_k} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D_k} \in \mathbb{R}^{n_o\times n_i}$. We will determine the conditions under which this system is (completely) controllable on a discrete time interval.

8.1.1 Controllability Condition

For discrete time systems, we are interested in whether the system is controllable on a finite time interval $[0, N]$. We want to know whether there exists a set of N inputs, $\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}$ such that system can reach any state $\boldsymbol{x_N}$ from an arbitrary initial state $\boldsymbol{x_0}$. At discrete time N, the state of the system is

$$\boldsymbol{x_N} = \prod_{i=0}^{N-1} \boldsymbol{A_i x_0} + \sum_{i=0}^{N-1} \prod_{j=i+1}^{N-1} \boldsymbol{A_j B_i u_i}.$$

We can rewrite the sum of products using the $n \times Nn_i$ matrix defined as

$$\boldsymbol{Q_N} := \begin{bmatrix} \boldsymbol{B_{N-1}} & \boldsymbol{A_{N-1}B_{N-2}} & \boldsymbol{A_{N-1}A_{N-2}B_{N-3}} & \dots & \boldsymbol{A_{N-1}}\cdots\boldsymbol{A_0 B_0} \end{bmatrix}.$$

With this definition, we can express the state of the system at time N as

$$\boldsymbol{x_N} = \prod_{i=0}^{N-1} \boldsymbol{A_i x_0} + \boldsymbol{Q_N} \begin{bmatrix} \boldsymbol{u_{N-1}} \\ \boldsymbol{u_{N-2}} \\ \vdots \\ \boldsymbol{u_0} \end{bmatrix}.$$

In order for this system to be controllable, the matrix $\boldsymbol{Q_N}$ must be able to map the sequence of inputs to any vector in the state space, $\mathbb{R}^n$. Therefore, the system is controllable if and only if the rank of this matrix is equal to n.

8.1.2 Controllable Subspaces

In the previous section, we showed that for a discrete LTV system with initial state $\boldsymbol{x_0}$ at discrete time 0, we can express the state of the system at time N as

$$\boldsymbol{x_N} = \prod_{i=0}^{N-1} \boldsymbol{A_i}\boldsymbol{x_0} + \boldsymbol{Q_N}\begin{bmatrix}\boldsymbol{u_{N-1}}\\ \boldsymbol{u_{N-2}}\\ \vdots \\ \boldsymbol{u_0}\end{bmatrix}, \text{ where}$$

$$\boldsymbol{Q_N} \coloneqq \begin{bmatrix}\boldsymbol{B_{N-1}} & \boldsymbol{A_{N-1}}\boldsymbol{B_{N-2}} & \boldsymbol{A_{N-1}}\boldsymbol{A_{N-2}}\boldsymbol{B_{N-3}} & \dots & \boldsymbol{A_{N-1}}\cdots\boldsymbol{A_0}\boldsymbol{B_0}\end{bmatrix}.$$

This tells us that the set of states reachable from the initial point $\boldsymbol{x_0}$ is

$$\prod_{i=0}^{N-1} \boldsymbol{A_i}\boldsymbol{x_0} + R(\boldsymbol{Q_N}).$$

Furthermore, a state $\boldsymbol{x_N}$ is reachable at discrete time N if and only if

$$\left(\boldsymbol{x_N} - \prod_{i=0}^{N-1} \boldsymbol{A_i}\boldsymbol{x_0}\right) \in R(\boldsymbol{Q_N}).$$

Recall that the system is controllable if and only if the range of $\boldsymbol{Q_N}$ is the entire state space, $\mathbb{R}^n$. If this is the case, every state $\boldsymbol{x_N}$ is reachable in N time steps.

8.1.3 Computing the Optimal Control

Suppose we want to find the best input sequence that allows us to move the state of our system from the state $\boldsymbol{x_0}$ at discrete time 0 to the state $\boldsymbol{x_N}$ at discrete time N. As shown in the previous section, if $\boldsymbol{x_1}$ is in the controllable subspace, then there exists a solution $u \in \mathcal{U}$ to the linear equation for $\boldsymbol{x_N}$.

We showed that if the system is controllable on the time interval $[0, N]$, then the matrix $\boldsymbol{Q_N}$ has full row rank and $R(\boldsymbol{Q_N}) = X = \mathbb{R}^n$. This means that if the system is controllable, then $\boldsymbol{x_N}$ is always in the controllable subspace, so there is always a solution to the linear equation. From linear matrix equations, the minimum norm solution is given by

$$u = \boldsymbol{Q_N^T}(\boldsymbol{Q_N}\boldsymbol{Q_N^T})^{-1}\left(\boldsymbol{x_N} - \prod_{i=0}^{N-1} \boldsymbol{A_i}\boldsymbol{x_0}\right).$$

If the system is not controllable on the time onterval $[O, N]$, then the map $\boldsymbol{Q_N}$ does not have full row rank and $R(\boldsymbol{Q_N}) \subset X = \mathbb{R}^n$. If $\boldsymbol{x_N}$ is still in the controllable subspace, then the minimum norm solution is now given by

$$u = \boldsymbol{Q_N^T}(\boldsymbol{Q_N}\boldsymbol{Q_N^T})^{\dagger}\left(\boldsymbol{x_N} - \prod_{i=0}^{N-1} \boldsymbol{A_i}\boldsymbol{x_0}\right).$$

If $\boldsymbol{x_N}$ is not in the controllable subspace, then there is no input $u \in \mathcal{U}$ that can steer the system from the initial state $\boldsymbol{x_0}$ to $\boldsymbol{x_N}$. Instead, we can choose a control that steers the system from $\boldsymbol{x_0}$ to a state $\boldsymbol{\hat{x}_N}$, which is the closest state to $\boldsymbol{x_N}$ within the controllable subspace. Now the minimum norm solution is

$$\hat{u} = \boldsymbol{Q_N^T}(\boldsymbol{Q_N}\boldsymbol{Q_N^T})^\dagger \left(\boldsymbol{x_N} - \prod_{i=0}^{N-1} \boldsymbol{A_i}\boldsymbol{x_0}\right).$$

8.2 Obseravability of Discrete LTV Systems

Consider the discrete linear time-varying (LTV) system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A_k}\boldsymbol{x_k} + \boldsymbol{B_k}\boldsymbol{u_k} \\ \boldsymbol{y_k} = \boldsymbol{C_k}\boldsymbol{x_k} + \boldsymbol{D_k}\boldsymbol{u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A_k} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B_k} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C_k} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D_k} \in \mathbb{R}^{n_o\times n_i}$. We will determine the conditions under which this system is (completely) observable on a discrete time interval.

8.2.1 Observability Condition

For discrete time systems, we are interested in whether the system is observable on a finite time interval $[0, N-1]$. We want to know whether the initial state of the system can be uniquely determined by observing its inputs $(\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}})$ and outputs $(\boldsymbol{y_0}, \boldsymbol{y_1}, \cdots, \boldsymbol{y_{N-1}})$ over time. Recall that we can express the output of the system at time step k as

$$\boldsymbol{y_k} = \boldsymbol{C_k}\prod_{i=0}^{k-1}\boldsymbol{A_i}\boldsymbol{x_0} + \sum_{i=0}^{k-1}\boldsymbol{C_k}\prod_{j=i+1}^{k-1}\boldsymbol{A_j}\boldsymbol{B_i}\boldsymbol{u_i} + \boldsymbol{D_k}\boldsymbol{u_k}.$$

We can then express all of the outputs on the time interval $[0, N-1]$ as

$$\begin{bmatrix} \boldsymbol{y_0} \\ \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{N-1}} \end{bmatrix} = \boldsymbol{O_N}\boldsymbol{x_0} + \boldsymbol{F_N}\begin{bmatrix} \boldsymbol{u_0} \\ \boldsymbol{u_1} \\ \vdots \\ \boldsymbol{u_{N-1}} \end{bmatrix}, \text{ where}$$

$$\boldsymbol{O_N} := \begin{bmatrix} \boldsymbol{C_0} \\ \boldsymbol{C_1}\boldsymbol{A_0} \\ \vdots \\ \boldsymbol{C_{N-1}}\boldsymbol{A_0}\cdots\boldsymbol{A_{N-2}} \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{F_N} := \begin{bmatrix} \boldsymbol{D_0} & 0 & & \\ \boldsymbol{C_1}\boldsymbol{B_0} & \boldsymbol{D_1} & & \\ \vdots & \vdots & \ddots & \\ \boldsymbol{C_{N-1}}\boldsymbol{A_0}\cdots\boldsymbol{A_{N-2}}\boldsymbol{B_0} & \boldsymbol{C_{N-2}}\boldsymbol{A_1}\cdots\boldsymbol{A_{N-2}}\boldsymbol{B_1} & \cdots & \boldsymbol{D_{N-1}} \end{bmatrix}.$$

In order to uniquely determine the initial state, $\boldsymbol{x_0}$, from the known inputs and outputs, the null space of $\boldsymbol{O_N}$ must be trivial. Therefore, the system is observable if and only if the rank of $\boldsymbol{O_N}$ is equal to n.

8.2.2 Observable Subspaces

In the previous section, we showed that for a discrete LTV system with initial condition $\boldsymbol{x_0}$, we can express all the outputs on the time interval $[0, N-1]$ as

$$\begin{bmatrix} \boldsymbol{y_0} \\ \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{N-1}} \end{bmatrix} = \boldsymbol{O_N x_0} + \boldsymbol{F_N} \begin{bmatrix} \boldsymbol{u_0} \\ \boldsymbol{u_1} \\ \vdots \\ \boldsymbol{u_{N-1}} \end{bmatrix}.$$

If $(\boldsymbol{x_0} - \boldsymbol{\bar{x}_0}) \in \mathcal{N}(\boldsymbol{O_N})$, then $\boldsymbol{O_N x_0} = \boldsymbol{O_N \bar{x}_0}$, so the initial conditions $\boldsymbol{x_0}$ and $\boldsymbol{\bar{x}_0}$ are not distinguishable for a set of inputs $\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}$ and set of outputs $\boldsymbol{y_0}, \boldsymbol{y_1}, \cdots, \boldsymbol{y_{N-1}}$. Furthermore, if $\boldsymbol{x_0} \in N(\boldsymbol{O_N})$, then the initial condition is not distinguishable from the zero vector $\mathbf{0}_n$ and thus cannot be uniquely determined. For this reason, this space is considered the unobservable subspace. The **observable subspace** is then $\mathbb{R}^n \setminus N(\boldsymbol{O_N})$.

8.2.3 Determining the Initial Condition

Suppose we want to determine the initial state of our system after observing a series of outputs resulting from known inputs. For convenience, we will define the following vectors composed of the known inputs and observed outputs:

$$\boldsymbol{Y_N} := \begin{bmatrix} \boldsymbol{y_0} \\ \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{N-1}} \end{bmatrix} \quad \text{and} \quad \boldsymbol{U_N} := \begin{bmatrix} \boldsymbol{u_0} \\ \boldsymbol{u_1} \\ \vdots \\ \boldsymbol{u_{N-1}} \end{bmatrix}.$$

Now we have the following linear matrix equation:

$$\boldsymbol{Y_N} = \boldsymbol{O_N x_0} + \boldsymbol{F_N U_N}.$$

A solution to this equation exists if and only if $\boldsymbol{Y_N} - \boldsymbol{F_N U_N} \in R(\boldsymbol{O_N})$. Furthermore, this solution is unique if and only if $\boldsymbol{O_N}$ has full column rank. In Section 8.2.1, we said that if the system is observable, then it has full column rank (i.e., rank n). Therefore, if the system is observable, the initial condition can always be uniquely determined by computing the following solution:

$$\boldsymbol{x_0} = (\boldsymbol{O_N^T O_N})^{-1} \boldsymbol{O_N^T} \Big(\boldsymbol{Y_N} - \boldsymbol{F_N U_N} \Big).$$

If the system is not (completely) observable, then the matrix $\boldsymbol{O_N}$ does not have full column rank and the initial state cannot be uniquely determined. We can define the best approximate of the initial state as the following optimal solution:

$$\boldsymbol{\hat{x}_0} = (\boldsymbol{O_N^T O_N})^{\dagger} \boldsymbol{O_N^T} \Big(\boldsymbol{Y_N} - \boldsymbol{F_N U_N} \Big).$$

Now suppose that the output of our system is described by

$$\boldsymbol{Y_N} = \boldsymbol{O_N x_0} + \boldsymbol{F_N U_N} + \boldsymbol{z},$$

where $\boldsymbol{z}$ is some unknown error or measurement noise. In this case, we may find that $\boldsymbol{Y_N} - \boldsymbol{F_N}\boldsymbol{U_N}$ is not in the range space of $\boldsymbol{O_N}$. In this case, let $\hat{\boldsymbol{Y}}_{\boldsymbol{N}} = \boldsymbol{O_N}\hat{\boldsymbol{x}}_{\boldsymbol{0}} + \boldsymbol{F_N}\boldsymbol{U_N}$. We aim to find the solution $\hat{\boldsymbol{x}}_{\boldsymbol{0}}$ such that $\hat{\boldsymbol{Y}}_{\boldsymbol{N}} - \boldsymbol{F_N}\boldsymbol{U_N}$ is as close to $\boldsymbol{Y_N} - \boldsymbol{F_N}\boldsymbol{U_N}$ as possible. Assuming the system is (completely) observable, the least squares solution is given by

$$\hat{\boldsymbol{x}}_{\boldsymbol{0}} = (\boldsymbol{O_N^T}\boldsymbol{O_N})^{-1}\boldsymbol{O_N^T}\Big(\boldsymbol{Y_N} - \boldsymbol{F_N}\boldsymbol{U_N}\Big).$$

If the system is not (completely) observable, then the matrix $\boldsymbol{O_N}$ does not have full column rank and the initial condition cannot be uniquely determined. The "best" solution in this case is given by

$$\hat{\boldsymbol{x}}_{\boldsymbol{0}} = (\boldsymbol{O_N^T}\boldsymbol{O_N})^{\dagger}\boldsymbol{O_N^T}\Big(\boldsymbol{Y_N} - \boldsymbol{F_N}\boldsymbol{U_N}\Big).$$

Chapter 9

Controllability/Observability: Discrete LTI Systems

9.1 Controllability of Discrete LTI Systems

Consider the discrete linear time-invariant (LTI) system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\ \boldsymbol{y_k} = \boldsymbol{C}\boldsymbol{x_k} + \boldsymbol{D}\boldsymbol{u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. We will determine the conditions under which this system is (completely) controllable on a discrete time interval.

9.1.1 Controllability Condition

Again, for discrete time systems, we are interested in whether the system is controllable on a finite time interval $[0, N]$, meaning that there exists a set of N inputs, $\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}$ such that system can reach any state $\boldsymbol{x_N}$ from an arbitrary initial state $\boldsymbol{x_0}$. At discrete time N, the state of the system is

$$\boldsymbol{x_N} = \boldsymbol{A}^N\boldsymbol{x_0} + \sum_{i=0}^{N-1} \boldsymbol{A}^{N-1-i}\boldsymbol{B}\boldsymbol{u_i}.$$

We can rewrite the sum of products using the $n \times Nn_i$ matrix defined as

$$\boldsymbol{Q_N} := \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \dots & \boldsymbol{A}^{N-1}\boldsymbol{B} \end{bmatrix}.$$

With this definition, we can express the state of the system at time N as

$$\boldsymbol{x_N} = \boldsymbol{A}^N\boldsymbol{x_0} + \boldsymbol{Q_N}\begin{bmatrix} \boldsymbol{u_{N-1}} \\ \boldsymbol{u_{N-2}} \\ \vdots \\ \boldsymbol{u_0} \end{bmatrix}.$$

In order for this system to be controllable, the matrix $\boldsymbol{Q_N}$ must be able to map the sequence of inputs to any vector in the state space, $\mathbb{R}^n$. Therefore, the system is controllable if and only if the rank of this matrix is equal to n.

9.1.2 Controllable Subspaces

In the previous section, we showed that for a discrete LTI system with initial state $\boldsymbol{x_0}$ at discrete time 0, we can express the state of the system at time N as

$$\boldsymbol{x_N} = \boldsymbol{A}^N \boldsymbol{x_0} + \boldsymbol{Q_N} \begin{bmatrix} \boldsymbol{u_{N-1}} \\ \boldsymbol{u_{N-2}} \\ \vdots \\ \boldsymbol{u_0} \end{bmatrix}, \text{ where}$$

$$\boldsymbol{Q_N} := \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \dots & \boldsymbol{A}^{N-1}\boldsymbol{B} \end{bmatrix}.$$

This tells us that the set of states reachable from the initial point $\boldsymbol{x_0}$ is

$$\boldsymbol{A}^N \boldsymbol{x_0} + R(\boldsymbol{Q_N}).$$

Furthermore, a state $\boldsymbol{x_N}$ is reachable at discrete time N if and only if

$$\left(\boldsymbol{x_N} - \boldsymbol{A}^N \boldsymbol{x_0}\right) \in R(\boldsymbol{Q_N}).$$

Recall that the system is controllable if and only if the range of $\boldsymbol{Q_N}$ is the entire state space, $\mathbb{R}^n$. If this is the case, every state $\boldsymbol{x_N}$ is reachable in N time steps.

9.1.3 Computing the Optimal Control

Suppose we want to find the best input sequence that allows us to move the state of our system from the state $\boldsymbol{x_0}$ at discrete time 0 to the state $\boldsymbol{x_N}$ at discrete time N. As shown in the previous section, if $\boldsymbol{x_N}$ is in the controllable subspace, there exists a solution $u \in \mathcal{U}$ to the linear equation for $\boldsymbol{x_N}$. To find the optimal solution, we can follow the same approach discussed in Section 8.1.3.

9.2 Observability of Discrete LTI Systems

Consider the discrete linear time-invariant (LTI) system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\ \boldsymbol{y_k} = \boldsymbol{C}\boldsymbol{x_k} + \boldsymbol{D}\boldsymbol{u_k} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. We will determine the conditions under which this system is (completely) observable on a discrete time interval.

9.2.1 Observability Condition

Again, for discrete time systems, we are interested in whether the system is observable on a finite time interval $[0, N-1]$, meaning that the initial state of the system can be uniquely determined by observing its inputs $(\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}})$ and outputs $(\boldsymbol{y_0}, \boldsymbol{y_1}, \cdots, \boldsymbol{y_{N-1}})$ over time. For the time-invariant case, recall that we can express the output of the system at time step k as

$$\boldsymbol{y_k} = \boldsymbol{C}\boldsymbol{A}^k\boldsymbol{x_0} + \sum_{i=0}^{k-1} \boldsymbol{C}\boldsymbol{A}^{k-1-i}\boldsymbol{B}\boldsymbol{u_i} + \boldsymbol{D}\boldsymbol{u_k}.$$

We can then express all of the outputs on the time interval $[0, N-1]$ as

$$\begin{bmatrix} \boldsymbol{y_0} \\ \boldsymbol{y_1} \\ \vdots \\ \boldsymbol{y_{N-1}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{N-1} \end{bmatrix} \boldsymbol{x_0} + \begin{bmatrix} \boldsymbol{D} & 0 & & \\ \boldsymbol{C}\boldsymbol{B} & \boldsymbol{D} & & \\ \vdots & \vdots & \ddots & \\ \boldsymbol{C}\boldsymbol{A}^{N-2}B & \boldsymbol{C}\boldsymbol{A}^{N-3}B & \cdots & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_{N-1}} \\ \boldsymbol{u_{N-2}} \\ \vdots \\ \boldsymbol{u_0} \end{bmatrix}.$$

Now let $\boldsymbol{O_N} \in \mathbb{R}^{Nn_o \times n}$ be the first matrix in the equation above defined as

$$\boldsymbol{O_N} := \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{N-1} \end{bmatrix}.$$

In order to uniquely determine the initial state, $\boldsymbol{x_0}$, from the known inputs and outputs, the null space of $\boldsymbol{O_N}$ must be trivial. Therefore, the system is observable if and only if the rank of $\boldsymbol{O_N}$ is equal to n.

9.2.2 Observable Subspaces

The unobservable and observable subspaces for a discrete LTI system can be found in the same way as in the time-varying case (Section 8.2.2), but the $\boldsymbol{O_N}$ matrix that should be used is the one defined in the previous section.

9.2.3 Determining the Initial Condition

Suppose we want to determine the initial state of our system after observing a series of outputs resulting from known inputs. To do so, we can follow the same approach discussed for discrete LTV systems in Section 8.2.3.

Part IV

Linear Control Design

Chapter 10

State Feedback & Observers

10.1 Full State Feedback

It is common to use **state feedback** to design a controller for a linear system that can drive the system to some desired state or set of states. State feedback involves the use of the state vector to compute the control action. We refer to controllers that use state feedback as **closed-loop** controllers and those that do not receive feedback about the state as **open-loop** controllers.

Full state feedback (FSF), or **pole placement**, is a method employed to place the closed-loop poles of the system in pre-determined locations in the complex plane. Placing poles is desirable because the location of the poles corresponds directly to the eigenvalues of the system, which determine the characteristics of the response of the system. The system must be controllable in order to implement this method, which we will discuss how to do for both **single-input single-output (SISO)** systems and **multiple-input multiple-output (MIMO)** systems. While this method cannot be implemented for systems that are not controllable, we will also discuss pole placement for stabilizable systems.

10.1.1 Feedback for Controllable SISO Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{b}u(t) \\ y(t) = \boldsymbol{c}^T x(t) + du(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, and $y(t) \in \mathbb{R}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{c} \in \mathbb{R}^n$, and $d \in \mathbb{R}$. To apply linear state feedback, we use an input function of the form

$$u(t) = -\boldsymbol{f}^T x(t) + r(t)$$

where $\boldsymbol{f} \in \mathbb{R}^n$ is a feedback vector and $r(t) \in \mathbb{R}$ is a reference signal. Under this control, the dynamics of the system are described by the differential equation:

$$\begin{aligned}\dot{x}(t) &= \boldsymbol{A}x(t) + \boldsymbol{b}\big(-\boldsymbol{f}^T x(t) + r(t)\big)\\ &= (\boldsymbol{A} - \boldsymbol{b}\boldsymbol{f}^T)x(t) + \boldsymbol{b}r(t).\end{aligned}$$

Now we effectively have a new system with the dynamics matrix $\boldsymbol{A} - \boldsymbol{b}\boldsymbol{f}^T$. We often refer to this matrix as the **closed-loop dynamics matrix**.

Theorem: If the system is controllable, there exists a feedback vector, $\boldsymbol{f}$, such that the eigenvalues of $\boldsymbol{A} - \boldsymbol{b}\boldsymbol{f}^T$ can be placed anywhere in the complex plane.

Proof: Recall from Section 7.3.1 that a continuous LTI SISO system is controllable if and only if there exists an invertible matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ such that

$$\tilde{\boldsymbol{A}} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{b}} = \boldsymbol{T}\boldsymbol{b} = \begin{bmatrix} 0\\ 0\\ \vdots\\ 0\\ 1\end{bmatrix},$$

where $a_1, a_2, \cdots, a_n$ are the coefficients in the denominator of the transfer function for the given system. Let $\tilde{x}(t) = \boldsymbol{T}x(t)$ be the system governed by the dynamics

$$\dot{\tilde{x}}(t) = \tilde{\boldsymbol{A}}\tilde{x}(t) + \tilde{\boldsymbol{b}}\tilde{u}(t).$$

Now consider the following linear state feedback for this new system:

$$\tilde{u}(t) = -\tilde{\boldsymbol{f}}^T\tilde{x}(t) + r(t),$$

where $\tilde{\boldsymbol{f}}^T = \boldsymbol{f}^T\boldsymbol{T}^{-1}$ is a feedback matrix whose ith element is $\tilde{f}_i$. Under this control law, the dynamics of our new system are governed by

$$\dot{\tilde{x}}(t) = (\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{b}}\tilde{\boldsymbol{f}}^T)\tilde{x}(t) + \tilde{\boldsymbol{b}}r(t).$$

The closed-loop dynamics matrix for this system can then be expressed as

$$\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{b}}\tilde{\boldsymbol{f}}^T = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\\ -a_n - \tilde{f}_1 & -a_{n-1} - \tilde{f}_2 & -a_{n-2} - \tilde{f}_3 & \cdots & -a_1 - \tilde{f}_n \end{bmatrix}.$$

Now we can see that the characteristic polynomial of the matrix $\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{b}}\tilde{\boldsymbol{f}}^T$ has the coefficients $-a_i - \tilde{f}_{n+1-i}$ for $i = 1, \cdots, n$. Because the elements of $\tilde{\boldsymbol{f}}$ are arbitrarily chosen real numbers, the coefficients of the closed-loop characteristic polynomial can be given any desired values, which means that the closed-loop poles can be assigned to arbitrary locations in the complex plane.

The closed-loop dynamics matrix for transformed system is related to the closed-loop dynamics matrix for the original matrix in the following way:

$$\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{b}}\tilde{\boldsymbol{f}}^T = \boldsymbol{T}\big(\boldsymbol{A}\boldsymbol{T}^{-1} - \boldsymbol{b}\boldsymbol{f}^T\big)\boldsymbol{T}^{-1}.$$

Because these matrices are similar, they have the same characteristic polynomial. Therefore, if the original system is controllable, we can arbitrarily select the poles of the both the transformed and the original system.

Designing the Feedback Vector: Suppose we want to choose the feedback vector, $\boldsymbol{f}$, such that the eigenvalues of the closed-loop dynamics matrix, $\boldsymbol{A}-\boldsymbol{b}\boldsymbol{f}^T$, are $\lambda_1, \cdots, \lambda_n$, where $\lambda_i \in \mathbb{C}_{--}$. Note that we generally choose the eigenvalues such that the closed-loop system is stable. We can design $\boldsymbol{f}$ as follows:

1. Compute the actual closed loop characteristic polynomial:

$$\begin{aligned}\chi_{\boldsymbol{A}-\boldsymbol{b}\boldsymbol{f}^T}(s) &= \det\big(s\boldsymbol{I_n} - (\boldsymbol{A} - \boldsymbol{b}\boldsymbol{f}^T)\big)\\ &= s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0\end{aligned}$$

2. Determine the desired closed loop characteristic polynomial:

$$\begin{aligned}\Pi(s) &= (s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_n)\\ &= s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0\end{aligned}$$

3. Equate the coefficients of the actual and desired characteristic polynomials and solve for the elements of $\boldsymbol{f}$:

$$\alpha_0 = \beta_0, \alpha_1 = \beta_1, \cdots, \alpha_{n-1} = \beta_{n-1}$$

10.1.2 Feedback for Controllable MIMO Systems

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases}\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t)\\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t)\\ x(t_0) = \boldsymbol{x_0}\end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. Linear state feedback is now of the form

$$u(t) = -\boldsymbol{F}x(t) + r(t)$$

where $\boldsymbol{F} \in \mathbb{R}^{n_i\times n}$ is a feedback matrix and $r(t) \in \mathbb{R}^{n_i}$ is a reference signal. Under this control, the dynamics of the system are described by

$$\begin{aligned}\dot{x}(t) &= \boldsymbol{A}x(t) + \boldsymbol{B}\big(-\boldsymbol{F}x(t) + r(t)\big)\\ &= (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{F})x(t) + \boldsymbol{B}r(t).\end{aligned}$$

Now we effectively have a new system with the dynamics matrix $\boldsymbol{A}-\boldsymbol{BF}$, which we refer to as the **closed-loop dynamics matrix**. If the original system is controllable, there exists a feedback matrix, $\boldsymbol{F}$, such that the eigenvalues of $\boldsymbol{A}-\boldsymbol{BF}$ can be placed anywhere in the complex plane.

As with the single-input single-output (SISO) system, $\boldsymbol{A}-\boldsymbol{BF}$ has n eigenvalues. However, for the multiple-input multiple-output (MIMO) system, the feedback matrix, $\boldsymbol{F}$, has more than n elements, so there are likely multiple valid solutions for $\boldsymbol{F}$. In practice, we often choose $\boldsymbol{F}$ based on the cost of using certain inputs over others and may choose the minimum norm solution.

10.1.3 Feedback for Stabilizable Systems

If a system is not controllable, then it must have some uncontrollable mode (refer to Section 5.4.2). We are unable to move this mode, so we can no longer place all of the eigenvalues of the closed-loop dynamics matrix anywhere in the complex plane. However, if the system is stabilizable, any uncontrollable modes are stable. While we cannot move these uncontrollable modes, we can still move all of the controllable modes such that the closed loop system is stable.

10.1.4 Closed Loop Transfer Function

Consider the continuous LTI system described by the equations

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o \times n_i}$. Recall that the transfer function for this system is

$$H_{OL}(s) = \frac{\hat{y}(s)}{\hat{u}(s)} = \boldsymbol{C}(s\boldsymbol{I_n} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D}.$$

Suppose we implement linear state feedback of the form $u(t) = -\boldsymbol{F}x(t) + r(t)$, where $\boldsymbol{F} \in \mathbb{R}^{n_i \times n}$ is a feedback matrix and $r(t) \in \mathbb{R}^{n_i}$ is a reference signal. Now the closed-loop system can be expressed in terms of the following equations:

$$\begin{cases} \dot{x}(t) = (\boldsymbol{A} - \boldsymbol{BF})x(t) + \boldsymbol{B}r(t) \\ y(t) = (\boldsymbol{C} - \boldsymbol{DF})x(t) + \boldsymbol{D}r(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases}.$$

Now we can see the transfer function for this closed-loop system is given by

$$H_{CL}(s) = \frac{\hat{y}(s)}{\hat{r}(s)} = (\boldsymbol{C} - \boldsymbol{DF})\big(s\boldsymbol{I_n} - (\boldsymbol{A} - \boldsymbol{BF})\big)^{-1}\boldsymbol{B} + \boldsymbol{D}.$$

10.2 Full State Observers

Often, we cannot observe the full state of a system directly and we seek to estimate it from the output. To do so, we can design an **observer**.

10.2.1 Observers for Observable SISO Systems

Consider a continuous LTI SISO system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{b}u(t) \\ y(t) = \boldsymbol{c}^T x(t) + du(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}$, and $y(t) \in \mathbb{R}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{b} \in \mathbb{R}^n$, $\boldsymbol{c} \in \mathbb{R}^n$, and $d \in \mathbb{R}$. Suppose we cannot directly access the full state, $x(t)$, but we can access the output, $y(t)$. We want to find the estimated state, $\hat{x}(t)$, using this output. We can design an observer with the observer gain vector $\boldsymbol{t} \in \mathbb{R}^n$ such that the estimated system can be described by the equations

$$\begin{cases} \dot{\hat{x}}(t) = \boldsymbol{A}\hat{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{t}\big(y(t) - \hat{y}(t)\big) \\ \hat{y}(t) = \boldsymbol{c}^T \hat{x}(t) + du(t) \end{cases}.$$

From these equations, we can see that the estimated state obeys the following:

$$\begin{aligned} \dot{\hat{x}}(t) &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{t}\big(y(t) - \hat{y}(t)\big) \\ &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{t}\Big(\big(\boldsymbol{c}^T x(t) + du(t)\big) - \big(\boldsymbol{c}^T \hat{x}(t) + du(t)\big)\Big) \\ &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{t}\boldsymbol{c}^T x(t) - \boldsymbol{t}\boldsymbol{c}^T \hat{x}(t) \\ &= (\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T)\hat{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{t}\boldsymbol{c}^T x(t) \end{aligned}$$

Let the error, $e(t)$, be the difference between the estimated state and the true state. The dynamics of the error are governed by the following equation:

$$\begin{aligned} \dot{e}(t) &= \dot{\hat{x}}(t) - \dot{x}(t) \\ &= \Big((\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T)\hat{x}(t) + \boldsymbol{b}u(t) + \boldsymbol{t}\boldsymbol{c}^T x(t)\Big) - \big(\boldsymbol{A}x(t) + \boldsymbol{b}u(t)\big) \\ &= (\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T)\hat{x}(t) + (\boldsymbol{t}\boldsymbol{c}^T - \boldsymbol{A})x(t) \\ &= (\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T)\big(\hat{x}(t) - x(t)\big) \\ &= (\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T)e(t) \end{aligned}$$

Theorem: If the system is observable, there exists a gain vector, $\boldsymbol{t}$, such that the eigenvalues of $\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T$ can be placed anywhere in the complex plane.

I will not include the proof of this theorem here because it is very similar to the one given for the state feedback theorem in Section 10.1.1. If we choose the gain matrix such that all of the eigenvalues of $\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T$ are in the open left half plane, then the error will approach zero exponentially. This then implies that the estimated state will approach the true state, irrespective of the input.

Designing the Gain Vector: Suppose we want to choose the observer gain vector, $\boldsymbol{t}$, such that the eigenvalues of the error dynamics matrix, $\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T$, are $\lambda_1, \cdots, \lambda_n$, where $\lambda_i \in \mathbb{C}_{--}$. We can design $\boldsymbol{t}$ in the following way:

1. Compute the actual error characteristic polynomial:
$$\begin{aligned}\chi_{\boldsymbol{A}-\boldsymbol{t}\boldsymbol{c}^T}(s) &= \det\big(s\boldsymbol{I_n} - (\boldsymbol{A} - \boldsymbol{t}\boldsymbol{c}^T)\big)\\ &= s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0\end{aligned}$$

2. Determine the desired error characteristic polynomial:
$$\begin{aligned}\Pi(s) &= (s-\lambda_1)(s-\lambda_2)\cdots(s-\lambda_n)\\ &= s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0\end{aligned}$$

3. Equate the coefficients of the actual and desired characteristic polynomials and solve for the elements of $\boldsymbol{t}$:
$$\alpha_0 = \beta_0, \alpha_1 = \beta_1, \cdots, \alpha_{n-1} = \beta_{n-1}$$

10.2.2 Observers for Observable MIMO Systems

Consider a continuous LTI MIMO system described by

$$\begin{cases}\dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t)\\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t)\\ x(t_0) = \boldsymbol{x_0}\end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$. As in the previous section, suppose we cannot directly access the full state, and we want to find the estimated state, $\hat{x}(t)$, using the output. We can design an observer with the observer gain matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n_o}$ such that the estimated system can be described by the equations

$$\begin{cases}\dot{\hat{x}}(t) = \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}\big(y(t) - \hat{y}(t)\big)\\ \hat{y}(t) = \boldsymbol{C}\hat{x}(t) + \boldsymbol{D}u(t)\end{cases}.$$

As with the SISO system, the state estimation error is given by $e(t) = \hat{x}(t) - x(t)$, and the error dynamics can be described by

$$\dot{e}(t) = (\boldsymbol{A} - \boldsymbol{T}\boldsymbol{C})e(t).$$

If the system is observable, there exists an observer gain matrix, $\boldsymbol{T}$, such that the eigenvalues of the error dynamics matrix, $\boldsymbol{A} - \boldsymbol{T}\boldsymbol{C}$, can be placed anywhere in the complex plane. As before, $\boldsymbol{A} - \boldsymbol{T}\boldsymbol{C}$ has n eigenvalues, but for the MIMO system, the gain matrix has more than n elements, so it likely has multiple valid solutions. In practice, we often choose $\boldsymbol{T}$ such that the gains are not too large.

10.2.3 Observers for Detectable Systems

If a system is not observable, then it must have some unobservable mode (refer to Section 5.4.3). We are unable to move this mode, so we can no longer place all of the eigenvalues of the error dynamics matrix anywhere in the complex plane. However, if the system is detectable, any unobservable modes are stable. While we cannot move these unobservable modes, we can still move all of the observable modes such that the estimated state converges to the true state.

10.2.4 Observer Transfer Function

Consider the continuous LTI system described by the equations

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o \times n_i}$. As we showed previously, we can design an observer with gain matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n_o}$ such that the estimated system can be described by

$$\begin{cases} \dot{\hat{x}}(t) = \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}\big(y(t) - \hat{y}(t)\big) \\ \hat{y}(t) = \boldsymbol{C}\hat{x}(t) + \boldsymbol{D}u(t) \end{cases}.$$

From these equations, we can see that the estimated state obeys the following:

$$\begin{aligned} \dot{\hat{x}}(t) &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}y(t) - \boldsymbol{T}\hat{y}(t) \\ &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}y(t) - \boldsymbol{T}\big(\boldsymbol{C}\hat{x}(t) + \boldsymbol{D}u(t)\big) \\ &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}y(t) - \boldsymbol{TC}\hat{x}(t) - \boldsymbol{TD}u(t) \\ &= (\boldsymbol{A} - \boldsymbol{TC})\hat{x}(t) + (\boldsymbol{B} - \boldsymbol{TD})u(t) + \boldsymbol{T}y(t) \end{aligned}$$

The observer receives both the control signal, $u(t)$, and the output of the original system, $y(t)$, as input. We can combine these into a single input variable:

$$\tilde{u}(t) := \begin{bmatrix} u(t) \\ y(t) \end{bmatrix}.$$

This now allows us the express the dynamics of the estimated system as

$$\dot{\hat{x}}(t) = (\boldsymbol{A} - \boldsymbol{TC})\hat{x}(t) + \begin{bmatrix} (\boldsymbol{B} - \boldsymbol{TD}) & \boldsymbol{T} \end{bmatrix} \tilde{u}(t).$$

With this new input, we can express the output of the observer as

$$\hat{y}(t) = \boldsymbol{C}\hat{x}(t) + \begin{bmatrix} \boldsymbol{D} & 0 \end{bmatrix} \tilde{u}(t).$$

Now we can see that the transfer function for the estimated system is

$$H(s) = \frac{y(s)}{\tilde{u}(s)} = \boldsymbol{C}\big(s\boldsymbol{I_n} - (\boldsymbol{A} - \boldsymbol{TC})\big)^{-1} \begin{bmatrix} (\boldsymbol{B} - \boldsymbol{TD}) & \boldsymbol{T} \end{bmatrix} + \begin{bmatrix} \boldsymbol{D} & 0 \end{bmatrix}.$$

10.3 Combining State Feedback and Observers

Often, we want to design a closed-loop controller using full state feedback but do not have access to the full state. In these cases, we need to use an observer to estimate the state. In this section, we will discuss how to combine full state feedback design (Section 10.1) with observer design (Section 10.2). An example of a system with an observer and state feedback is shown in Figure 10.1.

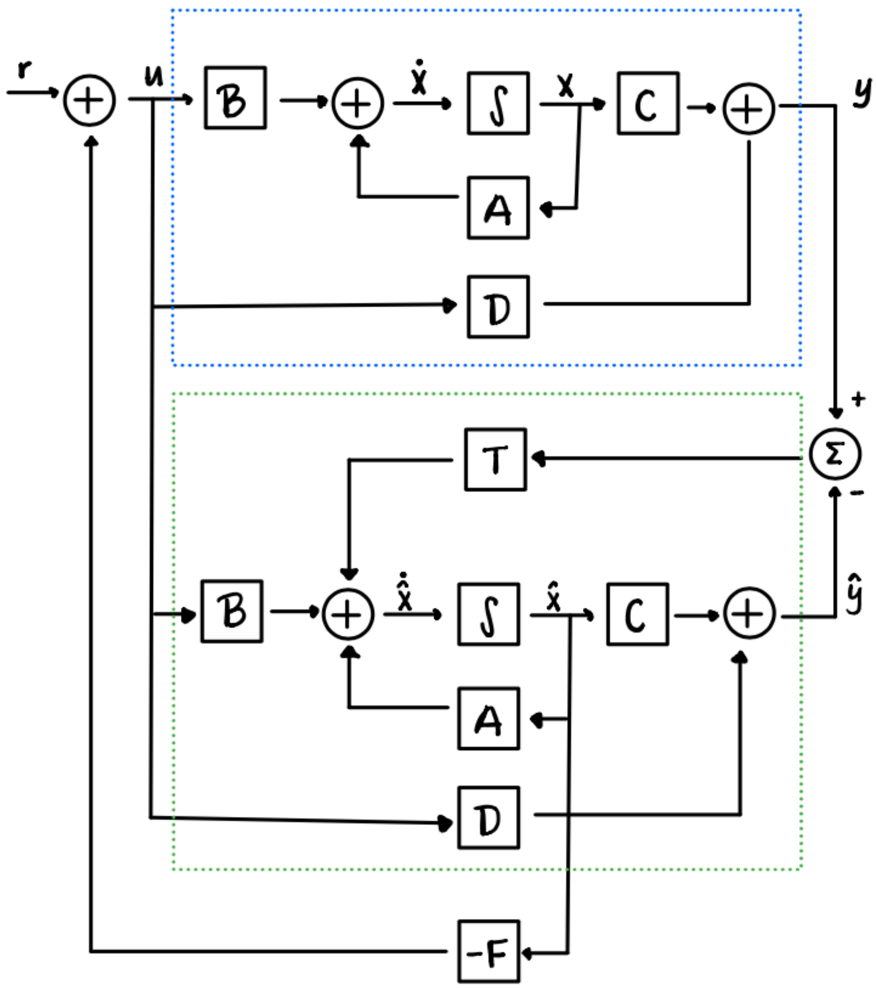

Figure 10.1: This is the complete configuration for a system with an observer and state feedback. The plant is shown in the blue dotted lines, the observer is shown in the green dotted lines, and the feedback matrix is at the bottom.

10.3.1 Closed-Loop System with an Observer

Consider the continuous LTI system described by the equations

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(t_0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o \times n_i}$. Suppose we want to implement linear state feedback but cannot access the full state directly. We can design an observer with the gain matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n_o}$ such that the estimated system can be described by

$$\begin{cases} \dot{\hat{x}}(t) = \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}\big(y(t) - \hat{y}(t)\big) \\ \hat{y}(t) = \boldsymbol{C}\hat{x}(t) + \boldsymbol{D}u(t) \end{cases}.$$

Now we can use this estimated state to implement linear state feedback with the input function $u(t) = -\boldsymbol{F}\hat{x}(t) + r(t)$, where $\boldsymbol{F} \in \mathbb{R}^{n_i \times n}$ and $r(t) \in \mathbb{R}^{n_i}$. The dynamics of the closed loop system can now be described by the following:

$$\begin{aligned} \dot{x}(t) &= \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ &= \boldsymbol{A}x(t) + \boldsymbol{B}\big(-\boldsymbol{F}\hat{x}(t) + r(t)\big) \\ &= \boldsymbol{A}x(t) - \boldsymbol{B}\boldsymbol{F}\hat{x}(t) + \boldsymbol{B}r(t) \end{aligned}$$

We can also describe the dynamics of the estimated state by the following:

$$\begin{aligned} \dot{\hat{x}}(t) &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}u(t) + \boldsymbol{T}(y(t) - \hat{y}(t)) \\ &= \boldsymbol{A}\hat{x}(t) + \boldsymbol{B}\big(-\boldsymbol{F}\hat{x}(t) + r(t)\big) + \boldsymbol{T}\Big(\big(\boldsymbol{C}x(t) + \boldsymbol{D}u(t)\big) - \big(\boldsymbol{C}\hat{x}(t) + \boldsymbol{D}u(t)\big)\Big) \\ &= \boldsymbol{A}\hat{x}(t) - \boldsymbol{B}\boldsymbol{F}\hat{x}(t) + \boldsymbol{B}r(t) + \boldsymbol{T}\boldsymbol{C}x(t) - \boldsymbol{T}\boldsymbol{C}\hat{x}(t) \\ &= (\boldsymbol{A} - \boldsymbol{B}\boldsymbol{F} + \boldsymbol{T}\boldsymbol{C})\hat{x}(t) + \boldsymbol{B}r(t) + \boldsymbol{T}\boldsymbol{C}x(t) \end{aligned}$$

Recall that we defined the error as the difference between the estimated and true state (i.e. $e(t) = \hat{x}(t) - x(t)$). The dynamics of the error are now given by

$$\begin{aligned} \dot{e}(t) &= \dot{\hat{x}}(t) - \dot{x}(t) \\ &= \Big((\boldsymbol{A} - \boldsymbol{B}\boldsymbol{F} + \boldsymbol{T}\boldsymbol{C})\hat{x}(t) + \boldsymbol{B}r(t) + \boldsymbol{T}\boldsymbol{C}x(t)\Big) - \Big(\boldsymbol{A}x(t) - \boldsymbol{B}\boldsymbol{F}\hat{x}(t) + \boldsymbol{B}r(t)\Big) \\ &= (\boldsymbol{A} - \boldsymbol{T}\boldsymbol{C})\hat{x}(t) + (\boldsymbol{T}\boldsymbol{C} - \boldsymbol{A})x(t) \\ &= (\boldsymbol{A} - \boldsymbol{T}\boldsymbol{C})(\hat{x}(t) - x(t)) \\ &= (\boldsymbol{A} - \boldsymbol{T}\boldsymbol{C})e(t) \end{aligned}$$

Notice that the error dynamics for this closed-loop system are exactly the same as they were when we did not consider linear state feedback. Now we can also write the dynamics of the true state in terms of the state and error:

$$\begin{aligned}\dot{x}(t) &= \boldsymbol{A}x(t) - \boldsymbol{BF}\hat{x}(t) + \boldsymbol{B}r(t)\\ &= \boldsymbol{A}x(t) - \boldsymbol{BF}(e(t) + x(t)) + \boldsymbol{B}r(t)\\ &= (\boldsymbol{A} - \boldsymbol{BF})x(t) - \boldsymbol{BF}e(t) + \boldsymbol{B}r(t)\end{aligned}$$

The output can also be expressed in terms of the true state and error:

$$\begin{aligned}y(t) &= \boldsymbol{C}x(t) + \boldsymbol{D}u(t)\\ &= \boldsymbol{C}x(t) + \boldsymbol{D}\big(-\boldsymbol{F}\hat{x}(t) + r(t)\big)\\ &= \boldsymbol{C}x(t) - \boldsymbol{DF}\hat{x}(t) + \boldsymbol{D}r(t)\\ &= Cx - DF(e + x) + Dr\\ &= (\boldsymbol{C} - \boldsymbol{DF})x(t) - \boldsymbol{DF}e(t) + \boldsymbol{D}r(t)\end{aligned}$$

Combining the dynamics of the actual state with the dynamics of the state estimation error, we can express the dynamics of the closed loop system as

$$\begin{bmatrix}\dot{x}(t)\\ \dot{e}(t)\end{bmatrix} = \begin{bmatrix}(\boldsymbol{A} - \boldsymbol{BF}) & -\boldsymbol{BF}\\ 0 & (\boldsymbol{A} - \boldsymbol{TC})\end{bmatrix}\begin{bmatrix}x(t)\\ e(t)\end{bmatrix} + \begin{bmatrix}\boldsymbol{B}\\ 0\end{bmatrix}r(t)$$

$$y(t) = \begin{bmatrix}(\boldsymbol{C} - \boldsymbol{DF}) & -\boldsymbol{DF}\end{bmatrix}\begin{bmatrix}x(t)\\ e(t)\end{bmatrix} + \boldsymbol{D}r(t)$$

Notice that the dynamics matrix of this system is a block triangular matrix, whose eigenvalues are those of the diagonal blocks: $\lambda(\boldsymbol{A} - \boldsymbol{BF}) \cup \lambda(\boldsymbol{A} - \boldsymbol{TC})$.

10.3.2 Transfer Function

Consider the LTI system given in the previous section with the full state observer and linear state feedback implemented as described. Let's define a new variable:

$$\tilde{x}(t) := \begin{bmatrix}x(t)\\ e(t)\end{bmatrix}.$$

To simplify our notation, let's also define the following matrices:

$$\tilde{\boldsymbol{A}} = \begin{bmatrix}(\boldsymbol{A} - \boldsymbol{BF}) & -\boldsymbol{BF}\\ 0 & (\boldsymbol{A} - \boldsymbol{TC})\end{bmatrix}, \qquad \tilde{\boldsymbol{B}} = \begin{bmatrix}\boldsymbol{B}\\ 0\end{bmatrix},$$

$$\tilde{\boldsymbol{C}} = \begin{bmatrix}(\boldsymbol{C} - \boldsymbol{DF}) & -\boldsymbol{DF}\end{bmatrix}, \qquad \tilde{\boldsymbol{D}} = \boldsymbol{D}.$$

Now the closed loop system with observer can be described by the equations

$$\begin{cases}\dot{\tilde{x}}(t) = \tilde{\boldsymbol{A}}\tilde{x}(t) + \tilde{\boldsymbol{B}}r(t)\\ y(t) = \tilde{\boldsymbol{C}}\tilde{x}(t) + \tilde{\boldsymbol{D}}r(t)\end{cases}.$$

We can now see that the closed loop transfer function for this system is

$$H_{CL}(s) = \frac{y(s)}{r(s)} = \tilde{\boldsymbol{C}}(s\boldsymbol{I_n} - \tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{B}} + \tilde{\boldsymbol{D}}.$$

10.3.3 Separation Principle

The **separation theorem** says that for a continuous LTI system, we can design the linear state feedback control matrix, $\boldsymbol{F}$, and the full state observer gain matrix, $\boldsymbol{T}$, independently. We can first design the full state observer gain matrix to provide asymptotically-accurate state estimates. Then we can design the linear state feedback matrix as though we can directly access the full state.

When choosing the desired poles of the closed-loop dynamics matrix, $(\boldsymbol{A}-\boldsymbol{BF})$, and the error dynamics matrix, $(\boldsymbol{A}-\boldsymbol{TC})$, we follow a few rules of thumb:

1. We typically design $\boldsymbol{F}$ such that the eigenvalues of $\boldsymbol{A}-\boldsymbol{BF}$ are not too far from the eigenvalues of $\boldsymbol{A}$. If we place the eigenvalues too far in the left half plane, we have large gains, which can lead to actuator saturation.

2. We typically design $\boldsymbol{T}$ such that the eigenvalues of $\boldsymbol{A}-\boldsymbol{TC}$ are also not too far from the eigenvalues of $\boldsymbol{A}$. If we place the eigenvalues too far in the left half plane, we again require large gains, which greatly amplifies any noise and/or errors in the system.

3. We tend to design $\boldsymbol{F}$ and $\boldsymbol{T}$ such that the eigenvalues of $\boldsymbol{A}-\boldsymbol{TC}$ are around 2-3 times farther to the left than the eigenvalues of $\boldsymbol{A}-\boldsymbol{BF}$. This is done so that the state estimation error converges to zero faster.

Chapter 11

Linear Quadratic Regulator (LQR)

11.1 Dynamic Programming

In the previous chapter, we discussed how to control a system using linear state feedback (with and without the use of an observer). Another very popular controller for both continuous and discrete LTI system is the **linear quadratic regulator (LQR)**, which relies on an optimization principle called **dynamic programming**. The idea of dynamic programming is to simplify a complicated problem by breaking it down into simpler sub-problems in a recursive manner.

Suppose we want to find the optimal control sequence over some time interval $[0, N]$. **Bellman's principle of optimality** is a dynamic programming principle that says if we have found the optimal sequence on the interval $[0, N]$, then the resulting sequence is also optimal on all subintervals of the form $[t, N]$, where $t > 0$. This means that the optimal control sequence over the entire time horizon remains optimal at intermediate points in time. In the following sections, I'll show how this principle is used to design the linear quadratic regulator.

11.2 Discrete Time LQR

The linear quadratic regulator (LQR) is applicable to both continuous and discrete LTI systems with some variations. I will begin by discussing how it is applied to discrete time systems using Bellman's principle of optimality.

11.2.1 Vanilla LQR Problem

I will begin with the "vanilla" LQR problem and then discuss some variations.

LQR Optimization Problem

Consider a discrete linear time-invariant (LTI) system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\ \boldsymbol{x_0} = x_{init} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, and $k \in [0, N]$. To design a linear quadratic regulator, we want to find the control sequence $U = \{\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}\}$ that minimizes the following quadratic cost function:

$$J(U, \boldsymbol{x_0}) \coloneqq \sum_{\tau=0}^{N-1} \left(\boldsymbol{x_\tau^T Q x_\tau} + \boldsymbol{u_\tau^T R u_\tau} \right) + \boldsymbol{x_N^T Q_f x_N},$$

where $\boldsymbol{Q}, \boldsymbol{Q_f} \in \mathbb{R}^{n \times n}$ are some positive semidefinite matrices and $\boldsymbol{R} \in \mathbb{R}^{n_i \times n_i}$ is a positive definite matrix. In this cost function, $\boldsymbol{x_\tau^T Q x_\tau}$ is the cost for deviating from the desired state (zero) at time τ, $\boldsymbol{u_\tau^T R u_\tau}$ is the cost for applying control at time τ, and $\boldsymbol{x_N^T Q_f x_N}$ is the cost for deviating from the desired terminal state. The cost $\boldsymbol{x_\tau^T Q x_\tau} + \boldsymbol{u_\tau^T R u_\tau}$ is called the **running cost**, and the cost $\boldsymbol{x_N^T Q_f x_N}$ is called the **final cost**. To find the control that minimizes this cost function, while obeying our dynamics, we set up the following optimization problem:

$$\begin{aligned} \hat{U} = \ & \arg\min_U \ J(U, \boldsymbol{x_0}) \\ \text{s.t.} \quad & \boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\ & \boldsymbol{x_0} = x_{init} \end{aligned}$$

Bellman's Principle

We can solve this problem using a dynamic programming approach. Recall Bellman's principle of optimality says that a sequence which is optimal on an entire time interval, $[0, N]$, is also optimal on all subintervals, $[k, N]$ for $k > 0$. Let's then define the control sequence on this interval as $U_k = \{\boldsymbol{u_k}, u_{k+1}, \cdots, u_{N-1}\}$. We can then express the quadratic cost on just this subinterval as

$$J(U_k, \boldsymbol{x_k}) \coloneqq \sum_{\tau=k}^{N-1} \left(\boldsymbol{x_\tau^T Q x_\tau} + \boldsymbol{u_\tau^T R u_\tau} \right) + \boldsymbol{x_N^T Q_f x_N}.$$

We will define the **minimum "cost-to-go"** from a given state, $\boldsymbol{x_k}$, as

$$\begin{aligned} J_k^*(\boldsymbol{x_k}) = & \min_{U_k} \ J(U_k, \boldsymbol{x_k}) \\ \text{s.t.} \quad & \boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\ & \boldsymbol{x_0} = x_{init} \end{aligned}$$

From Bellman's principle of optimality, we can recognize that the minimum cost-to-go from a given state, $\boldsymbol{x_k}$, is equal to the cost incurred at time k (the

stage cost) plus the minimum cost-to-go from the next state, $\boldsymbol{x_{k+1}}$. This then allows us to express the minimum cost-to-go from state $\boldsymbol{x_k}$ as

$$\begin{aligned} J_k^*(\boldsymbol{x_k}) = &\min_{\boldsymbol{u_k}} \left(\boldsymbol{x_k^T Q x_k} + \boldsymbol{u_k^T R u_k} + \hat{J}_{k+1}(\boldsymbol{x_{k+1}})\right) \\ &\text{s.t.} \quad \boldsymbol{x_{k+1}} = \boldsymbol{A x_k} + \boldsymbol{B u_k} \\ &\qquad\;\; \boldsymbol{x_0} = x_{init} \end{aligned}$$

Plugging the constraint into our objective, we can express the minimum cost as

$$J_k^*(\boldsymbol{x_k}) = \min_{\boldsymbol{u_k}} \left\{\boldsymbol{x_k^T Q x_k} + \boldsymbol{u_k^T R u_k} + \hat{J}_{k+1}(\boldsymbol{A x_k} + \boldsymbol{B u_k})\right\}.$$

Optimal Cost & Control

Now let's claim that the minimum cost-to-go and optimal control at time k are

$$J_k^*(\boldsymbol{x_k}) = \boldsymbol{x_k^T P_k x_k} \qquad \text{and} \qquad \boldsymbol{\hat{u}_k} = -\boldsymbol{K_k x_k},$$

where the matrices $\boldsymbol{P_k} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{K_k} \in \mathbb{R}^{n_i \times n}$ are defined such that

$$\boldsymbol{P_k} = \begin{cases} \boldsymbol{Q_f} & \text{for } k = N \\ \boldsymbol{Q} + \boldsymbol{K_k^T R K_k} + (\boldsymbol{A} - \boldsymbol{B K_k})^T \boldsymbol{P_{k+1}} (\boldsymbol{A} - \boldsymbol{B K_k}) & \text{for } k = 0, 1, \cdots, N-1 \end{cases}$$

$$\boldsymbol{K_k} = (\boldsymbol{R} + \boldsymbol{B}^T \boldsymbol{P_{k+1} B})^{-1} \boldsymbol{B}^T \boldsymbol{P_{k+1} A} \text{ for } k = 0, 1, \cdots, N-1$$

Proof of Optimality

We can prove that this claim is true by induction. For $k = N$, the minimum cost-to-go is simply the cost incurred at $k = N$, which is the final cost, $\boldsymbol{x_N^T Q_f x_N}$. Therefore, at $k = N$, the minimum cost-to-go aligns with the claim that the optimal cost is $J_k^*(\boldsymbol{x_k}) = \boldsymbol{x_k^T P_k x_k}$ with $\boldsymbol{P_k} = \boldsymbol{Q_f}$ for $k = N$.

We just showed that our claim is true for the initial condition $k = N$. Now we want to show that if it is true for $k = t$, then it is also true for $k = t - 1$. Recall that we previously defined the minimum cost-to-go such that

$$J_{t-1}^*(\boldsymbol{x_{t-1}}) = \min_{\boldsymbol{u_{t-1}}} \left\{\boldsymbol{x_{t-1}^T Q x_{t-1}} + \boldsymbol{u_{t-1}^T R u_{t-1}} + J_t^*(\boldsymbol{A x_{t-1}} + \boldsymbol{B u_{t-1}})\right\}.$$

If we assume our claim holds for $k = t$, the minimum cost-to-go at this time is

$$J_t^*(\boldsymbol{x_t}) = \boldsymbol{x_t^T P_t x_t}.$$

Using this assumption in our expression for the cost-to-go as time $k = t - 1$,

$$J_{t-1}^*(\boldsymbol{x_{t-1}}) = \min_{\boldsymbol{u_{t-1}}} \left\{\boldsymbol{x_{t-1}^T Q x_{t-1}} + \boldsymbol{u_{t-1}^T R u_{t-1}} + (\boldsymbol{A x_{t-1}} + \boldsymbol{B u_{t-1}})^T \boldsymbol{P_t} (\boldsymbol{A x_{t-1}} + \boldsymbol{B u_{t-1}})\right\}.$$

We can expand and simplify this expression for the optimal cost-to-go as

$$J_{t-1}^*(\boldsymbol{x_{t-1}}) = \min_{\boldsymbol{u_{t-1}}} \left\{\boldsymbol{x_{t-1}^T}\big(\boldsymbol{Q} + \boldsymbol{A}^T \boldsymbol{P_t A}\big)\boldsymbol{x_{t-1}} + \boldsymbol{u_{t-1}^T}\big(\boldsymbol{R} + \boldsymbol{B}^T \boldsymbol{P_t B}\big)\boldsymbol{u_{t-1}} + 2\boldsymbol{u_{t-1}^T B}^T \boldsymbol{P_t A x_{t-1}}\right\}.$$

Because this objective function is quadratic in $\boldsymbol{u_{t-1}}$, our optimization problem is convex, so the optimal solution, $\boldsymbol{\hat{u}_{t-1}}$, satisfies the following equality:

$$\nabla_{\boldsymbol{u_{t-1}}} J^*_{t-1}(\boldsymbol{x_{t-1}})|_{\boldsymbol{u_{t-1}}=\boldsymbol{\hat{u}_{t-1}}} = 0.$$

Taking the gradient of the minimum cost-to-go with respect to $\boldsymbol{u_{t-1}}$, we get

$$\nabla_{\boldsymbol{u_{t-1}}} J^*_{t-1}(\boldsymbol{x_{t-1}}) = 2\big(\boldsymbol{R} + \boldsymbol{B}^T\boldsymbol{P_t}\boldsymbol{B}\big)\boldsymbol{u_{t-1}} + 2\boldsymbol{B}^T\boldsymbol{P_t}\boldsymbol{A}\boldsymbol{x_{t-1}}.$$

Plugging in $\boldsymbol{\hat{u}_{t-1}}$ for $\boldsymbol{u_{t-1}}$ and setting this equation equal to zero, we find

$$\boldsymbol{\hat{u}_{t-1}} = -\big(\boldsymbol{R} + \boldsymbol{B}^T\boldsymbol{P_t}\boldsymbol{B}\big)^{-1}\boldsymbol{B}^T\boldsymbol{P_t}\boldsymbol{A}\boldsymbol{x_{t-1}} = -\boldsymbol{K_{t-1}}\boldsymbol{x_{t-1}}.$$

Now to find the minimum cost-to-go for $k = t-1$, we can substitute this optimal cost into our previous expression for $J^*_{t-1}(\boldsymbol{x_{t-1}})$, which gives us

$$\begin{aligned}
J^*_{t-1}(\boldsymbol{x_{t-1}}) &= \boldsymbol{x_{t-1}^T}\boldsymbol{Q}\boldsymbol{x_{t-1}} + \boldsymbol{\hat{u}_{t-1}^T}\boldsymbol{R}\boldsymbol{\hat{u}_{t-1}} + (\boldsymbol{A}\boldsymbol{x_{t-1}} + \boldsymbol{B}\boldsymbol{\hat{u}_{t-1}})^T\boldsymbol{P_t}(\boldsymbol{A}\boldsymbol{x_{t-1}} + \boldsymbol{B}\boldsymbol{\hat{u}_{t-1}}) \\
&= \boldsymbol{x_{t-1}^T}\boldsymbol{Q}\boldsymbol{x_{t-1}} + \boldsymbol{x_{t-1}^T}\boldsymbol{K_{t-1}^T}\boldsymbol{R}\boldsymbol{K_{t-1}}\boldsymbol{x_{t-1}} \\
&\qquad + \big(\boldsymbol{A}\boldsymbol{x_{t-1}} - \boldsymbol{B}\boldsymbol{K_{t-1}}\boldsymbol{x_{t-1}}\big)^T\boldsymbol{P_t}\big(\boldsymbol{A}\boldsymbol{x_{t-1}} - \boldsymbol{B}\boldsymbol{K_{t-1}}\boldsymbol{x_{t-1}}\big) \\
&= \boldsymbol{x_{t-1}^T}\boldsymbol{Q}\boldsymbol{x_{t-1}} + \boldsymbol{x_{t-1}^T}\boldsymbol{K_{t-1}^T}\boldsymbol{R}\boldsymbol{K_{t-1}}\boldsymbol{x_{t-1}} \\
&\qquad + \boldsymbol{x_{t-1}^T}\big(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{K_{t-1}}\big)^T\boldsymbol{P_t}\big(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{K_{t-1}}\big)\boldsymbol{x_{t-1}} \\
&= \boldsymbol{x_{t-1}^T}\Big(\boldsymbol{Q} + \boldsymbol{K_{t-1}^T}\boldsymbol{R}\boldsymbol{K_{t-1}} + \big(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{K_{t-1}}\big)^T\boldsymbol{P_t}\big(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{K_{t-1}}\big)\Big)\boldsymbol{x_{t-1}} \\
&= \boldsymbol{x_{t-1}^T}\boldsymbol{P_{t-1}}\boldsymbol{x_{t-1}}
\end{aligned}$$

Therefore, we have proven by induction that the expressions for $J^*_k(\boldsymbol{x_k})$ and $\boldsymbol{\hat{u}_k}$ given previously without proof are truly the minimum cost-to-go and optimal control. Notice that the optimal control is a linear function of the state, which is called linear state feedback. This is a rather nice and simple solution.

11.2.2 Variation 1: Output Cost

Now we will discuss some variations of the "vanilla" LQR problem. We will start with a simple variation, in which we aim to minimize the output of the system in place of the state. Consider a discrete LTI system described by

$$\begin{cases}
\boldsymbol{x_{k+1}} = \boldsymbol{A}\boldsymbol{x_k} + \boldsymbol{B}\boldsymbol{u_k} \\
\boldsymbol{y_k} = \boldsymbol{C}\boldsymbol{x_k} + \boldsymbol{D}\boldsymbol{u_k} \\
\boldsymbol{x_0} = x_{init}
\end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{y_k} \in \mathbb{R}^{n_o}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o\times n}$, $\boldsymbol{D} \in \mathbb{R}^{n_o\times n_i}$, and $k \in [0, N]$. Previously, we wanted to find the control sequence $U = \{\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}\}$ that minimized the quadratic cost function

$$J(U, \boldsymbol{x_0}) = \sum_{\tau=0}^{N-1}\Big(\boldsymbol{x_\tau^T}\boldsymbol{Q}\boldsymbol{x_\tau} + \boldsymbol{u_\tau^T}\boldsymbol{R}\boldsymbol{u_\tau}\Big) + \boldsymbol{x_N^T}\boldsymbol{Q_f}\boldsymbol{x_N}.$$

Now let's suppose that we to minimize the quadratic cost function

$$J(U, \boldsymbol{x_0}) = \sum_{\tau=0}^{N-1} \left(\boldsymbol{y_\tau^T Q y_\tau} + \boldsymbol{u_\tau^T R u_\tau} \right),$$

where $\boldsymbol{Q} \in \mathbb{R}^{n_o \times n_o}$ is a positive semidefinite matrix and $\boldsymbol{R} \in \mathbb{R}^{n_i \times n_o}$ is positive definite. In this modified cost function, $\boldsymbol{y_\tau^T Q y_\tau}$ is the cost for deviating from the desired output (zero) at time τ and $\boldsymbol{u_\tau^T R u_\tau}$ is the cost for applying control at time τ. This formulation of the cost function gives us nearly the same minimum cost-to-go and optimal control that we found previously:

$$J_k^*(\boldsymbol{x_k}) = \boldsymbol{x_k^T P_k x_k} \quad \text{and} \quad \boldsymbol{\hat{u}_k} = -\boldsymbol{K_k x_k},$$

where the matrices $\boldsymbol{P_k} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{K_k} \in \mathbb{R}^{n_i \times n}$ are now defined such that

$$\boldsymbol{P_k} = \begin{cases} 0 & \text{for } k = N \\ \boldsymbol{C^T Q C} + \boldsymbol{K_k^T R K_k} + (\boldsymbol{A} - \boldsymbol{B K_k})^T \boldsymbol{P_{k+1}} (\boldsymbol{A} - \boldsymbol{B K_k}) & \text{for } k = 0, 1, \cdots, N-1 \end{cases}$$

$$\boldsymbol{K_k} = (\boldsymbol{R} + \boldsymbol{B^T P_{k+1} B})^{-1} \boldsymbol{B^T P_{k+1} A} \text{ for } k = 0, 1, \cdots, N-1$$

We can prove that this is the minimum cost-to-go and optimal cost in the same way as we did for the previous formulation of the cost function.

11.2.3 Variation 2: Affine System

Consider a discrete affine time-invariant system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A x_k} + \boldsymbol{B u_k} + \boldsymbol{c} \\ \boldsymbol{x_0} = x_{init} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $c \in \mathbb{R}^n$, and $k \in [0, N]$. Suppose we want to find the control sequence $U = \{\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}\}$ that minimizes

$$J(U, \boldsymbol{x_0}) = \sum_{\tau=0}^{N-1} \left(\boldsymbol{x_\tau^T Q x_\tau} + \boldsymbol{u_\tau^T R u_\tau} \right) + \boldsymbol{x_N^T Q_f x_N}.$$

To derive the optimal control policy, we can define a new state vector:

$$\boldsymbol{z_k} := \begin{bmatrix} \boldsymbol{x_k} \\ 1 \end{bmatrix}.$$

This then allows us to express the affine system as a linear system:

$$\boldsymbol{z_{k+1}} = \boldsymbol{\tilde{A} z_k} + \boldsymbol{\tilde{B} u_k}, \text{ where}$$

$$\boldsymbol{\tilde{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{c} \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\tilde{B}} = \begin{bmatrix} \boldsymbol{B} \\ 0 \end{bmatrix}.$$

To express the cost function in terms of this new variable, we need to pad the cost matrices $\boldsymbol{Q}$ and $\boldsymbol{Q_f}$ with zeros to make them the appropriate size:

$$\tilde{\boldsymbol{Q}} := \begin{bmatrix} \boldsymbol{Q} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{\boldsymbol{Q}}_{\boldsymbol{f}} := \begin{bmatrix} \boldsymbol{Q_f} & 0 \\ 0 & 0 \end{bmatrix}.$$

Now our cost function can be expressed in terms of this new state as

$$J(U, \boldsymbol{z_0}) = \sum_{\tau=0}^{N-1} \left(\boldsymbol{z_\tau}^T \tilde{\boldsymbol{Q}} \boldsymbol{z_\tau} + \boldsymbol{u_\tau}^T \boldsymbol{R} \boldsymbol{u_\tau} \right) + \boldsymbol{z_N}^T \tilde{\boldsymbol{Q}}_{\boldsymbol{f}} \boldsymbol{z_N}.$$

This formulation of the cost function gives us nearly the same minimum cost-to-go and optimal control as we found for the original linear system:

$$J_k^*(\boldsymbol{z_k}) = \boldsymbol{z_k}^T \boldsymbol{P_k} \boldsymbol{z_k} \quad \text{and} \quad \hat{\boldsymbol{u}}_{\boldsymbol{k}} = -\boldsymbol{K_k} \boldsymbol{z_k},$$

where the matrices $\boldsymbol{P_k} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{K_k} \in \mathbb{R}^{n_i \times n}$ are now defined such that

$$\boldsymbol{P_k} = \begin{cases} \tilde{\boldsymbol{Q}}_{\boldsymbol{f}} & \text{for } k = N \\ \tilde{\boldsymbol{Q}} + \boldsymbol{K_k}^T \boldsymbol{R} \boldsymbol{K_k} + (\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{B}} \boldsymbol{K_k})^T \boldsymbol{P_{k+1}} (\tilde{\boldsymbol{A}} - \tilde{\boldsymbol{B}} \boldsymbol{K_k}) & \text{for } k = 0, 1, \cdots, N-1 \end{cases}$$

$$\boldsymbol{K_k} = (\boldsymbol{R} + \tilde{\boldsymbol{B}}^T \boldsymbol{P_{k+1}} \tilde{\boldsymbol{B}})^{-1} \tilde{\boldsymbol{B}}^T \boldsymbol{P_{k+1}} \tilde{\boldsymbol{A}} \text{ for } k = 0, 1, \cdots, N-1$$

We can prove that this is the minimum cost-to-go and optimal cost in the same way as we did for the original formulation of the cost function.

11.2.4 Variation 3: Reference Trajectory

Again, consider a discrete linear time-invariant (LTI) system described by

$$\begin{cases} \boldsymbol{x_{k+1}} = \boldsymbol{A} \boldsymbol{x_k} + \boldsymbol{B} \boldsymbol{u_k} \\ \boldsymbol{x_0} = x_{init} \end{cases},$$

where $\boldsymbol{x_k} \in \mathbb{R}^n$, $\boldsymbol{u_k} \in \mathbb{R}^{n_i}$, $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, and $k \in [0, N]$. Now suppose we want our system to follow a reference trajectory $(\boldsymbol{x_i}^{ref}, \boldsymbol{u_i}^{ref})$ for $i = 0, 1, \cdots, N$ as closely as possible. We can derive a control policy that minimizes the deviation of the system trajectory from the reference trajectory by finding the control sequence $U = \{\boldsymbol{u_0}, \boldsymbol{u_1}, \cdots, \boldsymbol{u_{N-1}}\}$ that minimizes the quadratic cost function

$$J(U, \boldsymbol{x_0}) = \sum_{\tau=0}^{N-1} \Big((\boldsymbol{x_\tau} - \boldsymbol{x_\tau}^{ref})^T \boldsymbol{Q} (\boldsymbol{x_\tau} - \boldsymbol{x_\tau}^{ref}) + (\boldsymbol{u_\tau} - \boldsymbol{u_\tau}^{ref})^T \boldsymbol{R} (\boldsymbol{u_\tau} - \boldsymbol{u_\tau}^{ref}) \Big) + (\boldsymbol{x_N} - \boldsymbol{x_N}^{ref})^T \boldsymbol{Q}_f (\boldsymbol{x_N} - \boldsymbol{x_N}^{ref}).$$

Case 1: Reference Trajectory Obeys Dynamics

Let's first consider the case where the reference trajectory obeys the dynamics of our system (i.e. $\boldsymbol{x_{k+1}^{ref}} = \boldsymbol{A x_k^{ref}} + \boldsymbol{B u_k^{ref}}$). We can define the new state and input variables $\boldsymbol{z_k} = \boldsymbol{x_k} - \boldsymbol{x_k^{ref}}$ and $\boldsymbol{v_k} = \boldsymbol{u_k} - \boldsymbol{u_k^{ref}}$. Now the dynamics of this new system can be described by the following equation:

$$\begin{aligned} \boldsymbol{z_{k+1}} &= \boldsymbol{x_{k+1}} - \boldsymbol{x_{k+1}^{ref}} = \left(\boldsymbol{A x_k} + \boldsymbol{B u_k}\right) - \left(\boldsymbol{A x_k^{ref}} + \boldsymbol{B u_k^{ref}}\right) \\ &= \boldsymbol{A}\left(\boldsymbol{x_k} - \boldsymbol{x_k^{ref}}\right) + \boldsymbol{B}\left(\boldsymbol{u_k} - \boldsymbol{u_k^{ref}}\right) = \boldsymbol{A z_k} + \boldsymbol{B v_k} \end{aligned}$$

Our cost function can now be expressed in terms of these new variables as

$$J(U, z_0) = \sum_{\tau=0}^{N-1} \left(\boldsymbol{z_\tau^T Q z_\tau} + \boldsymbol{v_\tau^T R v_\tau}\right) + \boldsymbol{z_N^T Q_f z_N}.$$

This formulation of the cost function gives us nearly the same minimum cost-to-go and optimal control as we found previously:

$$J_k^*(\boldsymbol{z_k}) = \boldsymbol{z_k^T P_k z_k} \qquad \text{and} \qquad \boldsymbol{\hat{v}_k} = -\boldsymbol{K_k z_k},$$

where the matrices $\boldsymbol{P_k} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{K_k} \in \mathbb{R}^{n_i\times n}$ are again defined such that

$$\boldsymbol{P_k} = \begin{cases} \boldsymbol{Q_f} & \text{for } k = N \\ \boldsymbol{Q} + \boldsymbol{K_k^T R K_k} + (\boldsymbol{A} - \boldsymbol{B K_k})^T \boldsymbol{P_{k+1}} (\boldsymbol{A} - \boldsymbol{B K_k}) & \text{for } k = 0, 1, \cdots, N-1 \end{cases}$$

$$\boldsymbol{K_k} = (\boldsymbol{R} + \boldsymbol{B}^T \boldsymbol{P_{k+1} B})^{-1} \boldsymbol{B}^T \boldsymbol{P_{k+1} A} \text{ for } k = 0, 1, \cdots, N-1$$

We can prove that this is the minimum cost-to-go and optimal cost in the same way as we did for the original formulation of the cost function. In terms of our original variables, the optimal control is given by

$$\boldsymbol{\hat{u}_k} = -\boldsymbol{K_k}(\boldsymbol{x_k} - \boldsymbol{x_k^{ref}}) + \boldsymbol{u_k^{ref}}.$$

Case 2: Reference Trajectory oes Not Obey Dynamics

Now let's consider the case where the reference trajectory does not obey the dynamics of our system (i.e. $\boldsymbol{x_{k+1}^{ref}} \neq \boldsymbol{A x_k^{ref}} + \boldsymbol{B u_k^{ref}}$). In this case, the reference trajectory must obey some time-varying affine model with $\boldsymbol{c_k} \in \mathbb{R}^n$:

$$\boldsymbol{x_{k+1}^{ref}} = \boldsymbol{A x_k^{ref}} + \boldsymbol{B u_k^{ref}} + \boldsymbol{c_k}.$$

As we did previously, we can define new state and input variables $\boldsymbol{z_k} = \boldsymbol{x_k} - \boldsymbol{x_k^{ref}}$ and $\boldsymbol{v_k} = \boldsymbol{u_k} - \boldsymbol{u_k^{ref}}$. The dynamics of this system can be described by

$$\begin{aligned} \boldsymbol{z_{k+1}} &= \boldsymbol{x_{k+1}} - \boldsymbol{x_{k+1}^{ref}} = \left(\boldsymbol{A x_k} + \boldsymbol{B u_k}\right) - \left(\boldsymbol{A x_k^{ref}} + \boldsymbol{B u_k^{ref}} + \boldsymbol{c_k}\right) \\ &= \boldsymbol{A}\left(\boldsymbol{x_k} - \boldsymbol{x_k^{ref}}\right) + \boldsymbol{B}\left(\boldsymbol{u_k} - \boldsymbol{u_k^{ref}}\right) - \boldsymbol{c_k} = \boldsymbol{A z_k} + \boldsymbol{B v_k} - \boldsymbol{c_k} \end{aligned}$$

Notice that this system is described by an affine time-varying model. To derive the optimal control policy, we can define another new state vector

$$\boldsymbol{w_k} := \begin{bmatrix} \boldsymbol{z_k} \\ 1 \end{bmatrix}.$$

This then allows us to express the affine time-varying system as a linear one:

$$\boldsymbol{w_{k+1}} = \boldsymbol{\tilde{A}_k w_k} + \boldsymbol{\tilde{B} v_k}, \text{ where}$$

$$\boldsymbol{\tilde{A}_k} = \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{c_k} \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\tilde{B}} = \begin{bmatrix} \boldsymbol{B} \\ 0 \end{bmatrix}.$$

To express the cost function in terms of this new variable, we need to pad the cost matrices $\boldsymbol{Q}$ and $\boldsymbol{Q_f}$ with zeros to make them the appropriate size:

$$\boldsymbol{\tilde{Q}} := \begin{bmatrix} \boldsymbol{Q} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\tilde{Q}_f} := \begin{bmatrix} \boldsymbol{Q_f} & 0 \\ 0 & 0 \end{bmatrix}.$$

Now our cost function can be expressed in terms of this new state as

$$J(U, \boldsymbol{z_0}) = \sum_{\tau=0}^{N-1} \left(\boldsymbol{w_\tau^T \tilde{Q} w_\tau} + \boldsymbol{v_\tau^T R v_\tau} \right) + \boldsymbol{w_N^T \tilde{Q}_f w_N}.$$

This formulation of the cost function gives us nearly the same minimum cost-to-go and optimal control as we found previously:

$$J_k^*(\boldsymbol{w_k}) = \boldsymbol{w_k^T P_k w_k} \quad \text{and} \quad \boldsymbol{\hat{v}_k} = -\boldsymbol{K_k w_k},$$

where the matrices $\boldsymbol{P_k} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{K_k} \in \mathbb{R}^{n_i \times n}$ are now defined such that

$$\boldsymbol{P_k} = \begin{cases} \boldsymbol{\tilde{Q}_f} & \text{for } k = N \\ \boldsymbol{\tilde{Q}} + \boldsymbol{K_k^T R K_k} + (\boldsymbol{\tilde{A}_k} - \boldsymbol{\tilde{B} K_k})^T \boldsymbol{P_{k+1}} (\boldsymbol{\tilde{A}_k} - \boldsymbol{\tilde{B} K_k}) & \text{for } k = 0, 1, \cdots, N-1 \end{cases}$$

$$\boldsymbol{K_k} = (\boldsymbol{R} + \boldsymbol{\tilde{B}}^T \boldsymbol{P_{k+1} \tilde{B}})^{-1} \boldsymbol{\tilde{B}}^T \boldsymbol{P_{k+1} \tilde{A}_k} \text{ for } k = 0, 1, \cdots, N-1$$

We can prove that this is the minimum cost-to-go and optimal cost in the same way as we did for the original formulation of the cost function. In terms of our original variables, the optimal control is given by

$$\boldsymbol{\hat{u}_k} = -\boldsymbol{K_k} \begin{bmatrix} \boldsymbol{x_k} - \boldsymbol{x_k^{ref}} \\ 1 \end{bmatrix} + \boldsymbol{u_k^{ref}}.$$

11.3 Continuous Time LQR

LQR is also applicable to continuous LTI systems. We will now discuss how to set up and solve the LQR optimization problem for continuous systems.

11.3.1 Vanilla LQR Problem

Again, I will begin with the "vanilla" LQR problem and then discuss variations.

LQR Optimization Problem

Consider a continuous linear time-invariant (LTI) system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ x(0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$. Suppose we want to find the control function, u, that minimizes the quadratic cost function

$$J(u) = \int_0^{\infty} \Big(x(t)^T \boldsymbol{Q} x(t) + u(t)^T \boldsymbol{R} u(t) \Big) dt,$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is a positive semidefinite matrix and $\boldsymbol{R} \in \mathbb{R}^{n_i\times n_i}$ is positive definite. In this cost function, $x(t)^T \boldsymbol{Q} x(t)$ is the cost for deviating from the desired state (zero) at time t and $u(t)^T \boldsymbol{R} u(t)$ is the cost for applying control at time t. To find the control that minimizes this cost function, while obeying our dynamics, we set up the following optimization problem:

$$\begin{aligned} \hat{u} = \arg\min_u \quad & J(u) \\ \text{s.t.} \quad & \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ & x(0) = \boldsymbol{x_0} \end{aligned}$$

Optimal Control

For this optimization problem, we claim the optimal control is given by

$$\hat{u}(t) = -\boldsymbol{R}^{-1} \boldsymbol{B}^T \boldsymbol{P} x(t),$$

where $\boldsymbol{P}$ is the positive definite solution to the Algeabraic Ricatti Equation:

$$\boldsymbol{PA} + \boldsymbol{A}^T \boldsymbol{P} - \boldsymbol{PBR}^{-1} \boldsymbol{B}^T \boldsymbol{P} + \boldsymbol{Q} = 0.$$

Proof of Optimality

To prove this is actually the optimal control, we will first use the fundamental theorem of calculus to write the following integral derivative:

$$\frac{d}{dt} \int_0^{\infty} x(t)^T \boldsymbol{P} x(t) dt = \lim_{t\to\infty} x(t)^T \boldsymbol{P} x(t) - x(0)^T \boldsymbol{P} x(0).$$

Additionally, using the Leibniz rule, we can write this integral as

$$\frac{d}{dt} \int_0^{\infty} x(t)^T \boldsymbol{P} x(t) dt = \int_0^{\infty} \frac{d}{dt} x(t)^T \boldsymbol{P} x(t) dt.$$

Using the product rule to compute the inner derivative, we now have

$$\frac{d}{dt}\int_0^\infty x(t)^T \boldsymbol{P} x(t) dt = \int_0^\infty \dot{x}(t)^T \boldsymbol{P} x(t) + x(t)^T \boldsymbol{P} \dot{x}(t) dt.$$

Given the condition on the dynamics of the system, this expression becomes

$$\frac{d}{dt}\int_0^\infty x(t)^T \boldsymbol{P} x(t) dt = \int_0^\infty \big(\boldsymbol{A}x(t)+\boldsymbol{B}u(t)\big)^T \boldsymbol{P} x(t) + x(t)^T \boldsymbol{P}\big(\boldsymbol{A}x(t)+\boldsymbol{B}u(t)\big) dt.$$

Expanding the expression inside the integral, we have

$$\begin{aligned}\frac{d}{dt}\int_0^\infty x(t)^T \boldsymbol{P} x(t) dt &= \int_0^\infty x(t)^T \boldsymbol{A}^T \boldsymbol{P} x(t) + 2u(t)^T \boldsymbol{B}^T \boldsymbol{P} x(t) + x(t)^T \boldsymbol{P}\boldsymbol{A} x(t) dt \\ &= \int_0^\infty x(t)^T \big(\boldsymbol{A}^T \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\big) x(t) + 2u(t)^T \boldsymbol{B}^T \boldsymbol{P} x(t) dt.\end{aligned}$$

We can now take our previous expression for the cost function and essentially add zero to it by subtracting and adding two equivalent expressions:

$$\begin{aligned}J(u) = \int_0^\infty \Big(x(t)^T \boldsymbol{Q} x(t) + u(t)^T \boldsymbol{R} u(t)\Big) dt - \lim_{t\to\infty} x(t)^T \boldsymbol{P} x(t) + x(0)^T \boldsymbol{P} x(0) \\ \int_0^\infty x(t)^T \big(\boldsymbol{A}^T \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\big) x(t) + 2u(t)^T \boldsymbol{B}^T \boldsymbol{P} x(t) dt.\end{aligned}$$

We can write this expression for the cost a bit more neatly:

$$\begin{aligned}J(u) = \int_0^\infty x(t)^T \big(\boldsymbol{Q} + \boldsymbol{A}^T \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}\big) x(t) + 2u(t)^T \boldsymbol{B}^T \boldsymbol{P} x(t) + u(t)^T \boldsymbol{R} u(t) dt \\ + x(0)^T \boldsymbol{P} x(0) - \lim_{t\to\infty} x(t)^T \boldsymbol{P} x(t).\end{aligned}$$

Assuming $\boldsymbol{P}$ is the positive definite solution to the Algeabraic Ricatti Equation, we can substitute $\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}$ for $\boldsymbol{Q} + \boldsymbol{A}^T \boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}$ in our previous expression:

$$\begin{aligned}J(u) = \int_0^\infty x(t)^T \big(\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\big) x(t) + 2u(t)^T \boldsymbol{B}^T \boldsymbol{P} x(t) + u(t)^T \boldsymbol{R} u(t) dt \\ + x(0)^T \boldsymbol{P} x(0) - \lim_{t\to\infty} x(t)^T \boldsymbol{P} x(t).\end{aligned}$$

Rearranging the expression inside the integral, we can write the cost as

$$\begin{aligned}J(u) = \int_0^\infty \big(u(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} x(t)\big)^T \boldsymbol{R}\big(u(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} x(t)\big) dt \\ + x(0)^T \boldsymbol{P} x(0) - \lim_{t\to\infty} x(t)^T \boldsymbol{P} x(t).\end{aligned}$$

If the system is stable under the optimal control, then the state approaches the zero vector as time goes to infinity. Under this condition, the cost is simply

$$J(u) = x(0)^T \boldsymbol{P} x(0) + \int_0^\infty \big(u(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} x(t)\big)^T \boldsymbol{R}\big(u(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} x(t)\big) dt.$$

We previously assumed that the optimal control is given by $\hat{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}x(t)$. If we plug this into our previous expression for the cost, we find

$$\begin{aligned} J(\hat{u}) &= x(0)^T\boldsymbol{P}x(0) + \int_0^\infty \left(\hat{u}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}x(t)\right)^T \boldsymbol{R}\left(\hat{u}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}x(t)\right)dt \\ &= x(0)^T\boldsymbol{P}x(0) + \int_0^\infty \mathbf{0}_{n_i}^T\boldsymbol{R}\mathbf{0}_{n_i}\,dt \\ &= x(0)^T\boldsymbol{P}x(0) \end{aligned}$$

This is the cost incurred for the initial state, $x(0)$, without applying any control or considering any other states. Therefore, this must be the minimum cost. Now we have proven that the optimal control is in fact $\hat{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}x(t)$. Notice that this control is a linear function of the state, which is called linear state feedback. Again, LQR gives us a simple solution for the optimal control.

11.3.2 Variation 1: Output Cost

Now we will discuss some variations of the "vanilla" LQR problem. We will start with a simple variation, in which we aim to minimize the output of the system in place of the state. Consider a discrete LTI system described by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \\ x(0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, and $y(t) \in \mathbb{R}^{n_o}$. Assume that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n_i}$, $\boldsymbol{C} \in \mathbb{R}^{n_o \times n}$, and $\boldsymbol{D} \in \mathbb{R}^{n_o \times n_i}$. Now suppose we want to find the control function u that minimizes the quadratic cost function in terms of the output:

$$J(u) = \int_0^\infty \left(y(t)^T\boldsymbol{Q}y(t) + u^T(t)\boldsymbol{R}u(t)\right)dt,$$

where $\boldsymbol{Q} \in \mathbb{R}^{n_o \times n_o}$ is a positive semidefinite matrix and $\boldsymbol{R} \in \mathbb{R}^{n_i \times n_i}$ is positive definite. In this cost function, $y^T(t)\boldsymbol{Q}y(t)$ is the cost for deviating from the desired output (zero) at time t and $u^T(t)\boldsymbol{R}u(t)$ is the cost for applying control at time t. To find the control that minimizes this cost function, while obeying our dynamics, we set up the following optimization problem:

$$\begin{aligned} \hat{u} = \arg\min_u \quad & J(u) \\ \text{s.t.} \quad & \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ & y(t) = \boldsymbol{C}x(t) + \boldsymbol{D}u(t) \end{aligned}$$

The solution to this optimization problem is very similar to the one given for the cost function that penalized state deviations, instead of output deviations. For this optimization problem, the optimal control is again given by

$$\hat{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}x(t),$$

but now $\boldsymbol{P}$ is the positive definite solution to the Algeabraic Ricatti Equation:

$$\boldsymbol{PA} + \boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{PBR}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{C}^T\boldsymbol{QC} = 0.$$

We can prove that this is the optimal control in the same way as we did for the previous formulation of the cost function.

11.3.3 Variation 2: Reference Trajectory

Again, consider a continuous LTI system described by the differential equation

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{B}u(t) \\ x(0) = \boldsymbol{x_0} \end{cases},$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^{n_i}$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and $\boldsymbol{B} \in \mathbb{R}^{n\times n_i}$. Suppose we want our system to follow a reference trajectory $(x^{ref}(t), u^{ref}(t))$ for $t \geq 0$ as closely as possible. We can derive a control policy that minimizes the deviation of the system trajectory from the reference trajectory by finding the control function, u, that minimizes the following quadratic cost function:

$$J(u) = \int_0^\infty (x(t)-x^{ref}(t))^T\boldsymbol{Q}(x(t)-x^{ref}(t))+(u(t)-u^{ref}(t))^T\boldsymbol{R}(u(t)-u^{ref}(t))dt.$$

We can define the new state and input variables such that $z(t) = x(t) - x^{ref}(t)$ and $v(t) = u(t) - u^{ref}(t)$. Now our quadratic cost function can be expressed as

$$J(u) = \int_0^\infty \Big(z(t)^T\boldsymbol{Q}z(t) + \boldsymbol{v}(t)^T\boldsymbol{R}v(t)\Big)dt.$$

This formulation of the cost gives us nearly the same optimal control as before:

$$\hat{v}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}z(t),$$

where $\boldsymbol{P}$ is the positive definite solution to the Algeabraic Ricatti Equation:

$$\boldsymbol{PA} + \boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{PBR}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{Q} = 0.$$

We can prove that this is the optimal cost in the same way as we did for the original formulation of the cost function. In terms of our original system variables and reference signals, the optimal control is given by

$$\hat{u}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\Big(x(t) - x^{ref}(t)\Big) + u^{ref}(t).$$

11.4 Choosing Q and R Matrices

So far, we discussed how to solve the LQR optimization problem for both the standard discrete and continuous LTI cases, along with a few variations. We'll end by discussing how to choose the parameters of this optimization problem.

We typically choose the cost matrices, $\boldsymbol{Q}$ and $\boldsymbol{R}$, to be diagonal matrices, so that the ith diagonal element of $\boldsymbol{Q}$ represents the cost of deviating from the ith element of the state (or output), and the ith diagonal element of $\boldsymbol{R}$ represents the cost for using the ith element of the input. If the elements of $\boldsymbol{R}$ are larger than those of $\boldsymbol{Q}$, then we place a greater cost on using input than on state/output deviations, and control is said to be expensive. In this case, we will see a slower response and more overshoot before the system reaches the desired state. If the elements of $\boldsymbol{R}$ are smaller than $\boldsymbol{Q}$, then we place a greater cost on state/output deviations than on using input, and control is said to be cheap. In this case, we will see a faster response and less overshoot before the system settles.