

1-12-2023

Introduction to Control Engineering

Xiangyu Meng

Louisiana State University at Baton Rouge, xmeng5@lsu.edu

Follow this and additional works at: <https://digitalcommons.lsu.edu/etext>



Part of the [Controls and Control Theory Commons](#)

Recommended Citation

Meng, Xiangyu, "Introduction to Control Engineering" (2023). *E-Textbooks*. 1.
<https://digitalcommons.lsu.edu/etext/1>

This Book is brought to you for free and open access by LSU Digital Commons. It has been accepted for inclusion in E-Textbooks by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.

Introduction to Control Engineering

Xiangyu Meng

January 12, 2023



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License by Xiangyu Meng, except where otherwise noted.

© 2023 Xiangyu Meng
Division of Electrical and Computer Engineering
Louisiana State University
Baton Rouge, LA 70803

If you are an instructor reviewing, adopting, or adapting this textbook, please help us understand your use and share suggestions by contacting me at xmeng5@lsu.edu.

This project was supported with Course Transformation Program funding from LOUIS: The Louisiana Library Network, a program of the Louisiana Board of Regents:
<https://www.louislibraries.org/alearningla/projects/transformations>

Contents

1	Introduction	7
1.1	What is control?	7
1.2	Open loop and Closed-loop Control	10
1.3	Examples of Control Systems	11
1.3.1	Flush Toilet	11
1.3.2	Buck Converters	11
1.3.3	Unmanned Aerial Vehicles	13
1.3.4	Autonomous vehicles	14
1.4	Control System Design Process	16
1.5	Summary	16
I	Mathematics and Physics for Control Engineering	19
2	Mathematics	21
2.1	Trigonometric functions	21
2.2	Complex Variables	21
2.3	Derivative	22
2.4	Laplace Transforms	23
2.4.1	Properties of Laplace Transforms	23
2.4.2	Table of selected Laplace transforms	25
2.4.3	Using Laplace Transforms to Solve Differential Equations	25
2.5	Cramer's Rule	26
3	Physics	27
3.1	Newton's laws of motion	27
3.1.1	Inertial Reference Frame	27

3.1.2	Spring and Damper	27
3.1.3	Translational and Rotational Motions	28
3.2	Circuits	30
3.2.1	Electric circuits	30
3.2.2	Kirchhoff's circuit laws	31
3.2.3	Operational Amplifier	32
3.3	Mechatronics	36
3.3.1	Law of Motors	36
3.3.2	Law of Generator	36
II	Analysis and Design of Control Systems	39
4	Modeling of Dynamic Systems	41
4.1	Linear Time-Invariant Systems	42
4.1.1	Linearity	42
4.1.2	Time Invariance	42
4.1.3	Linear System Representations	42
4.2	Modeling the Process	43
4.2.1	Autonomous Vehicle Cruise Control	43
4.2.2	Quadcopter Drones	44
4.2.3	The 3D Printer	45
4.3	Modeling controllers	46
4.4	Modeling sensors and actuators	48
4.4.1	Sensors	48
4.4.2	Actuator	50
5	Analysis of Control Systems	53
5.1	Block Diagrams	53
5.1.1	Terminologies	54
5.1.2	Elementary Block Diagram	54
5.2	Transfer Functions	57
5.2.1	Impulse Response	57
5.2.2	General Time Response	60
5.2.3	Partial Fraction Expansion	60
5.2.4	Zero and Poles	61
5.3	Stability	62
5.3.1	Stability from Impulse Response	62

5.3.2	Stability from Input-Output Relationship	63
5.3.3	Routh Array	64
5.4	Dynamic Response of an LTI System	65
5.4.1	Test Signals	66
5.5	Steady State Performance	66
5.5.1	Frequency Response	66
5.5.2	Step Response	68
5.6	Transient Performance	69
5.6.1	First Order Systems	69
5.6.2	Second Order Systems	73
5.6.3	Higher Order Systems	79
6	Design of Control Systems	81
6.1	What is a good control design?	82
6.1.1	Stability	83
6.1.2	Tracking	84
6.1.3	Regulation	84
6.1.4	Robustness	84
6.2	Tracking Problems	85
6.2.1	Test Signals	85
6.2.2	System Types for Reference Input	86
6.2.3	System Types for Disturbance Input	88
6.3	PID Controller	89
6.3.1	P Controller	90
6.3.2	I Controller	91
6.3.3	D Controller	94
6.3.4	PID Controller	94
6.4	PID Controller Tuning	94
6.4.1	The Step Response Method	95
6.4.2	The Frequency Response Method	96
III	Computer Aid Tools	99
7	Matlab	101
7.1	Control System Toolbox	102
7.1.1	Test Signals	102
7.1.2	Model Interconnections	103

7.1.3	PID Tuner	103
7.2	Symbolic Math Toolbox	103
7.3	Simscape	105
7.3.1	Simscape Electrical	106
8	LabView	107
8.1	Getting Started Modules	107
8.2	MathScript	107
8.3	Mathematics	108
8.4	Control and Simulation	108
8.5	Examples:	109
8.5.1	System Modeling	109
8.5.2	Block Diagram Reduction	110
8.5.3	Dynamic Response	110
8.5.4	System Type	110
8.5.5	PID Design	112
Appendices		
A	Image Credits	115

Chapter 1

Introduction

The learning objectives of this chapter:

- Appreciate the role and importance of control systems in our daily lives.
- Understand the basic components of a control system.
- Understand the difference between the open-loop and closed-loop systems, and the role of feedback in a closed-loop control system.
- Gain a practical sense of real-life control problems.

1.1 What is control?

According to Collins English Dictionary, as a verb, to control a piece of equipment, process, or system means to make it work in the way that you want it to work. To start our control design journey, let us look at the concept of systems. The system defines the relationship between the inputs and the outputs, where the inputs are forcing signals and the outputs are the responses of the system. The system can be almost anything, such as your car, your AC, or your vacuum robot.

What physical variables can be controlled? They could be position, speed, acceleration, or torque in vehicle control problems. They could be temperature, pressure, liquid level, or liquid flow rate in a chemical reaction process control problem. The list goes on and on. Now you realize that control is ubiquitous in almost all automated processes, from car suspensions, to self-driving vehicles, from quadcopters to commercial airplanes, and from

home vacuum robots to advanced surgical robots. Control is referred to as a hidden technology in engineering because it is crucial to many devices and automated processes while being mainly out of sight.

The main principle underlying all control systems is feedback, which is a concept we use all the time in our daily lives. For instance, consider the steps involved in regulating water temperature before taking a shower. Water comes out when the knob is turned. There is a desired water temperature we would like to reach before stepping into the shower. We use our hand to test the temperature and compare it with the desired temperature. If the water temperature is too cold, we turn the knob towards the hot water side. Conversely, if it is too hot, we turn the knob towards the cold water side. This process continues until the water temperature is within range of the preferred shower temperature. When the control system nomenclature is used, the desired temperature can be referred to as “reference input”, “set point”, or “desired input”; while the actual water temperature is referred to as the “output”. Furthermore, our hand in this example serves as a “sensor” for our brain to interpret temperature readings, and based on the difference between the input and output, our brain is able to determine whether or not the water temperature needs to be adjusted. The brain is seen as a “controller” and our hand and the knob known as “actuator” execute the command sent by the brain. The shower system being controlled is called the “process”. The overall goal here is to let the output to be close to the reference input. Applying all these terms in a block diagram results in a generic control system diagram shown in Fig. 1.1.

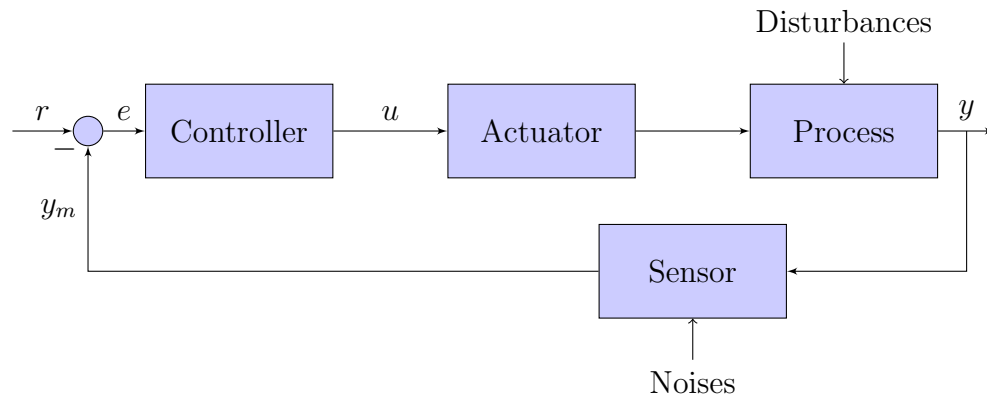


Figure 1.1: Control System Block Diagram

In a control system, there are four major signals

- Reference input (r): it is also called set-point.
- Output (y): it is the variable we want to control.
- Error (e): it is the difference between the reference input and the sensor output. Error signals play the central role in analysis of control systems.
- Control signal (u): it is the signal computed based on the error and applied to the actuator.

and four major components:

- Process: it is the central component, the one being controlled. The process is usually subject to disturbances, which tend to adversely affect the output of a system.
- Sensor: the role of a sensor is to measure the output of a process. The basic requirements of a sensor are accuracy, reliability, and sensitivity. The measurements are required to be accurate even though there are always sensor noises. It is desired that the sensor output has a linear relationship with the input and it is sensitive to the input change.
- Controller: the aim of the controller is to compute a control signal based on the error. In traditional control, the controller is a circuit. In modern digital control, the controller is usually a micro-computer.
- Actuator: an actuator is a device that performs the action to control the process directly. The actuator needs to be powerful, speedy and reliable. It should be able to influence the process output in a large range and in a short period of time.

When we use the feedback principle to perform certain tasks in daily life, it is called “manual” control. When we teach the machine to use the feedback principle to perform certain tasks, it is called “automatic” control. No matter what the control is, manual control or automatic control, the process can always be abstracted as the block-diagram shown in Fig. 1.1. We can use the room temperature control as an example. This is a case of automatic control in keeping the room temperature at a desired temperature for comfort. In this example, the process is the house. Thermostat serves as both the sensor

and the controller. The actuator is the valve, and the control signal is on and off of the heating and cooling systems as well as any ventilation fans that might be part of the system. The set point is the desired room temperature. Nowadays, most digital thermostats are programmable, and different setpoints can be set for different times of a day.

Further control system will incorporate the artificial intelligent, computer logic and sensor fusion into the feedback control to achieve intelligent control.

1.2 Open loop and Closed-loop Control

A control system can be open-loop or closed-loop. We provide an input to the plant and it gives us an output. We can then modify the input to try and get a desired output. This control strategy is called open-loop control. And as we have seen sometimes open-loop control does not work very well. It can be difficult to manually find the ideal input which gives the desired output. To resolve the issue, a commonly used control strategy is closed-loop control which is also called feedback control.

Even though the closed-loop control is the main stream for automatic control. Open-loop control has also been used in applications with less accuracy requirements. The open-loop control structure is shown in Fig. 1.2. In the open-loop control, the controller is calibrated based on many exper-

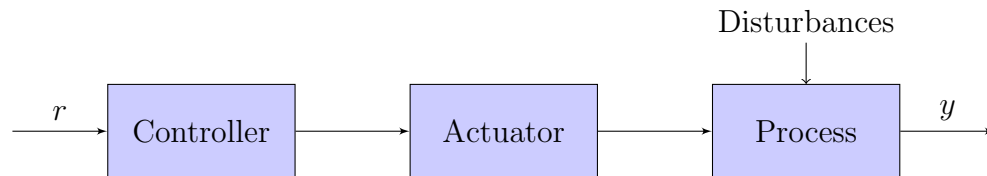


Figure 1.2: Open-loop Control System

iments to achieve a desired output for a given input signal. The open-loop control utilizes an actuating device to control the process directly without using feedback. Open-loop control is simple, inexpensive, and is used in the scenario when the output is hard to measure. However, open-loop control is sensitive to disturbances and is unable to reject these disturbances. Home appliances usually use open-loop control, such as toasters, washing machines, and microwaves. The toaster cannot adjust its toast time according to the thickness of the toast. Anyone with burnt toast can attest that toasters are

open-loop systems. In addition, there is no sensor to measure the color of the toast. There does not exist a sensor to measure the cleanliness of clothes.

Closed-loop control uses a measurement of the output and feedback of the output signal to compare with the desired reference input. The control signal is computed based on the error between the two. Closed-loop control is less sensitive to noises, disturbances, and changes in the environment. The disadvantages are that closed-loop control has a complex structure and is expensive when compared with open-loop control. Closed-loop control can be found in solar collectors, temperature control systems, speed control systems, etc.

1.3 Examples of Control Systems

Control systems have a long history, which can trace back to antiquity.

1.3.1 Flush Toilet

This mechanism of the liquid-level control was invented in ancient times, and it is still used today in the flush toilet. Figure 1.3 shows how a manual flush toilet operates. The “float ball” also known as the ballcock rises as the tank fills with water. The float is attached to a float rod. The opposite end of the rod is connected to an inlet valve. As the float ball rises, the rod attached to it presses against the inlet valve. When the tank is full, the rod presses against the inlet valve hard enough to shut water off. This stops the tank from overflowing. The flushing toilet was invented in 1596 and has made many improvements over the period of time.

1.3.2 Buck Converters

Power electronics are everywhere in the increasingly electrified world today. But not all power is created equal. For example, the power output from a wind farm is not immediately usable by your toaster. Power must first be converted from one form to another, such as between AC and DC. Power electronics devices in electric vehicles, or laptop computers, perform these conversions.

Power electronic devices convert electricity what you have at a source and what you need at a load. The conversions can go from DC to AC, AC

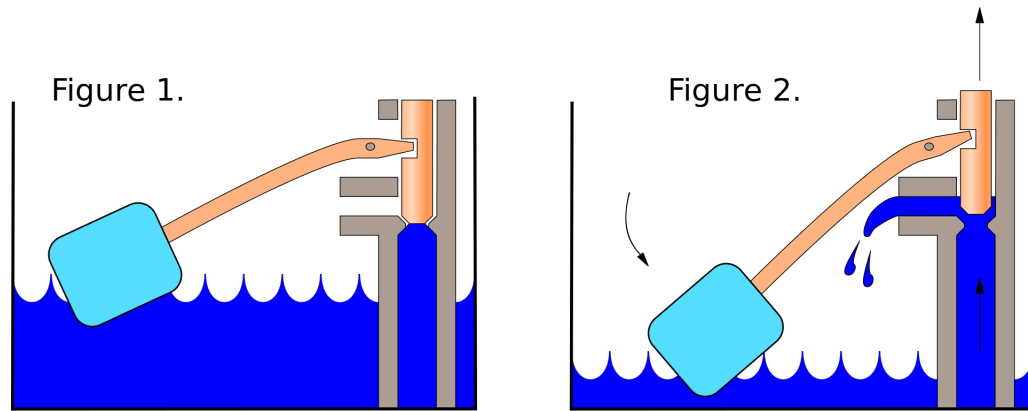


Figure 1.3: Flush toilet. Source: https://en.wikipedia.org/wiki/Flush_toilet

to DC, one DC voltage to another, and one AC magnitude and frequency to another. The key feature of all power electronic devices is that they are controlled by high-frequency switching. These switches are built using semiconductor devices, such as transistors and metal–oxide–semiconductor field-effect transistors (MOSFETs). Switching minimizes the power consumption of the device since power only flows when the switch is on.

For example, a buck converter is a DC to DC converter that decreases the voltage from the input source to the output load. Buck converters are common in personal electronics, charging devices, hybrid vehicles and more. Imagine, for instance, that you are using a solar charger to charge your mobile device. This solar panel has an open-circuit voltage of 15 volts. However, USB charging has a maximum voltage of 5 volts. It is important to buck the voltage to a lower value before connecting your device. Otherwise, it might just explode. Let us briefly review how a buck converter works. Again, the goal is to convert a high DC voltage to a lower DC voltage. Of course, a simple way to do this is to just use a voltage divider, but this is pretty inefficient. Instead, by rapidly opening and closing a switch, an average output voltage can be got that is lower than the input. This is a technique called pulse width modulation or PWM. For a PWM signal, the fraction of that the signal is high is called the duty cycle. When the duty cycle is constant, the average voltage across the switch is just duty cycle times the period. Now the average voltage is lower than the input voltage. But the maximum voltage is still higher than our load, the phone can tolerate. Plus, these fluctuations do not make a particular useful signal. What we

need is something to smooth the high frequency pulses into something that looks close to a constant DC value, something like a low pass filter. Adding an inductor and a capacitor to your circuit smooths out the high frequency switching leaving behind a signal that is nearly DC. The complete buck converter circuit also includes a diode, so current can continue to flow where the switch is open. The relationship between input and output voltage, switching frequency, and components sizes all factors into the design of the buck converter.

In practice, all power electronic components must be regulated using feedback control. Feedback controller allows the converter to appropriately handle fluctuations in the input or load. Standard techniques, such as PID or lead-lag compensation, can be used to regulate a buck converter. A PID controller can be implemented to maintain a specified load voltage, although the same modelling techniques apply to current controllers. The controller automatically adjusts the duty cycle of the switch based on the error between the desired and measured voltage.

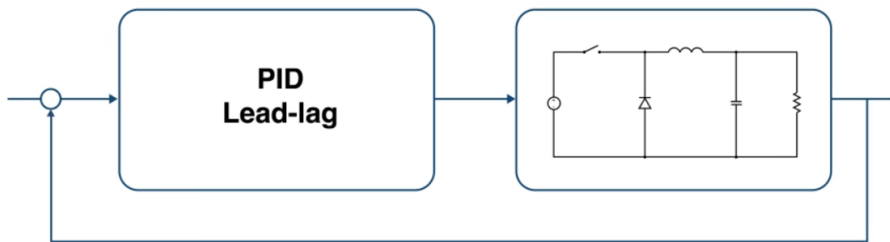


Figure 1.4: Closed-loop control of a buck converter

1.3.3 Unmanned Aerial Vehicles

Unmanned aerial vehicles became popular in the 2010s. Specifically, quadcopters have been used both commercially and non-commercially with applications in real estate, tourism, exploration, and performing arts. Control systems are used in quadcopters to achieve various motions, such as tilting to a certain roll angle. Here the reference input is the desired roll angle. The output is the actual roll angle. We use a sensor to measure the actual roll angle. An electronic control system is used to generate a control signal based on the error. Actuator in this case is the power systems and rotors. The process is the quadcopter.



Figure 1.5: Quadcopter

1.3.4 Autonomous vehicles

Control systems are expected to play a major role in self-driving cars, also known as autonomous vehicles. More vehicles running on the road are incorporating some degree of automation with technologies, such as lane keeping, adaptive cruise control, etc.

Lane Keeping

Lane keeping provides automatic steering and braking to keep a vehicle in its travel lane. This feature relies on painted lane markings to operate. These include the markings between lanes and along the edges of the roads. This feature may help the driver from driving off the road and avoid collisions. Integrated video cameras track the road markings and detect any change in the distance to either side of the vehicle. Whether cameras, sensors or radars are used, all of these technologies can make our roads safer. After

all, according to the German Federal Statistical Office, one third of accidents occurring outside built-up areas are the result of vehicles “veering off the road”. Lane Keeping Assist can help eliminate this type of accident.

Adaptive Cruise Control

You might have used cruise control on your car while driving on the highway. You change the speed by adjusting the amount of fuel going to the engine. Usually, you do this by stepping on the pedal. But if you are using cruise control, this is done by the computer automatically. Here, the car speed is the output, the amount of fuel going to the engine is the input, and everything in between is the plant. The goal of cruise control is to keep the vehicle speed at a set point, for example, 60 miles per hour. The desired speed is called the reference. Here, the microcomputer is the controller. Adaptive cruise control is an advanced cruise control technique for road vehicles that automatically adjust the vehicle speed to maintain a safe distance from vehicles ahead (see Fig. 1.6). Control is based on sensor information from on-board sensors.

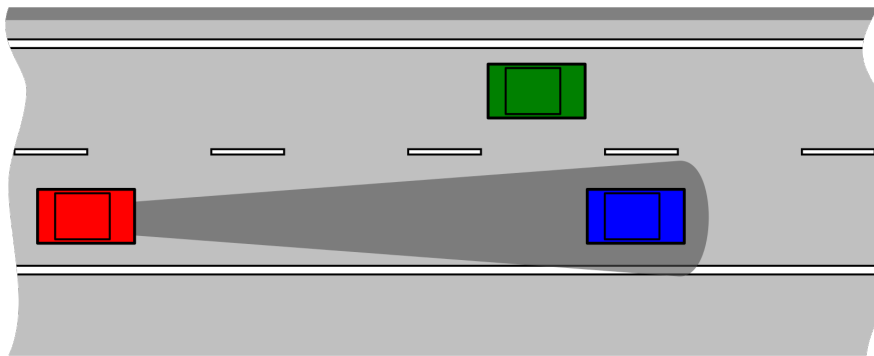


Figure 1.6: Adaptive Cruise Control. Source: https://en.wikipedia.org/wiki/Adaptive_cruise_control

Such systems may use a radar or laser sensor or a camera setup allowing the vehicle to brake when it detects the car is approaching another vehicle

ahead, then accelerate when traffic allows it to. When combined with another driver assist feature such as lane centering, the vehicle is considered a Level 2 autonomous car.

1.4 Control System Design Process

To provide control criteria, or specification: disturbance rejection, steady-state tracking error, transient response characteristics, sensitivity to parameter changes in the plant.

The general steps to solve a real control problem:

- Choose hardware (sensors, and actuators).
- Develop mathematical models for all components (plant, sensor, actuator, etc.).
- Design controller based on the models.
- Simulate the designed controller using computer aid software. If the performance is unsatisfactory, identify the problem and re-design the control system.
- Test the control system on real plants. If the physical testing is unsatisfactory, redo all steps.

Control System Design is generally a trial-and-error process in which the above steps are used iteratively to determine the design parameters of an “acceptable” system. A typical design involves juggling with conflicting design objectives until an acceptable controller is found. Acceptable performance is generally defined in terms of time and frequency domain criteria such as rise time, settling time, peak overshoot, gain and phase margin, and bandwidth.

1.5 Summary

Control is the process of making a system work with rules determined by a control engineer.

There are typically four signals in a control system

- Reference input (r): it is also called set-point.

- Output (y): it is the variable we want to control.
- Error (e): it is the difference between the reference input and the sensor output.
- Control signal (u): it is the signal computed based on the error and applied to the actuator.

The main components of a controls system are **Process**, **Sensor**, **Controller**, and **Actuator**:

- Process: it is the central component, the one being controlled.
- Sensor: the role of the sensor is to measure the output of the process.
- Controller: it is used to compute a control signal based on the error.
- Actuator: it is the one that acts to control the process directly.

There are two types of control systems: **open-loop control** and **closed-loop control**.

- The open-loop control utilizes an actuating device to control the process directly without using feedback. Open-loop control is simple and inexpensive.
- Closed-loop control uses a measurement of the output and feedback of the output signal to compare with the desired reference input.

Problems

1. Explain the differences and similarities between closed-loop control and open-loop control.
2. Draw a component block diagram for each of the following feedback control systems:
 - (a) Aircraft landing systems
 - (b) A manual control system for regulating the level of fluid in a tank by adjusting the output valve.

- (c) Chamber temperature-control system
 - (d) Student learning process
 - (e) Steering control of an automobile
 - (f) Idle-speed control of an automobile
3. List five electronic devices in your home that are closed-loop control systems.
 4. List five electronic devices in your home that are open-loop control systems.

Part I

Mathematics and Physics for Control Engineering

Chapter 2

Mathematics

2.1 Trigonometric functions

The linear combination of sine and cosine functions is equivalent to a single sine or cosine function with a phase shift and scaled amplitude,

$$a \cos \theta + b \sin \theta = c \cos(\theta - \varphi_1) = c \sin(\theta + \varphi_2) \quad (2.1)$$

where c , φ_1 and φ_2 are defined as

$$\begin{aligned} c &= \sqrt{a^2 + b^2}, \\ \varphi_1 &= \arctan(b/a), \\ \varphi_2 &= \arctan(a/b). \end{aligned}$$

2.2 Complex Variables

A complex variable $z = a + bj$ is the sum of a real number and an imaginary number, where $a = \operatorname{Re}\{z\}$ is the real part of z , and $b = \operatorname{Im}\{z\}$ is the imaginary part of z . A complex variable can be written in the polar form as

$$z = r \exp(j\theta)$$

where $r = \sqrt{a^2 + b^2}$ and $\theta = \arctan(b/a)$. Note that $a = r \cos(\theta)$ and $b = r \sin(\theta)$. Here r is the magnitude of z , denoted as $|z|$, and the angle θ can be denoted by the form $\angle z$. The conjugate of the complex variable

$z = a + bj$ is denoted by z^* and is defined as $z^* = a - bj$. The multiplication and division of two complex numbers $z_1 = r_1/\underline{\theta_1}$ and $z_2 = r_2/\underline{\theta_2}$ are

$$z_1 z_2 = (r_1/\underline{\theta_1})(r_2/\underline{\theta_2}) = r_1 r_2 / \underline{\theta_1 + \theta_2}$$

and

$$\frac{z_1}{z_2} = \frac{r_1/\underline{\theta_1}}{r_2/\underline{\theta_2}} = \frac{r_1}{r_2} \underline{\theta_1 - \theta_2}.$$

2.3 Derivative

Here are the rules for the derivatives of the most common basic functions.

Derivative of powers

$$\frac{d}{dx} x^a = a x^{a-1},$$

where a is an integer and x is positive.

Exponential functions

$$\frac{d}{dx} \exp(x) = \exp(x).$$

Trigonometric functions

$$\frac{d}{dx} \sin(x) = \cos(x).$$

$$\frac{d}{dx} \cos(x) = -\sin(x).$$

Rules for combined functions

- Constant rule: if $f(x)$ is constant, then

$$f^{(1)}(x) = 0.$$

- Sum rule: if $f(x)$ and $g(x)$ are functions, α and β are constants, then

$$(\alpha f + \beta g)^{(1)} = \alpha f^{(1)} + \beta g^{(1)}.$$

- Product rule: if $f(x)$ and $g(x)$ are functions, then

$$(fg)^{(1)} = f^{(1)}g + fg^{(1)}.$$

- Quotient rule: if $f(x)$ and $g(x)$ are functions, then

$$\left(\frac{f}{g}\right)^{(1)} = \frac{f^{(1)}g - fg^{(1)}}{g^2}.$$

- Chain rule for composite functions: if $f(x) = h(g(x))$, then

$$f^{(1)}(x) = \frac{\partial h(g(x))}{\partial g(x)} \cdot g^{(1)}(x).$$

Example 1 Calculate the derivative of the function

$$f(x) = x^4 + \sin(x^2) + 7.$$

2.4 Laplace Transforms

The Laplace transform is a linear operator of a function $f(t)$ with a real argument t ($t \geq 0$) that transforms $f(t)$ to a function $F(s)$ with a complex argument s , given by the integral

$$F(s) \triangleq \int_0^\infty f(t)e^{-st}dt,$$

where $s = \sigma + \omega j$ is a complex frequency parameter with real numbers σ and ω . An alternate notation for the Laplace transform is $\mathcal{L}\{f\}$.

Example 2 Find the Laplace transform of the impulse function $\delta(t)$.

Example 3 Find the Laplace transform of the step $f(t) = a$ and ramp $f(t) = bt$ functions.

2.4.1 Properties of Laplace Transforms

Superposition

The Laplace transform is linear, which means that the principle of superposition applies:

$$\mathcal{L}\{\alpha_1 f_1(t) + \alpha_2 f_2(t)\} = \alpha_1 F_1(s) + \alpha_2 F_2(s).$$

Differentiation

The transform of the derivative of a signal is related to its Laplace transform and its initial condition as follows:

$$\mathcal{L}\{f^{(1)}\} = sF(s) - f(0) \quad (2.2)$$

where $f^{(1)}$ is the first derivative of f . We denote the derivative of $f^{(1)}$ by $f^{(2)}$ and call $f^{(2)}$ the second derivative of f . Continuing in this manner, we obtain the function $f^{(n)}$, which is called the n th derivative of f . Another application of (2.2) leads to

$$\mathcal{L}\{f^{(2)}\} = s^2F(s) - sf(0) - f^{(1)}(0). \quad (2.3)$$

Repeated application of (2.2) leads to

$$\mathcal{L}\{f^{(n)}\} = s^nF(s) - s^{n-1}f(0) - s^{n-2}f^{(1)}(0) - \dots - s^0f^{(n-1)}(0).$$

Integration

The Laplace transform of the integral of a signal $f(t)$ is simply the division of the Laplace transform of the function by s , that is,

$$\mathcal{L}\left\{\int_0^t f(\tau)d\tau\right\} = \frac{F(s)}{s}.$$

Convolution

The convolution of two functions f and g is written $f(t) * g(t)$, and its definition is

$$f(t) * g(t) = \int_0^t f(\tau)g(t-\tau)d\tau.$$

Convolution in the time domain corresponds to multiplication in the frequency domain, that is,

$$\mathcal{L}\{f_1(t) * f_2(t)\} = F_1(s)F_2(s),$$

where $\mathcal{L}\{f_1(t)\} = F_1(s)$ and $\mathcal{L}\{f_2(t)\} = F_2(s)$.

Time Shifting

Let $f(t - \tau)$ be a function which is delayed by τ seconds of $f(t)$. Then,

$$\mathcal{L}\{f(t - \tau)\} = \exp(-\tau s)F(s).$$

2.4.2 Table of selected Laplace transforms

The following table provides Laplace transforms for many common functions of a single variable.

Table 2.1: Table of Laplace Transforms

Function	Time domain	Frequency domain
impulse	$\delta(t)$	1
step	1	$\frac{1}{s}$
ramp	t	$\frac{1}{s^2}$
parabola	t^2	$\frac{2!}{s^3}$
exponential	e^{-at}	$\frac{1}{s+a}$
sine	$\sin(\omega t)$	$\frac{\omega}{s^2+\omega^2}$
cosine	$\cos(\omega t)$	$\frac{s}{s^2+\omega^2}$
	$t \sin(\omega t)$	$\frac{2\omega s}{(s^2+\omega^2)^2}$
	$t \cos(\omega t)$	$\frac{s^2-\omega^2}{(s^2+\omega^2)^2}$

2.4.3 Using Laplace Transforms to Solve Differential Equations

Laplace transforms can be used to solve differential equations using the above properties.

Example 4 Find the solution to the differential equation

$$x^{(2)}(t) + x(t) = 0,$$

where $x(0) = 0$ and $x(\pi/2) = 1$.

Solution: Using (2.3), the Laplace transform of the differential equation is

$$s^2 X(s) - sx(0) - x^{(1)}(0) + X(s) = 0.$$

Using the initial condition $x(0) = 0$ and solving $X(s)$, we have

$$X(s) = \frac{x^{(1)}(0)}{s^2 + 1}.$$

Taking the inverse Laplace transform yields $x(t) = x^{(1)}(0) \sin t$. Based on the condition $x(\pi/2) = 1$, we have $x^{(1)}(0) = 1$. Therefore, we get $x(t) = \sin(t)$.

2.5 Cramer's Rule

Cramer's rule is an explicit formula for the solution of a system of linear equations with as many equations as unknowns, valid whenever the system has a unique solution. Consider a system of n linear equations for n unknowns, represented in matrix multiplication form as follows:

$$Ax = b,$$

where the $n \times n$ matrix A has a nonzero determinant, and the vector $x = [x_1, \dots, x_n]^T$ is the column vector of the variables. Then Cramer's rule states that the system has a unique solution, whose individual values for the unknowns are given by

$$x_i = \frac{\det(A_i)}{\det(A)},$$

where A_i is the matrix formed by replacing the i -th column of A by the column vector b .

Chapter 3

Physics

3.1 Newton's laws of motion

3.1.1 Inertial Reference Frame

In classical physics, an inertial frame of reference is a frame of reference that is not undergoing acceleration. The motion of a body can only be described relative to something else - other bodies, observers, or a set of space-time coordinates. If the coordinates are chosen badly, the laws of motion may be more complex than necessary. Take the example of a person driving a car. If we take the road as the inertial reference frame, the car is moving forward, but trees along the roadside are not moving. However, if we take the car as the inertial reference frame, then the car is not moving, but the trees are moving behind.

3.1.2 Spring and Damper

Spring

Hooke's law of elasticity states that the force F_s needed to extend or compress a spring by some distance x scales linearly with respect to that distance, that is,

$$F_s = kx$$

where k is a constant factor characteristic of the spring, and x is the distance of the free end of the spring from its "relaxed" position.

Damper

A damper is a mechanical device which resists motion via viscous friction. The resulting force is proportional to the velocity, but acts in the opposite directions, slowing the motion and absorbing energy. It is commonly used in conjunction with a spring. The viscous friction is

$$F_v = bv,$$

where b is the viscous friction constant, and v is the velocity. The symbol for a damper is shown in Fig. 3.1.

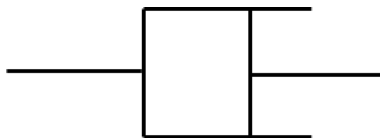


Figure 3.1: Damper

3.1.3 Translational and Rotational Motions

Newton's laws of motion are three physical laws that, together, laid the foundation for classical mechanics.

First law

In an inertial frame of reference, an object either remains at rest or continues to move at a constant velocity, unless acted upon by an outside force.

Second law

In an inertial frame of reference, the sum of all forces f on an object is equal to the mass m of that object multiplied by the acceleration a of the object, that is,

$$f = ma.$$

Since

$$a = \frac{dv}{dt} = \frac{d^2x}{dt^2}$$

and

$$v = \frac{dx}{dt},$$

the following relationships hold by applying the Laplace transform $F(s) = mA(s)$, $F(s) = m s V(s)$, and $F(s) = m s^2 X(s)$ by assuming zero initial conditions.

In an inertial frame of reference, the sum of all torques τ on a one-dimensional rotational system is equal to the moment of inertia I multiplied by the angular acceleration α , that is,

$$\tau = I\alpha.$$

Here the torque is the turning effect of a force

$$\tau = fr,$$

where f is the force applied perpendicular to the rotational system, r is the distance from the place where the force is applied to the rotational center, and the moment of inertia I is a measure of the object's resistance to changes to its rotation, where $I = mr^2$. Similar to the translational motion, we have the relationships

$$\alpha = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}$$

and

$$\omega = \frac{d\theta}{dt},$$

where θ is the angle, ω is the angular velocity. By applying the Laplace transform, the above relationships can be written as

$$A(s) = s\Omega(s) = s^2\Theta(s)$$

in the complex domain by assuming zero initial conditions.

Third law

When one body exerts a force on a second body, the second body simultaneously exerts a force equal in magnitude and opposite in direction on the first body.

In this activity, you will model the bungee jump as a mass-spring-damper system. Project: Modeling a Bungee Jump.

In Italy, thrill seekers can bungee jump off the Colossus bridge which is 152 meters (488 feet) high. Bungee jumping uses a long elastic cord attached to the jumper. When the person jumps, the cord extends then stretches, slowing the jumper's descent. The jumper slows to a stop, then rebounds upwards. The jumper oscillates up and down until the kinetic energy dissipates, coming to rest at an equilibrium position. A simple model of the bungee jump is a mass-spring-damper system. The bungee jumper is modeled as a mass, m , the bungee cord as a spring (with spring constant k), and the dissipation of energy as a damper (with damping coefficient b).

3.2 Circuits

3.2.1 Electric circuits

Below we show some elements of electric circuits: Resistor, Capacitor and Inductor. The relationships between the voltage v across and current i through the elements are shown in Table 3.1. It is more convenient to use the impedance rather than the differential equations. In Table 3.1, C is the

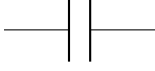
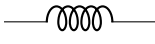
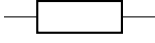
	Symbol	Equation	Impedance
Capacitor		$i = C \frac{dv}{dt}$	$\frac{V}{I} = \frac{1}{Cs}$
Inductor		$v = L \frac{di}{dt}$	$\frac{V}{I} = Ls$
Resistor		$v = Ri$	$\frac{V}{I} = R$

Table 3.1: Basic Electric Elements

capacitance, L is the inductance, R is the resistance, and s is a complex variable.

3.2.2 Kirchhoff's circuit laws

Kirchhoff's circuit laws include Kirchhoff current law and Kirchhoff voltage law.

Kirchhoff's current law

This law states that, for any node in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node; or equivalently

The algebraic sum of currents leaving a node equals to the algebraic sum of currents entering that node.

Recalling that current is a signed quantity reflecting direction towards or away from a node, this principle can be succinctly stated

$$\sum_{k=1}^n I_k = 0,$$

where n is the total number of branches with currents flowing towards or away from the node. For the circuit shown in Fig. 3.2, Kirchhoff's current law suggests that

$$i_2 + i_3 = i_1 + i_4.$$

Kirchhoff's voltage law

This law states

The algebraic sum of all voltages taken around a closed loop in a circuit is zero.

Similarly to Kirchhoff's current law, the voltage law can be stated as

$$\sum_{k=1}^n V_k = 0,$$

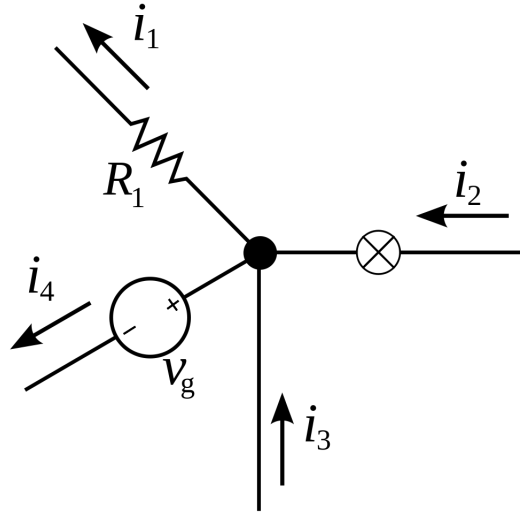


Figure 3.2: Kirchhoff's current law. Source: https://en.wikipedia.org/wiki/Kirchhoff's_circuit_laws

where n is the total number of voltages measured. For the circuit shown in Fig. 3.3, Kirchhoff voltage law suggests that

$$v_1 + v_2 + v_3 - v_4 = 0.$$

If the voltage is positive, the current flows from a high voltage to a low voltage, and it is negative otherwise.

3.2.3 Operational Amplifier

In control systems, one additional circuit element is usually present - the operational amplifier. Operational amplifier also known as op-amp is a power amplifier. It is normally used in sensor circuits to amplify a weak signal, and is also used in compensation circuits. Op-amps are the basis for many circuits, like audio amplifiers, waveform generators, and filters. These circuits are made up of individual op-amps that amplify basic computations, like sums and differences, or integration and differentiation. Op-amp is an electronic amplifier used as a building block to implement transfer functions. The op-amp consists of two inputs: non-inverting input (V_+) and inverting input (V_-). The op-amp is designed and constructed such that the input

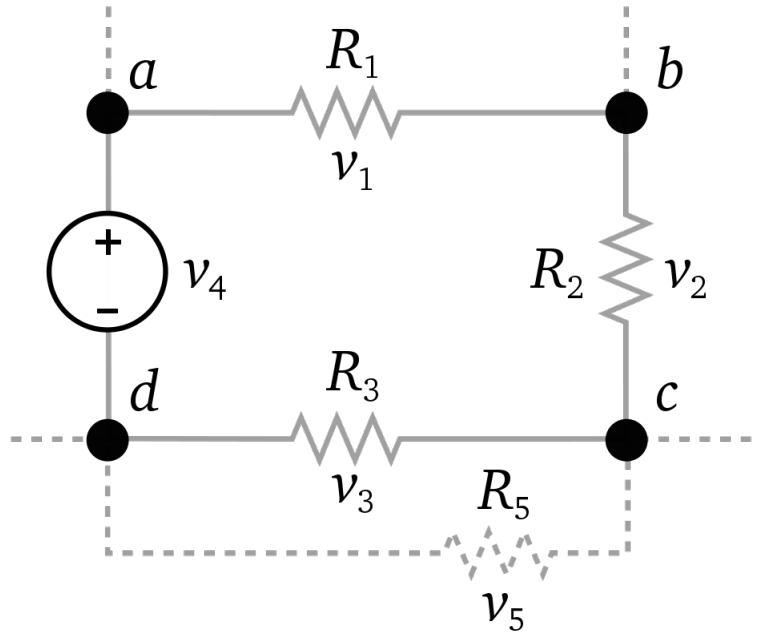


Figure 3.3: Kirchhoff's voltage law. Source: https://en.wikipedia.org/wiki/Kirchhoff's_circuit_laws

impedance is very high, resulting in i_+ and i_- being very small. Op-amp chips can have many different architectures.

Open-loop Operational Amplifier

A simplified circuit of an open-loop op-amp is shown in Fig. 3.4. It has a differential input. The input voltage is the difference between the positive and negative terminals. The output voltage of the op-amp

$$V_{out} = A_{OL}(V_+ - V_-),$$

where A_{OL} is the open-loop gain of the amplifier. Additionally, the amplifier gain is very large. However, the magnitude of A_{OL} is not well controlled by the manufacturing process, and so it is impractical to use an open-loop amplifier as a stand-alone differential amplifier.

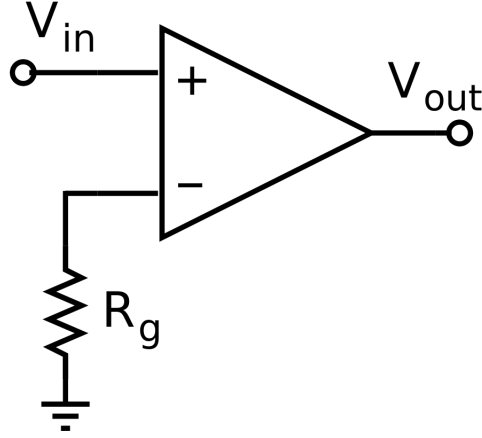


Figure 3.4: Open-loop op-amp. Source: https://en.wikipedia.org/wiki/Operational_amplifier

Closed-loop Operational Amplifier

In practice, op-amps are most commonly used in a negative feedback circuit. That is, a portion of the output is fed back to the negative terminal. This feedback stabilizes the op-amp's behavior. In the non-inverting amplifier shown in Fig. 3.5, the negative feedback is achieved via the voltage divider R_f and R_g . Equilibrium is established when the positive and negative terminals of the op-amp have the same voltage, that is, $V_+ = V_- = V_{in}$. Based on the voltage division, we have

$$\frac{V_{in}}{R_g} = \frac{V_{out}}{R_g + R_f}.$$

Therefore,

$$V_{out} = \left(1 + \frac{R_f}{R_g}\right) V_{in}.$$

Ideal Operational Amplifier Characteristics

We assume that all op-amps used in control circuits are ideal. The ideal op-amp is a standard set of assumptions used to describe most op-amp behavior. It has the following characteristics:

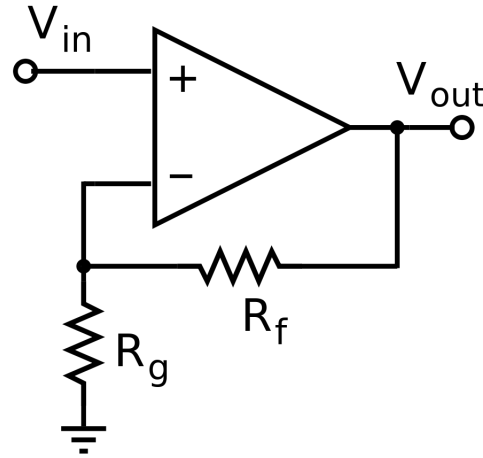


Figure 3.5: Closed-loop op-amp. Source https://en.wikipedia.org/wiki/Operational_amplifier

- **Infinite Input Impedance:** The op-amp has infinite input impedance which makes it useful in signal conditioning applications to avoid the attenuation of weak signals, in instrumentation and many more. The infinite input impedance of the op-amp is very desirable as it results in zero input current.
- **Zero Output Impedance:** Another important feature of the ideal op-amp is the zero output impedance which makes it very desirable in the signal amplifier. Because of the zero output impedance of the op-amp, it acts as an Ideal Voltage Source. With any amount of current flowing through the op-amp, the signal strength will not be disturbed. As the zero output impedance will result zero voltage drop at the output terminals of the op-amp so the signal will not get affected irrespective of the amount of current that flows through the op-amp.

Infinite input impedance means that the amplifier does not draw current (load) from the previous stage. Output voltage resulting from the previous stage applies to the amplifier input without current, so the previous stage does not consume additional power from power supply (higher efficiency), and the amplifier does not make input signal distorted. In the same way, the amplifier with zero output impedance will load all output voltage to the next

stage and the output signal will not be distorted. The output voltage is not affected by the load connected to the output terminal.

The formulas for an ideal op-amp are

$$i_+ = i_- = 0, \quad (3.1)$$

$$v_+ = v_-. \quad (3.2)$$

The amplification factor is based on the ratio of the feedback resistance to the input resistance. *Note that the current flowing into the op-amp is zero due to the infinite impedance. However, the current out of the op-amp is NOT zero.*

3.3 Mechatronics

3.3.1 Law of Motors

Consider a conductor of length l m is arranged in a magnetic field of flux density B teslas carrying a current I amp. Then there is a force on the conductor. The size of this force is

$$F = Bli \sin(\theta),$$

where θ is the angle made by the conductor with the magnetic field. This equation is the basis of conversion of electric energy to mechanical work and is called the law of motors. The direction of the force can be found by application of the right hand rule as follows: thumb follows the direction of current flow, the middle figure points to the polarities from N to S . The direction where the palm faces is the direction of the force.

3.3.2 Law of Generator

If a straight conducting wire is placed in a plane perpendicular to a uniform magnetic field, and is moving in a direction perpendicular to the field, then each charge q in the wire experiences a magnetic force of magnitude $F = qvB$. The negatively charged electrons will accelerate in response to this force. Since they cannot leave the wire, negative charges will accumulate on one end of the wire, while positive charges will be left behind on the other end. The separated charges produce an electric field, which exerts a force on the

other charges in the wire. This electric force opposes the magnetic force. If we place the wire on a conducting rail, a current will start to flow in the circuit formed by the rail and the wire. The voltage across the wire is

$$e = Blv,$$

where B is the magnetic field density, l is the length of conductor, and v is the velocity. The direction of the current follows the right hand rule: thumb follows the direction of motion, the middle points to the direction of the magnetic field from N to S , the palm faces the direction of the current.

Part II

Analysis and Design of Control Systems

Chapter 4

Modeling of Dynamic Systems

The learning objectives of this chapter:

- Recognize that transfer functions can describe the dynamic behavior of physical systems;
- Understand the application of Laplace transforms and their role in obtaining transfer functions;
- Model the basic mechanical systems;
- Model the basic electrical systems;
- Model the basic mechatronic systems;
- Understand the important role of modeling in the control system design process.

Dynamic systems are everywhere and you interact with them every day. You can find multiple dynamic systems in a car and the car itself is a dynamic system. The circuits in your electronic devices are also dynamic systems. Even how global climate changes over time is a dynamic process. A dynamic system contains variables that change over time. There are three types of variables: the inputs, the outputs, and the internal states. The input is a stimulus to the system, and it causes other variables to change. An internal state is a system variable that evolves over time. Its value depends on the inputs, other internal states and the properties of the system. Output is a result or variable that is of interest to you and similarly depends on the input, the internal states, and the system properties.

For example, let us look at your house's thermal system. The outdoor temperature is the input and it causes an internal state, your room temperature, to change. How does that room temperature change? Well, that depends on the outdoor temperature, the room temperature itself, the house's insulation.

4.1 Linear Time-Invariant Systems

4.1.1 Linearity

A system is linear if when you multiply the input by a , the output will scale up by a as well. For example, if you slide a dimmer on a light up and down, you expect the light's intensity to scale up and down linearly with the input. If the slider is halfway, the light is at half intensity.

Mathematically, a system is linear if a linear combination of signals sent through the system results in the same linear combination of corresponding outputs.

4.1.2 Time Invariance

A system is time invariant when whether you apply an input to a system now or N seconds from now, the output will be identical except for a delay of N seconds.

Mathematically, a system is time invariant if a time shift of the input sequence causes a corresponding shift in the output sequence.

4.1.3 Linear System Representations

A mathematical modeling is a description of a system using mathematical concepts and language. Linear systems can be represented using

- Linear differential equations
- Convolution with impulse response
- Transfer function with poles-zeros.

To model the dynamics of a control system, we can use differential equations. In traditional control theory, it is more convenient using a transfer

function to represent the input-output relationship in the frequency domain. The transfer function $H(s)$ of an LTI system is defined as the ratio of the Laplace transform of the output of the system to its input assuming all zero initial conditions. The standard form of the transfer function is given as

$$H(s) = \frac{b(s)}{a(s)}$$

where $a(s)$ and $b(s)$ are polynomials in s and $a(s)$ is *monic*. A polynomial is called monic if the coefficient of the highest power in s is 1. The *relative degree* of a transfer function is by definition the difference between the degrees of the denominator and numerator polynomials. A rational transfer function is called *proper* if its relative degree is at least 0, and *strictly proper* if its relative degree is at least 1.

A control system consists of four major components: process, sensor, controller and actuator. The process combined with the actuator is called the “plant”. Here we will discuss how to model each component using the underlying physics.

4.2 Modeling the Process

4.2.1 Autonomous Vehicle Cruise Control

Here we use a simplified model for the vehicle dynamics. Let us use x to denote the distance from the reference 0 distance point. The inertia reference frame is defined as the positive direction to the right. To find the transfer function, we first draw the free body diagram. A free body diagram is a graphical illustration used to visualize the applied forces, movements, and resulting reactions on a body in a given condition. The free body diagram of the autonomous vehicle cruise control is shown in Fig. 4.1. Here u is the engine torque. It is the input. According to Newton’s second law, we have

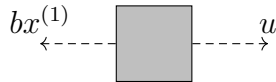


Figure 4.1: Free Body Diagram

$$u - bx^{(1)} = mx^{(2)},$$

where b is constant which is related to the air resistance. Here x is the output and u is the input. Applying the Laplace transform, we have

$$U(s) - bsX(s) = ms^2X(s)$$

if we assume all initial conditions are zeros. The transfer function is

$$\frac{X(s)}{U(s)} = \frac{1}{ms^2 + bs}.$$

To obtain the transfer function from the engine torque to the velocity, we can use the relationship $sX(s) = V(s)$. Then

$$\frac{V(s)}{U(s)} = \frac{sX(s)}{U(s)} = \frac{1}{ms + b}.$$

4.2.2 Quadcopter Drones

Figure 4.2 shows a quadcopter with four rotors. Two of them rotate in the clockwise direction while the other two rotate in the counterclockwise direction. The rotor blades are arranged so that the rotating rotor blades generate lifting forces. The quadcopter drone hovers in the air when the lifting force is balanced with the gravity force applied to the quadcopter. By changing the rotational speed of the four rotors, various translational and rotational motions can be made by quadcopters. The torques to the four rotors are defined as T_A , T_B , T_C , and T_D . For the rotational motion, the quadcopter can make the roll, pitch, and yaw motions. The definitions are given below:

- Roll: rotate around the front-to-back axis;
- Pitch: rotate around the side-to-side axis;
- Yaw: rotate around the vertical axis.

To produce the pitch motion, we can increase the rotational speeds of rotors C and D, and decrease the rotational speeds of rotors A and B. To produce the roll motion, we can increase the rotational speeds of A and D, and decrease the rotational speeds of B and C. To produce the yaw motion, we can increase the rotational speeds of A and C, and decrease the rotational speeds of B and D. To make the drone hovering in the air, the increase amount of the speed

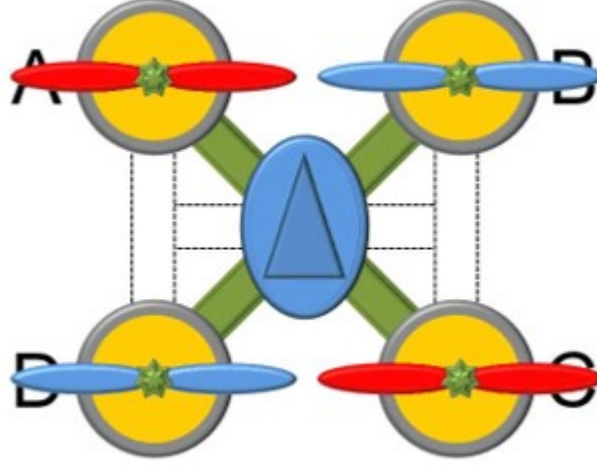


Figure 4.2: Quadcopter. Source: <https://en.wikipedia.org/wiki/Quadcopter>

should be the same with the decrease amount of the speed. Let us denote the control torque for positive θ motion as T_θ . Then the transfer function from the torque to the pitch angle θ is

$$\frac{\Theta(s)}{T_\theta(s)} = \frac{1}{I_y} \frac{1}{s^2}.$$

This model is called a double integrator model because it contains the product of two $1/s$, which corresponds to the Laplace transform of an integration.

4.2.3 The 3D Printer

In this example, let us model the motion for a motor-driven gantry used for a 3D printer as shown in Fig. 4.3. The printhead can move independently on the X and Y axis. For each axis, there is a motor that provides the torque to move the gantry along it. The rotational motion of the motor is converted to linear motion of the gantry using a linear actuator. The motion of the

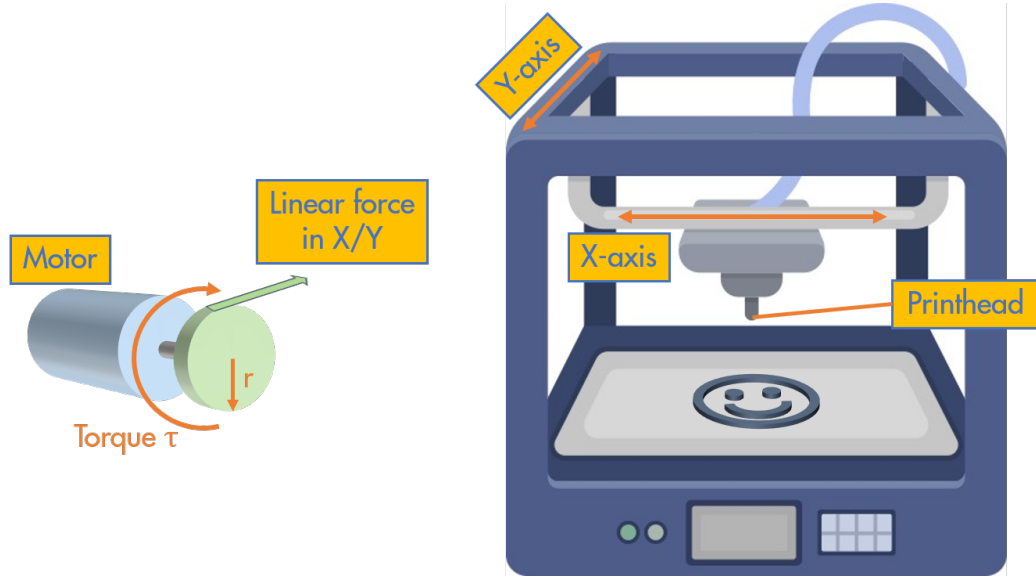


Figure 4.3: 3D Printer. Source: <https://matlabacademy.mathworks.com/>

printhead on either axis is governed by the differential equations:

$$\begin{aligned}\ddot{x} &= \frac{1}{m} (F_a - F_d), \\ F_a &= \frac{\tau}{r}, \\ F_d &= B\dot{x},\end{aligned}$$

where x is the printhead position, B is the damping coefficient, r is the radius of the gear, and τ is the torque produced by the motor.

4.3 Modeling controllers

In the last section, Newton's law is used to model the process, which is usually a mechanical system. This section is devoted to the modeling of electric circuits. Electric circuits are used frequently in control systems, especially in controllers. This is because electric signals are easy to be processed and manipulated. Although the controller is more and more implemented with digital logic by microprocessors and computers, analog circuits are still used to perform certain functions. Analog circuits have two main advantages

over digital circuits. They are faster and generally cheaper than a digital implementation for a simple controller. A sensor is a device whose purpose is to convert the process output into an electrical signal so that it can be easily manipulated by controllers. Furthermore, analog circuits must be used in power amplifier for mechatronic control and anti-alias pre-filters for digital control.

Electric circuits consist of the intersections of the power sources such as voltage and current and other electronic components, such as resistor, capacitor, inductor. A very important building block in control system design is the operational amplifier, which is short for op-amp. The op-amp itself is a negative feedback control system.

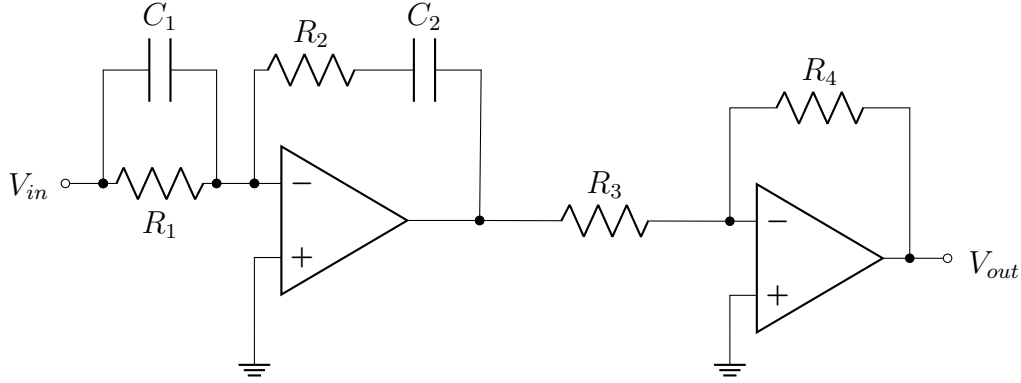


Figure 4.4: PID Controller

A PID controller circuit is shown in Fig. 4.4. To find the transfer function of the PID controller, we can find the impedance first. The impedance of the parallel connection of R_1 and C_1 is

$$Z_1 = \frac{R_1}{R_1 C_1 s + 1}.$$

The impedance of the series connection of R_2 and C_2 is

$$Z_2 = \frac{1 + C_2 R_2 s}{C_2 s}.$$

Let us denote the output voltage of the first operation amplifier is V . Then the transfer function from V_{in} to V is

$$\frac{V(s)}{V_{in}(s)} = -\frac{Z_2(s)}{Z_1(s)} = -\frac{1 + C_2 R_2 s}{C_2 s} \frac{R_1 C_1 s + 1}{R_1}.$$

The transfer function from V to V_{out} is

$$\frac{V_{out}(s)}{V(s)} = -\frac{R_4}{R_3}.$$

Therefore, the PID controller can be written as

$$\frac{V_{out}}{V_{in}} = k_P + \frac{k_I}{s} + k_D s,$$

where

$$k_P = \frac{R_4}{R_2} \frac{C_1}{C_2} + \frac{R_2 R_4}{R_1 R_3} \quad (4.1)$$

$$k_I = \frac{R_4}{R_1 C_2 R_3} \quad (4.2)$$

$$k_D = \frac{C_1 R_2 R_4}{R_3} \quad (4.3)$$

By tuning the electric elements of resistor and capacitor, the parameters of the PID controller are adjusted.

4.4 Modeling sensors and actuators

Electric current and magnetic fields interact in two ways that are particularly important to the understanding of the operation of most electromechanical actuators and sensors.

4.4.1 Sensors

Position Sensor

A position sensor facilitates measurement of mechanical positions, such as absolute position (location) or relative position (displacement) in terms of linear travel, rotational angle, or three-dimensional space. Common types of position sensors include potentiometer and encoder. A potentiometer is a three-terminal resistor with a sliding or rotating contact that forms an adjustable voltage divider, as illustrated in Fig. 4.5. Let us assume the

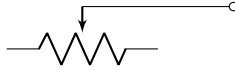


Figure 4.5: Position Sensor

resistance of the resistor is R , and the right part of the voltage divider is R_1 . Then the voltage across R_1 is

$$V_s(t) = \frac{R_1}{R}V,$$

where V_s is the sensor output, V is the applied DC voltage, and R_1 is proportional to the displacement. Therefore, the mechanical position is translated into the voltage.

Similar to optical position sensors, *ultrasonic position sensors* emit a high-frequency sound wave generated typically from a piezoelectric crystal transducer. The ultrasonic waves generated from the transducer are reflected from the object being measured, or target, back to the transducer where an output signal is generated. Ultrasonic sensors can function to perform as proximity sensors, where they report on an object being within a specified range of the sensor, or as a position sensor which provides ranging information. The advantages of ultrasonic position sensors are that they can work with target objects of different materials and surface characteristics, and can detect small objects over larger distances than other types of position sensors.

A passive infrared sensor (PIR sensor) is an electronic sensor that measures infrared (IR) light radiating from objects in its field of view. They are most often used in PIR-based motion detectors. PIR sensors are commonly used in security alarms and automatic lighting applications.

Throttle position sensors provide feedback to the fuel injection system of an automobile. Motion sensors detect the movement of an object and can be used to trigger action (such as illuminating a floodlight or activating a security camera). Proximity sensors as well can detect that an object has come within range of the sensor. Both sensors, therefore, might be considered as a specialized form of position sensors.

Velocity Sensor

A special DC generator, called tachometer, can be used to measure velocity. A tachometer is an instrument measuring the rotation speed of a shaft or

disk, as in a motor or other machine. The word tach comes from Greek which means speed. Tachometers or revolution counters on cars, aircraft, and other vehicles show the rate of rotation of the engine's crankshaft, and typically have markings indicating a safe range of rotation speeds. According to the law of generator, the output of the sensor

$$e = K_e \omega.$$

The tachometer output voltage is directly proportional to its shaft velocity.

Acceleration Sensor

An accelerometer is an electromechanical device used to measure acceleration forces. Such forces may be static, like the continuous force of gravity or, as is the case with many mobile devices, dynamic to sense movement or vibrations. Acceleration is the measurement of the change in velocity.

4.4.2 Actuator

The DC motor is one of the most common actuators. An electric motor is an electrical machine that converts electrical energy into mechanical energy. Most electric motors operate through the interaction between the motor's magnetic field and electric current in a wire winding to generate force in the form of torque applied on the motor's shaft. Electric motors are the most common actuators in a control system. Electric motors can be found in most electronic devices, such as computer hard-disk drive control, electric vehicles, quadcopters, washing machines, etc. Electric motors work under the principles of law of motors and law of generators. The stator is the part which does not move, having two magnets on the side. The armature in the middle is also called a rotor, which is the part that spins. The axle that sticks out of the motor is a shaft. The input of the motor is usually a voltage source. The circuit connects the commutator with two brushes on the side. The current flows through the brush, commutator, and armature loop and back through the other side. The commutator will spin along the armature. The brushes slide along as the commutator spin, but they always maintain contact with the commutator. According to the law of motor, there is a force on the armature when a current carry armature is placed in a magnetic field at the right angle.

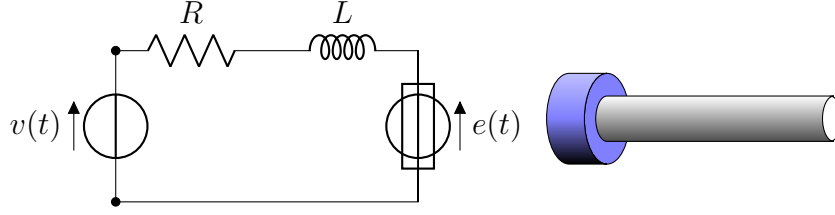


Figure 4.6: Electric Motors

An equivalent electric circuit of a DC motor is shown in Fig. 4.6. According to the motor equations, the torque T on the rotor and the back electromotive force voltage are given by

$$T = K_t i_a \quad (4.4)$$

$$e = K_e \theta_m^{(1)} \quad (4.5)$$

where K_t is the torque constant, and K_e is the electric constant. Based on Newton's laws,

$$K_t i_a - b \theta_m^{(1)} = J_m \theta_m^{(2)}.$$

Based on KCL law,

$$L i_a^{(1)} + R i_a = v - K_e \theta_m^{(1)}.$$

By applying Laplace transform to both equations, we have

$$J_m s^2 \Theta_m + b s \Theta_m = K_t I_a(s) \quad (4.6)$$

$$K_e s \Theta_m + L s I_a(s) + R I_a(s) = V(s) \quad (4.7)$$

The voltage is the input, and the rotational speed is the output. The transfer function from the voltage $V(s)$ to the rotational speed $\Omega(s)$ is

$$\frac{\Omega(s)}{V(s)} = \frac{s \Theta(s)}{V(s)} = \frac{K_t}{J_m L s^2 + (R J_m + b L) s + K_t K_e}.$$

Chapter 5

Analysis of Control Systems

The learning outcomes in this chapter are shown as follows:

- Be aware of block diagrams and their role in analyzing control systems
- Understand the concept of stability of dynamic systems
- Be aware of the impact of zeros and poles on the system response.
- Recognize the direct relationship between the pole locations of second-order systems and the transient response.

The chapter and the next chapter are concerned with the analysis and design of closed-loop control systems. System analysis is the study and testing of a system to determine certain properties of the system, while system design is the process of constructing a system to obtain the desired response.

5.1 Block Diagrams

In Chapter 4, the transfer function is obtained by first applying the Laplace transform to differential equations, and then algebraically solving the set equations. The algebraic equations after the Laplace transform can be arranged in a way that two blocks do not interact unless the output of a block is the input of another block. Each block represents the relationship between its input and output. The relationships between blocks are indicated by lines and arrows. When all components in a component block diagram are replaced by its corresponding transfer function, it becomes a block diagram.

By viewing the transfer function inside each block as a number, the overall transfer function can be obtained by graphical simplification. It provides a different way to solve the algebraic equation.

5.1.1 Terminologies

Path: a sequence of connected blocks, the route passing from one node to another in the direction of signal flow of the blocks without including any block more than once.

Forward Path: a path from the input to output such that no node is included more than once.

Loop: Any closed path that returns to its starting node without passing through any node more than once.

Path Gain: The path gain is the product of transfer functions making up the path.

Loop Gain: the path gain associated with a loop - that is, the product of gains in a loop.

5.1.2 Elementary Block Diagram

Series Connection

Two subsystems G_1 and G_2 are connected in series as shown in Fig. 5.1, that is, the output of G_1 is the input of G_2 . The transfer function for the overall system is

$$\frac{Y(s)}{R(s)} = G_1 G_2.$$

The overall transfer function is the path gain from input $R(s)$ to output $Y(s)$.

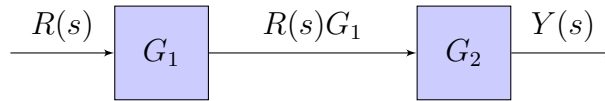


Figure 5.1: Series connection

To show this, the algebraic approach can be used. The output of G_1 is $R(s)G_1$, which is the input of G_2 . Then, $Y(s) = R(s)G_1G_2$.

Parallel Connection

Two subsystems G_1 and G_2 are connected in parallel as shown in Fig. 5.2, that is, the subsystems G_1 and G_2 have the same input $R(s)$ and the sum of their outputs is the output of the overall system. The transfer function for the overall system is

$$\frac{Y(s)}{R(s)} = G_1 + G_2.$$

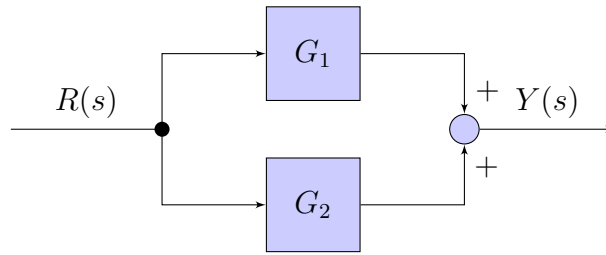


Figure 5.2: Parallel connection

The outputs of G_1 and G_2 are $R(s)G_1$ and $R(s)G_2$, respectively. The output $Y(s) = G_1R(s) + G_2R(s) = (G_1 + G_2)R(s)$.

Feedback Connection

Two subsystems are connected in a feedback arrangement as shown in Fig. 5.3. The output $Y(s)$ is

$$Y(s) = (R(s) - G_2Y(s))G_1.$$

By arranging the terms, we have

$$Y(s)(1 + G_2G_1) = R(s)G_1.$$

Therefore, the transfer function is

$$\frac{Y(s)}{R(s)} = \frac{G_1}{1 + G_1G_2}.$$

A general rule for the feedback connection is that the transfer function is a ratio of the forward path gain and 1 minus the loop gain. Here the forward path gain is G_1 and the loop gain is $-G_1G_2$. Note that when the loop gain is calculated, all the signs along the loop must be included.

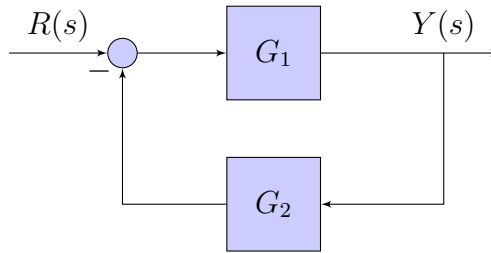


Figure 5.3: Block Diagram

Block Diagram Algebra

In block diagram algebra, the pick-off point can be considered as parentheses, the parallel connection of two blocks is like a summation. We will discuss three algebraic operations of block diagrams.

- Moving a block across a pick-off point: The block diagram shown in Fig. 5.4(a) is equivalent to the block diagram shown in Fig. 5.4(b). From both block diagrams, we have $Y_1(s) = G_1R(s)$ and $Y_2(s) = G_2R(s)$.
- Moving a block across a summer: The block diagram shown in Fig. 5.5(a) is equivalent to the block diagram shown in Fig. 5.5(b). From both block diagrams, we have $Y(s) = G_1(U_1(s) + G_2U_2(s))$ and $Y(s) = G_2G_1(U_1(s)/G_2 + U_2(s))$. When a block is moved from the input side of a summer to the output side of a summer, all input paths are divided by the moving block, and all output paths are multiplied by the moving block.

Block-diagram reduction is a way to change the block-diagram to different forms by moving blocks across pick-off points and summers so that it can be reduced to a simpler form to represent the input-output relationship.

Tips: In general, the block diagram reduction process is easier when pick-off points or summing points are combined into one by moving them. There are no general rules of shifting a pick-off point before and after a summer.

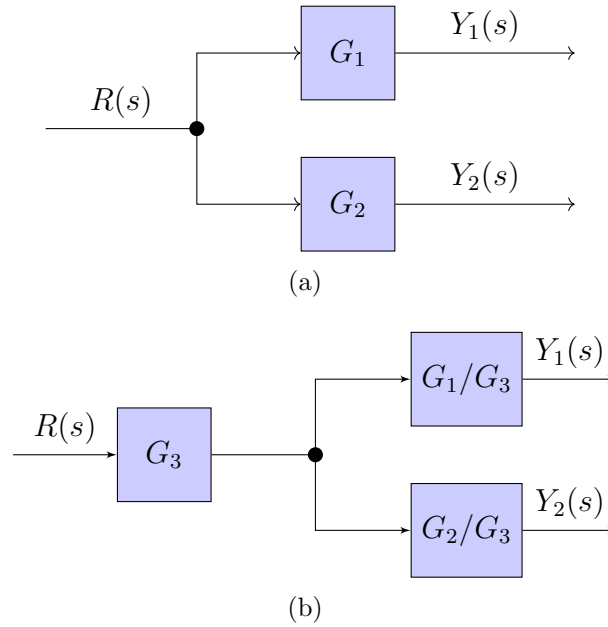


Figure 5.4: Moving a pick-off point

5.2 Transfer Functions

5.2.1 Impulse Response

Two important properties of an LTI system are:

- The system satisfies the principle of superposition.
- If the input signal is delayed, the output is unchanged except that it is delayed by the same amount of time.

If $y_1(t)$ and $y_2(t)$ are the outputs of an LTI system with the inputs $u_1(t)$ and $u_2(t)$, respectively, what is the output of the LTI system with the input $\alpha_1 u_1(t - \tau_1) + \alpha_2 u_2(t - \tau_2)$? The output is $\alpha_1 y_1(t - \tau_1) + \alpha_2 y_2(t - \tau_2)$.

Is there a “standard signal” so that if we know the response of the standard signal, then we know the response of every signal based on the two properties of the LTI system by scaling and shifting the response of the “standard signal”? The answer is yes, and such “standard signal” is the

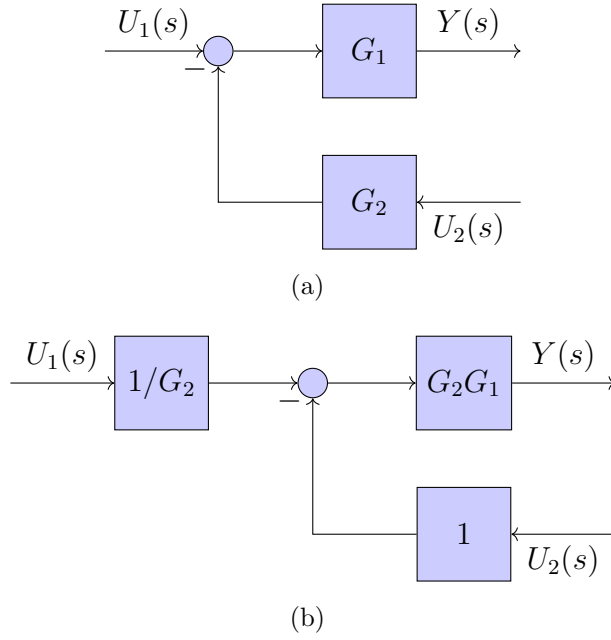


Figure 5.5: Moving a summer

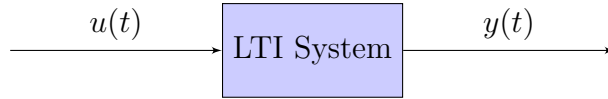


Figure 5.6: LTI System

impulse signal. Mathematically, the unit impulse signal is defined as

$$\delta(t) = 0 \text{ for } t \neq 0, \text{ and } \int_0^{+\infty} \delta(t)dt = 1. \quad (5.1)$$

Any input signal $u(t)$ can be written as

$$u(t) = \int_0^t u(\tau)\delta(t - \tau)d\tau$$

based on the sifting property or the sampling property of the delta function. The input signal is expressed as the integration (infinitely many sum) of impulses signals delayed by τ and scaled by $u(\tau)d\tau$ for $0 \leq \tau < +\infty$. The output $h(t)$ is called *impulse response* when the input signal is $\delta(t)$. Then

according to the superposition principle and time-invariant property, the output of a system is

$$y(t) = \int_0^t u(\tau)h(t - \tau)d\tau \quad (5.2)$$

when the input is $u(t)$.

The output signal $y(t)$ of an LTI system with an input signal $u(t)$ can be described by the convolution integral of the input $u(t)$ with the unit impulse response of the system $h(t)$.

With the assumption of zero initial conditions of the system, (5.2) becomes

$$Y(s) = H(s)U(s)$$

when the Laplace transform is applied to both sides of (5.2), and

$$H(s) = \frac{Y(s)}{U(s)}.$$

The transfer function $H(s)$ of an LTI system is the ratio of the Laplacian transform of the output of the system to its input assuming all zero initial conditions.

The Laplace transform of the delta function is 1. Therefore, the Laplace transform of the impulse response of an LTI system is the transfer function, that is,

$$H(s) = Y(s).$$

Thus, another definition of the transfer function of an LTI system is given as

The transfer function $H(s)$ is the Laplace transform of the unit impulse response $h(t)$ of an LTI system assuming all zero initial conditions.

5.2.2 General Time Response

The steps to calculate the general time response of a system.

1. Determine the model of the system in the form of transfer functions;
2. Determine the Laplace transform of the input signal $U(s)$;
3. Calculate the Laplace transform of the output signal $Y(s) = H(s)U(s)$;
4. Find the inverse Laplace transform of the output signal $Y(s)$, that is, the time response $y(t)$.

5.2.3 Partial Fraction Expansion

If the Laplace transform of the output signal is a rational function, that is, the ratio of two polynomials in s , then a simple method called “partial-fraction expansion” can be used to calculate the time response easily. Consider the general rational form of $Y(s)$ as

$$Y(s) = \frac{b_ms^m + b_{m-1}s^{m-1} + \cdots + b_0}{a_ns^n + a_{n-1}s^{n-1} + \cdots + a_0}.$$

The polynomial in the denominator can be factored out as

$$Y(s) = \frac{1}{a_n} \frac{b_ms^m + b_{m-1}s^{m-1} + \cdots + b_0}{\prod_{i=1}^n (s - p_i)}$$

where p_i are the roots of the denominator polynomial.

Depending on p_i s, we have different cases to discuss:

p_i real and distinct:

In this case, $Y(s)$ can be written as

$$Y(s) = \frac{C_1}{s - p_1} + \frac{C_2}{s - p_2} + \cdots + \frac{C_n}{s - p_n}.$$

To determine the coefficients of C_i , we can use the cover-up method. Let us multiply $s - p_i$ at both sides of the above equation. Then, we have

$$(s - p_i)Y(s) = \frac{C_1}{s - p_1}(s - p_i) + \frac{C_2}{s - p_2}(s - p_i) + \cdots + \frac{C_n}{s - p_n}(s - p_i).$$

Then we let $s = p_i$ which leads to the following formula

$$C_i = (s - p_i)Y(s)|_{s=p_i}.$$

p_i complex:

When the denominator polynomial contains complex roots, the complex roots appear in a pair. The partial fraction expansion includes such terms

$$\frac{C_1 s + C_2}{s^2 + as + b}.$$

To find the coefficients of C_1 and C_2 , the approach of balancing the coefficients of the equation can be used. For the case of purely imaginary roots when $a = 0$, the same approach can be applied.

5.2.4 Zero and Poles

The transfer function $T(s)$ of an LTI system can be written as the ratio of two polynomials in s , that is,

$$T(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \cdots + b_0}{a^n s^n + a_{n-1} s^{n-1} + \cdots + a_0} = \frac{N(s)}{D(s)}.$$

where $N(s) = b_m s^m + b_{m-1} s^{m-1} + \cdots + b_0$ and $D(s) = a^n s^n + a_{n-1} s^{n-1} + \cdots + a_0$ are simple polynomials with real coefficients in the complex variable $s = \sigma + j\omega$. The degree n of the polynomial $D(s)$ is no less than the degree m of the polynomial $N(s)$, that is, $n \geq m$.

The roots of the numerator polynomials $N(s)$ in the complex domain are called **zeros** of the transfer function. The degree of $N(s)$ is m . Therefore, it has m roots being denoted z_1, \dots, z_m such that $N(z_i) = 0$ for $i = 1, \dots, m$. The polynomial $N(s)$ can also be written as $N(s) = b_m \prod_{i=1}^m (s - z_i)$.

The roots of the denominator polynomials $D(s)$ in the complex domain are called **poles** of the transfer function. The poles of a transfer function determine the stability property, and the modes of the system response. The degree of $D(s)$ is n . Therefore, it has n roots being denoted p_1, \dots, p_n such that $D(p_i) = 0$ for $i = 1, \dots, n$. The polynomial $D(s)$ can also be written as $D(s) = a_n \prod_{i=1}^n (s - p_i)$.

Therefore, the transfer function $T(s)$ can be written in the zero-pole form as

$$T(s) = K \frac{\prod_{i=1}^m (s - z_i)}{\prod_{i=1}^n (s - p_i)}$$

where $K = \frac{b_m}{a_n}$ is called the gain of the transfer function.

5.3 Stability

The concept of system stability can be characterized from two perspectives: impulse response or input-output relationship. The current in an electric motor cannot exceed a certain value without damaging the windings.

5.3.1 Stability from Impulse Response

A system is stable if the impulse response approaches zero as time approaches infinity. A system is unstable if the impulse response approaches infinity as time approaches infinity. A system is marginally stable if the impulse response neither decays nor grows but remains constant or oscillates. The poles of a transfer function determine the stability property of a system. Let us consider the following transfer function in the zero pole form:

$$T(s) = K \frac{\prod_{i=1}^m (s - z_i)}{\prod_{i=1}^n (s - p_i)}.$$

Using the partial fraction expansion, $T(s)$ can be written as

$$T(s) = \frac{C_1}{s - p_1} + \frac{C_2}{s - p_2} + \cdots + \frac{C_n}{s - p_n}.$$

Let us consider the case that all poles $\{p_i\}_{i=1}^n$ are real or complex but distinct. The inverse Laplace transform of the term corresponding to

$$\frac{C_i}{s - p_i}$$

is

$$C_i \exp(p_i t).$$

If the pole p_i is real or complex but repeated, the inverse Laplace transform is

$$t C_i \exp(p_i t).$$

Then we know that this term decays to zero when p_i is strictly in the LHP, and diverges when it is strictly in the RHP. A system is stable if and only if $\exp(p_i t) \rightarrow 0$ for all p_i as $t \rightarrow \infty$, that is, the real part of p_i is negative.

Therefore, a system is stable if all poles are strictly in the LHP, and is unstable if any poles are in the RHP. The system is called neutrally stable if

there are non-repeated $j\omega$ poles and all other poles are in the LHP. Note that the double integrator model $1/s^2$ is unstable, and a system with repeated $j\omega$ poles is unstable. This is because the inverse Laplace transform for the term of repeated $j\omega$ poles is $t \exp(\pm j\omega t)$, which is unbounded with time.

A LTI system is stable if all the poles of the transfer function are strictly in the LHP.

A polynomial is called Hurwitz if its roots lie in the open left-half plane. A rational transfer function is called stable if its denominator polynomial is Hurwitz.

5.3.2 Stability from Input-Output Relationship

A system is stable if every bounded input yields a bounded output. A system is unstable if any bounded input yields an unbounded output. The stability in this sense is called BIBO stable. The output of a system can be written as the convolution of the impulse response and input signal, that is,

$$y(t) = \int_0^t h(\tau)u(t - \tau)d\tau.$$

To derive a condition for BIBO stable, let us assume that the input $|u(t)| \leq M$ is bounded, where M is a finite constant. Then, we can find a bound for the output

$$\begin{aligned} |y(t)| &= \left| \int_0^t h(\tau)u(t - \tau)d\tau \right| \\ &\leq M \int_0^t |h(\tau)| d\tau. \end{aligned}$$

Therefore, a system with the impulse response $h(t)$ is BIBO-stable if and only if the integral

$$\int_0^\infty |h(\tau)| d\tau < \infty.$$

A real-life example about stability is the original Tacoma Narrows Bridge, built in 1940. The bridge is notorious for collapsing four months later and

being an example of an improperly designed control system. The winds passing through the Tacoma Narrows and coming in contact with the bridge acted as an external periodic frequency that matched the natural structural frequency of the bridge. As a result of this match, resonance phenomena occur, increasing the sway of the bridge over time until it cannot withstand the vibration. This shows that the process of ensuring that a system is BIBO stable must be rigorous and is one of the most common examples of an engineering disaster as a result of improper design.

5.3.3 Routh Array

Using MATLAB can easily determine the locations of all poles of a transfer function, and thus the stability of an LTI system. There are alternative ways to determine the locations of the poles without actually solving the roots of the denominator polynomial of a transfer function. Here we introduce one approach, called “Routh Array”, which can determine the stability of an LTI system based on the coefficients of the denominator polynomial. Routh Array is also useful in the control design, where the controller is parameterized by several parameters. It is not easy to find the roots when there are symbolic coefficients in the polynomial. However, Routh Array is able to tell us the range of the parameters so that the closed-loop system is stable.

Before introducing the Routh stability criterion, let us discuss the necessary conditions on the coefficients of the denominator polynomial for stability of an LTI system. A necessary condition for stability is that all coefficients of the denominator polynomial are positive. Then based on the law of contraposition, it is equivalent to say that a system is unstable if any of the coefficients of the denominator are zero or negative. Once the denominator polynomial passes the necessary condition test, then you can further inspect the coefficients to determine if the closed-loop system is stable or not.

The first step is to construct the Routh Array. The Routh Array will tell us how many poles are located in the LHP, in the RHP, and on the imaginary axis, but not the locations. But it is enough to determine the stability of the system. For a transfer function of with n th-degree polynomial in the denominator,

$$D(s) = a_n s^n + a_{n-1} s^{n-1} + \cdots + a_1 s + a_0,$$

a table with $n + 1$ arrows needs to be calculated. The first two rows are the

Row 1	a_n	a_{n-2}	a_{n-4}	\dots
Row 2	a_{n-1}	a_{n-3}	a_{n-5}	\dots
Row 3	b_1	b_2	b_3	\dots
Row 4	c_1	c_2	c_3	\dots
\vdots	\vdots	\vdots	\vdots	\ddots

Table 5.1: Routh Array

coefficients of the polynomial $D(s)$. We start the calculation from the third row. For each element in the table, it needs the elements in the first column and the next column from two previous rows. Then, the four elements form a matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The element can be calculated as

$$\frac{bc - ad}{c}.$$

For example, for $i = 1, 2, 3, \dots$,

$$b_i = \frac{a_{n-1}a_{n-2i} - a_n a_{n-(2i+1)}}{a_{n-1}}$$

and

$$c_i = \frac{b_1 a_{n-(2i+1)} - b_{i+1} a_{n-1}}{b_1}.$$

The calculations will continue until Row $n + 1$. When completed, the system is stable if and only if the elements of the first column of the Routh array are all positive. The number of sign changes in the first column is the number of poles in the RHP. For example, $(+ - +)$ is counted as two sign changes. Therefore, there are two positive poles in the RHP.

5.4 Dynamic Response of an LTI System

The dynamic response of an LTI system is the output signal of the system when an input signal is applied to the system. Commonly used control input signals include impulse signal, sinusoid signal, and polynomial inputs

(step signal, ramp signal, parabola, etc.). When the inputs are impulse and step signals, the outputs are called *impulse response* and *step response*, respectively. When the input is a sinusoidal signal, the output in the steady state is called *frequency response*.

5.4.1 Test Signals

Step signal is a very common test input signal to control systems. Whenever a system is activated, a step signal is applied. For example, a constant speed is commanded in vehicle cruise control. It is frequently done by adding a step function to the input of the system. In addition, step response displays important characteristics of the system. Sinusoidal signal is another common input signal to control systems.

A system dynamic response is a composite response of transient response and steady-state response, that is,

$$y(t) = y_{tr}(t) + y_{ss}(t),$$

where the transient response $y_{tr}(t)$ approaches to zero when t is very large, and the steady-state value $y_{ss}(t)$ when t is very large.

5.5 Steady State Performance

5.5.1 Frequency Response

The steady state response of a system for an input sinusoidal signal is known as the frequency response.

The input signal is a sinusoidal function $u(t) = A \cos(\omega t)$. The output is the convolution integral of the input and the impulse response of the system

$$y(t) = A \int_0^t h(\tau) \cos(\omega(t - \tau)) d\tau$$

Then, the integral can be written as

$$y(t) = A \int_0^\infty h(\tau) \cos(\omega(t - \tau)) d\tau - A \int_t^\infty h(\tau) \cos(\omega(t - \tau)) d\tau$$

The first term is called the steady-state response; and the second term will decay to zero as t grows when the system is stable. Based on the principle

of superposition, we can study the steady-state response when $A = 1$. When $A \neq 1$, the steady-state response is the steady-state response when $A = 1$ scaled by A .

Using the Euler's relation:

$$u(t) = \cos(\omega t) = \frac{1}{2} (\exp(j\omega t) + \exp(-j\omega t))$$

the steady-state response is

$$y_{ss}(t) = \frac{1}{2} \int_0^\infty h(\tau) (\exp(j\omega(t - \tau))) d\tau + \frac{1}{2} \int_0^\infty h(\tau) (\exp(-j\omega(t - \tau))) d\tau$$

Based on the definition of Fourier transform, the steady-state response can be written as

$$\begin{aligned} y_{ss}(t) &= \frac{1}{2} \exp(j\omega t) H(j\omega) + \frac{1}{2} \exp(-j\omega t) H(-j\omega) \\ &= \operatorname{Re}(H(j\omega)) \cos(\omega t) - \operatorname{Im}(H(j\omega)) \sin(\omega t) \\ &= |H(j\omega)| \cos(\omega t + \angle H(j\omega)). \end{aligned}$$

The steady-state response of a stable LTI system to a sinusoidal input is a sinusoidal with the same frequency as the input and with magnitude amplified by $|H(j\omega)|$ and the phase shifted by the angle $\angle H(j\omega)$.

Frequency Response

$$y_{ss}(t) = A |H(j\omega)| \cos(\omega t + \angle H(j\omega))$$

We can plot the magnitude ratio and phase difference as a function of the input frequency. This is called “Bode plot”, which is a frequency response of a stable LTI system. It includes a Bode magnitude plot and a Bode phase plot. Bode plots are found useful in understanding the stability of negative feedback. With bode plots, it is easy to understand the principle of signal processing. Let us consider the example of low pass filters.

Example 5 *Passive Low Pass Filter* A low pass filter is a circuit that can be designed to filter out unwanted high frequencies of an electrical signal and pass signals with a frequency lower than a selected cutoff frequency.

5.5.2 Step Response

Step signal is a special case of the sinusoidal input $u(t) = A \cos(\omega t)$ when $\omega = 0$. Therefore, the steady-state response of a step signal can be found by setting $\omega = 0$ in the frequency response, that is,

$$y_{ss}(t) = A|H(0)| \cos(0t + \angle H(0)).$$

The steady-state response $y_{ss}(t)$ of a stable LTI system to a step input $u(t) = A$ is $A|H(0)| \cos(\angle H(0))$. It is not difficult to figure out that

$$|H(0)| \cos(\angle H(0)) = H(0).$$

The value of $H(0)$ is known as “DC gain”. DC gain is defined as the ratio of the steady-state output of a stable LTI system to its step inputs, that is,

$$\text{DC gain} = \lim_{s \rightarrow 0} H(s).$$

In a tracking problem, the DC gain must be one so that the tracking error in the steady-state is zero. DC gain is useful in quickly determining if the steady-state tracking error to a step signal is zero or not without using the partial fraction expansion.

Example 6 Consider an LTI system with the transfer function

$$H(s) = \frac{5}{s + 10}.$$

Find the DC gain and determine if the steady-state tracking error is zero.

We can also obtain the steady-state step response of a stable LTI system by using the partial fraction expansion. The Laplace transform of the output is $Y(s) = H(s)/s$. The output $Y(s)$ can be written as

$$Y(s) = \frac{C_0}{s} + \frac{C_1}{s - p_1} + \cdots + \frac{C_n}{s - p_n}.$$

Since the system is stable, all p_i s are in the left half of the s -plane. If the inverse Laplace transform is applied, all time functions corresponding to the terms $C_i/(s - p_i)$ for $i = 1, \dots, n$ will decay to zero, and time function of the

term C_0/s is C_0 . Therefore, the steady-state response of $y(t)$ is $y_{ss}(t) = C_0$. We can use the cover-up method to calculate C_0 , that is,

$$C_0 = \lim_{s \rightarrow 0} sY(s)$$

Therefore, we have the final value theorem of a step response of a stable LTI system.

If an LTI system is stable, then the final value of a step response A is

$$y_{ss}(t) = \lim_{s \rightarrow 0} sH(s) \frac{A}{s} = AH(0).$$

The above result is called the “final value theorem”.

5.6 Transient Performance

5.6.1 First Order Systems

The first order system without zeros can be written as

$$H(s) = \frac{b}{s + a}.$$

In order to show the physical meaning of the parameters, it is common to write the first order system without zeros as

$$H(s) = \frac{K}{\tau s + 1},$$

where K is the DC gain, and $\tau > 0$. Therefore, the pole $-1/\tau$ is in the LHP, which means the stability of the system. Here we will discuss the transient performance of first order systems in terms of impulse response and step response.

Impulse Response

The impulse response is also called the “natural response”. Rewriting the transfer function as

$$H(s) = \frac{\frac{K}{\tau}}{s + \frac{1}{\tau}},$$

the inverse Laplace transform indicates the impulse response is

$$h(t) = \frac{K}{\tau} \exp\left(-\frac{t}{\tau}\right).$$

The impulse response decays to zero for stable systems. To measure the rate of decay, the concept “time constant” is usually used for first order systems, which is defined as the time when the impulse response decays to $1/e$ of the initial value, and e is known as Euler’s number approximately equal to 2.71828. If we draw a tangent line at the initial time, the slope of the line is $-K/\tau^2$ since

$$\left. \frac{d}{dt} \frac{K}{\tau} \exp\left(-\frac{t}{\tau}\right) \right|_{t=0} = -\frac{K}{\tau^2}.$$

This straight line intersects with the $y = 0$ line at $t = \tau$. Therefore, the time constant measures the rate of decay.

Step Response

Since the Laplace transform of the unit step signal is $1/s$, the Laplace transform of the step response is

$$Y(s) = \frac{K}{s(\tau s + 1)},$$

Using the partial fraction expansion

$$Y(s) = \frac{K}{s} - \frac{K}{s + \frac{1}{\tau}},$$

and the inverse Laplace transform, the step response is

$$y(t) = K - K \exp\left(-\frac{t}{\tau}\right)$$

The first term originates in the pole of the input and is called the forced response; since this term does not decay to zero with increasing time, it is also called the steady-state response. The second term originates in the pole of the transfer function and is called the natural response; since this term decays to zero with increasing time, it is also called the transient response. The time constant in the step response can be described as the time for the response to rise to 63% of its final value. At two time constants, the response reaches 86.5% of its final value. At $t = 3\tau$, 4τ , and 5τ , the response reaches 95%, 98.2%, and 99.3%, respectively, of its final value. As we see here, the time constant is a metric of the speed of the system response. After about 5 time constants 5τ , the response remains within 1% of its final value, and it is considered that the response is at steady-state.

Example 7 Consider the transfer function

$$H(s) = \frac{5}{s + 2}$$

What is the steady-state value, and when does the unit step response reach the steady state?

What is the use of all the observations above? They can be used to identify the parameters K and τ in a transfer function when a model is not available. To do so, we can use a unit step signal to test the system, and plot the step response. Let us see this example below.

Example 8 The unit step response of a first order system is shown in Fig. 5.7. What is the DC gain K and time constant τ in the standard first order transfer function?

For step responses, it is also common to define the transient performance using settling time, overshoot, and rise time. In the following, we will give the definitions and calculate their values for first order systems.

- **Settling Time:** Time for the response to reach and stay within 1% of its final value, denoted by t_s . According to the definition, we have

$$K - K \exp\left(-\frac{t_s}{\tau}\right) = 0.99K.$$

By solving the above equation for t_s , we find the settling time

$$t_s = 2 \ln(10)\tau \approx 4.6\tau.$$

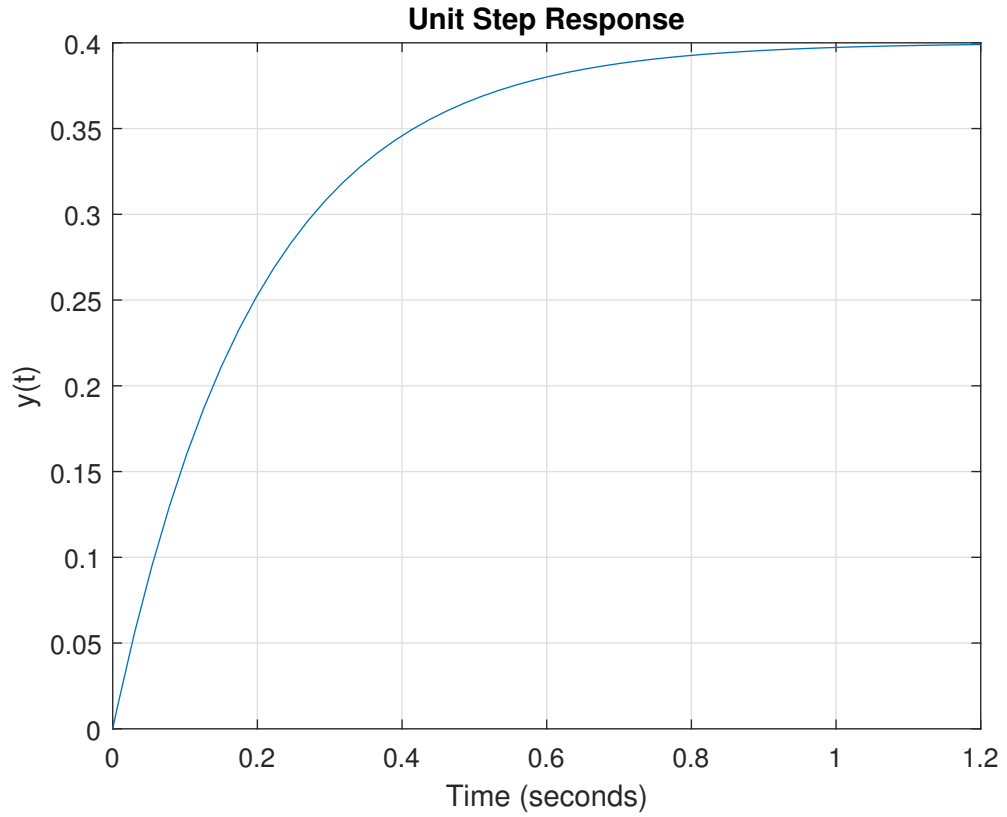


Figure 5.7: Unit Step Response Test

- **Rise Time:** Time for the response to go from 10% to 90% of its final value, denoted by t_r . According to the definition, we have $K - K \exp\left(-\frac{t_{0.9}}{\tau}\right) = 0.9K$ and $K - K \exp\left(-\frac{t_{0.1}}{\tau}\right) = 0.1K$. The rise can be found as

$$t_r = [\ln(0.9) - \ln(0.1)]\tau \approx 2.2\tau.$$

- **Overshoot:** The maximum amount the system overshoots its final value divided by its final value, denoted by PO . For first order systems, the percentage overshoot is $PO = 0\%$.

Damping ratio	System response
$\zeta = 0$	undamped
$0 < \zeta < 1$	underdamped
$\zeta = 1$	critically damped
$\zeta > 1$	over-damped

Table 5.2: Second-order system response

5.6.2 Second Order Systems

A standard second-order transfer function with unity DC gain can be written as

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}.$$

Depending the value of ζ , the system response shows different behaviors as shown in Table 5.2.

Real Roots

When $\zeta > 1$, the transfer function $H(s)$ has two distinct real roots. Therefore, it can be written in the following form

$$H(s) = \frac{\alpha_1\alpha_2}{(s + \alpha_1)(s + \alpha_2)}.$$

Let us assume $\alpha_2 > \alpha_1 > 0$. The transfer function has two poles at $-\alpha_1$ and $-\alpha_2$. A partial fraction expansion of $H(s)$ indicates that the impulse response contains two components

$$H(s) = \frac{C_1}{s + \alpha_1} + \frac{C_2}{s + \alpha_2},$$

where C_1 and C_2 are determined by the β in the numerator, and they determine the sizes of each components. The inverse Laplace transform results in the impulse response

$$h(t) = C_1 \exp(-\alpha_1 t) + C_2 \exp(-\alpha_2 t).$$

The impulse response of a second order system can be regarded as a weighted sum of the impulse responses of two first order systems. The locations of the

poles of the transfer function determine the shapes of the components of the natural response and the numerator influences the size of the coefficient that multiplies each component. This claim also applies to higher order complex systems. The component $\exp(-\alpha_2 t)$ decays faster than the component $\exp(-\alpha_1 t)$ since $\alpha_2 > \alpha_1$. Then, we can say $-\alpha_2$ is a faster pole than $-\alpha_1$. In general, a pole which is further to the left in the s -plane are associated with faster components in the impulse response than those associated with poles closer to the imaginary axis.

Example 9 Find the impulse response of the transfer function

$$H(s) = \frac{2}{s^2 + 3s + 2}.$$

The denominator is $D(s) = s^2 + 3s + 2 = (s + 1)(s + 2)$. The poles of $H(s)$ are at $s = -1$ and $s = -2$. A partial fraction expansion of $H(s)$ results in

$$H(s) = \frac{2}{s + 1} - \frac{2}{s + 2}.$$

The inverse Laplace transform gives the time function

$$h(t) = 2 \exp(-t) - 2 \exp(-2t).$$

The impulse response is shown in Fig. 5.8. At the beginning, the component $\exp(-2t)$ decays faster, and therefore the difference between the two component becomes large and large. After some time, the difference between them becomes smaller and smaller. Eventually, the overall response decays to zero as the system is stable.

Complex Roots

As in the first order system case, let us write the second order system with a pair of complex poles in the form

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

so that it is easy to identify the physical meaning of the coefficients $0 < \zeta < 1$ and $\omega_n > 0$, where ζ and ω_n are defined to be the “damping ratio” and “undamped natural frequency”, respectively. First, let us determine the

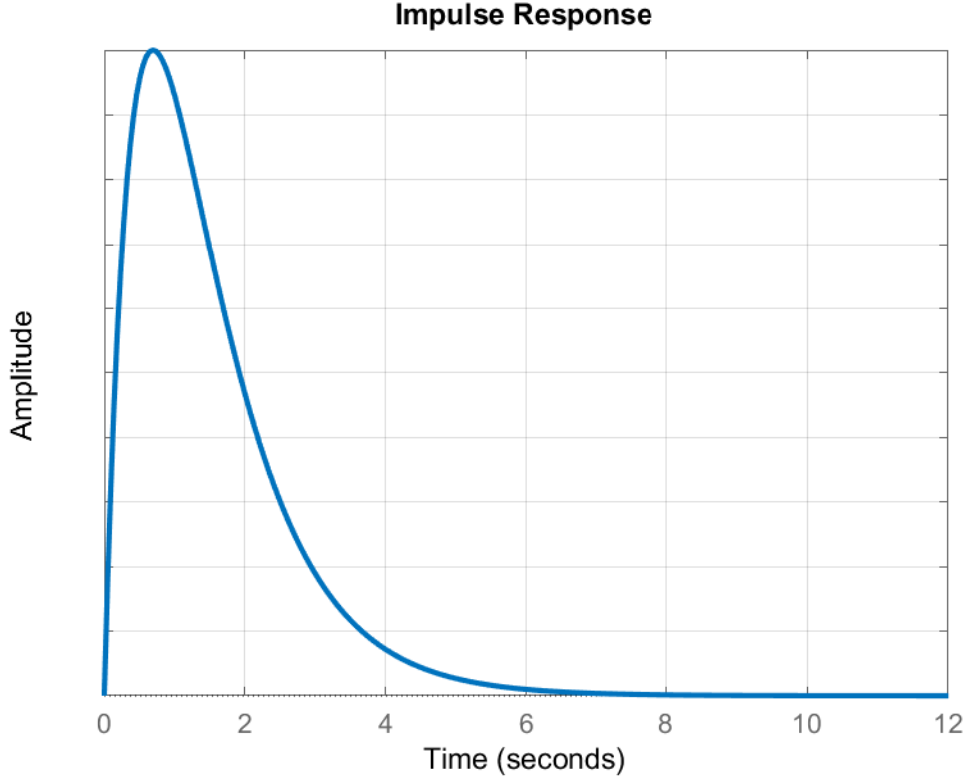


Figure 5.8: Impulse Response

location of the poles in the s -plane. Let us write the denominator of the polynomial as $s^2 + 2\zeta\omega_n s + \zeta^2\omega_n^2 + (1 - \zeta^2)\omega_n^2 = (s + \sigma)^2 + w_d^2$, where $\sigma = \zeta\omega_n$ and $\omega_d = \omega_n\sqrt{1 - \zeta^2}$. Here ω_d is called “damped natural frequency”. Let $(s + \sigma)^2 + w_d^2 = 0$. We obtain a pair of complex poles at $s = -\sigma \pm j\omega_d$, where $-\sigma$ is the real part of the complex poles, and $\pm\omega_d$ are the imaginary parts of the complex poles. Therefore, ω_n is the magnitude of the complex poles, and the phase is $\arcsin(\zeta)$ for one of the poles.

When $\zeta = 0$, the transfer function becomes

$$H(s) = \frac{\omega_n^2}{s^2 + \omega_n^2}.$$

The inverse Laplace transform shows that

$$h(t) = \frac{1}{\omega_n} \sin(\omega_n t).$$

In this case, we have a pair of imaginary poles at $\pm j\omega_n$, and the impulse response is a sinusoid with frequency ω_n . Therefore, there is no damping. That is the reason why ω_n is called “undamped natural frequency”. With the increase of ζ , the poles move away from the imaginary axis to the left. The impulse response is

$$h(t) = \frac{\omega_n}{\sqrt{1-\zeta^2}} \exp(-\sigma t) \sin(\omega_d t).$$

Even though the response will oscillate, it will decay to zero eventually. Therefore, ω_d is called “damped natural frequency”. The variable $1/\sigma$ is also defined as the time constant. After four time constants, it is generally considered that the system is at the steady-state.

The unit step response can be found by the inverse Laplace transform of $Y(s)$, that is,

$$Y(s) = H(s) \frac{1}{s}.$$

The time function of the step response is thus

$$y(t) = 1 - \exp(-\sigma t) \left(\cos(\omega_d t) + \frac{\sigma}{\omega_d} \sin(\omega_d t) \right)$$

Use the trigonometric function (2.1),

$$\cos(\omega_d t) + \frac{\sigma}{\omega_d} \sin(\omega_d t) = \sqrt{1 + \frac{\sigma^2}{\omega_d^2}} \cos \left(\omega_d t - \arctan \left(\frac{\sigma}{\omega_d} \right) \right).$$

Using the definition $\sigma = \zeta\omega_n$ and $\omega_d = \omega_n\sqrt{1-\zeta^2}$, we have

$$\cos(\omega_d t) + \frac{\sigma}{\omega_d} \sin(\omega_d t) = \frac{1}{\sqrt{1-\zeta^2}} \cos(\omega_d t - \arcsin(\zeta))$$

since

$$\arctan \left(\frac{\zeta}{\sqrt{1-\zeta^2}} \right) = \arcsin(\zeta).$$

Finally, we can obtain the step response in a different form

$$y(t) = 1 - \frac{\exp(-\sigma t)}{\sqrt{1 - \zeta^2}} \cos(\omega_d t - \arcsin(\zeta)).$$

Performance specifications for a control system design often involve certain requirements associated with the time response of the system. We can characterize some key features of the step response of second order systems used for control design. The requirements for a step response are expressed in terms of the standard quantities: rise time, overshoot, and settling time. These performance specifications have the same definitions as those of the first order system case. Because of their importance, we would like to define them again:

- Rise time (t_r): The time it takes for the response to rise from 10% to 90% of its steady-state value.

Unlike the first order system case, it is difficult to give an analytical solution of the rise time t_r in terms of the parameters ω_n and ζ . Since ζ is a value between 0 and 1, we can use the average value of $\zeta = 0.5$ and the rise time is roughly

$$t_r \approx \frac{1.8}{\omega_n}.$$

- Peak time (t_p): The time it takes for the response to reach the maximum overshoot point.

To find the peak time, let us take the time derivative of $y(t)$ and set it to be zero. The time derivative of $y(t)$ is

$$\begin{aligned} y^{(1)}(t) &= \sigma \exp(-\sigma t) \left(\cos(\omega_d t) + \frac{\sigma}{\omega_d} \sin(\omega_d t) \right) \\ &\quad + \exp(-\sigma t) (\omega_d \sin(\omega_d t) - \sigma \cos(\omega_d t)) \\ &= \exp(-\sigma t) \left(\frac{\sigma}{\omega_d} + \omega_d \right) \sin(\omega_d t) \end{aligned}$$

All the peaks points correspond to $y^{(1)}(t) = 0$, $y^{(1)}(t-) > 0$, and $y^{(1)}(t+) < 0$. Therefore, we know that $\omega_d t = k\pi$ for $k = 1, 2, \dots$ and the $\omega_d t_p = \pi$. Therefore, the peak time

$$t_p = \frac{\pi}{\omega_d}.$$

- **Overshoot (OP):** The maximum amount the system overshoots its final value divided by its final value.

The overshoot can be calculated using the peaking value $y(t_p)$ at the peak time t_p , where

$$y(t_p) = 1 - \frac{\exp(-\sigma t_p)}{\sqrt{1 - \zeta^2}} \cos(\omega_d t_p - \arcsin(\zeta)).$$

It can be calculated that

$$\exp(-\sigma t_p) = \exp\left(-\zeta \omega_n \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}}\right) = \exp\left(-\pi \frac{\zeta}{\sqrt{1 - \zeta^2}}\right),$$

and

$$\cos(\omega_d t_p - \arcsin(\zeta)) = \cos\left(\omega_d \frac{\pi}{\omega_d} - \arcsin(\zeta)\right) = -\sqrt{1 - \zeta^2}.$$

Then, we have

$$y(t_p) = 1 + \exp\left(-\pi \frac{\zeta}{\sqrt{1 - \zeta^2}}\right).$$

Since the DC gain $H(0)$ is 1. The final value of $y(t)$ is 1. Therefore, we have

$$OP = 100 \exp\left(-\pi \frac{\zeta}{\sqrt{1 - \zeta^2}}\right)\%.$$

- **Settling time (t_s):** The time it takes for the response to reach 1% of the steady-state value and stay within the 1% range.

This means that $0.99 < y(t) < 1.01$ for $t \geq t_s$, that is,

$$0.99 < 1 - \frac{\exp(-\sigma t)}{\sqrt{1 - \zeta^2}} \cos(\omega_d t - \arcsin(\zeta)) < 1.01.$$

This is equivalent to say that

$$\left| \frac{\exp(-\sigma t)}{\sqrt{1 - \zeta^2}} \cos(\omega_d t - \arcsin(\zeta)) \right| < 0.01$$

To obtain an analytical expression for the setting time in which

$$\left| \frac{\exp(-\sigma t_s)}{\sqrt{1-\zeta^2}} \cos(\omega_d t - \arcsin(\zeta)) \right| \leq 0.01$$

is really hard. Since the response is bounded by the envelop determined by the exponential term $\exp(-\sigma t)$, we then borrow the concept of time constant. According to the definition of time-constant,

$$\tau = \frac{1}{\sigma}.$$

Similar to the first order case, to decay within the 1% range of steady-state value, the settling time is

$$t_s \approx \frac{4.6}{\sigma}.$$

5.6.3 Higher Order Systems

Effects of Zeros

Now let us assume we have a transfer function $H_1(s)$ and the corresponding response $Y_1(s) = H_1(s)U(s)$ is $y(t)$ for an input $U(s)$. Now let us consider another transfer function

$$H_2(s) = \frac{s+z}{z} H_1(s).$$

The transfer function has an additional zero compared with the transfer function $H_1(s)$. The output of the second system with the same input is $Y_2(s) = H_2(s)U(s)$, where $Y_2(s)$ can be written as

$$Y_2(s) = H_1(s)U(s) + \frac{s}{z} Y_1(s).$$

By taking the inverse Laplace transform and using the property (2.2), we have

$$y_2(t) = y_1(t) + \frac{1}{z} y_1^{(1)}(t)$$

by assuming that $y_1(0) = 0$.

When there is an extra zero, the response consists of two parts: the original response (the response without an extra zero) and the weighted derivative

of the original response. When z is very large, the zero has little effect to the original response. When z is small, the zero has substantial influence to the original response.

For step responses, the derivative is typically positive at the start of a step response. The positive part will be added to the original step response. Therefore, the major effect of zeros to step response is to increase overshoot and reduce rise time.

It is interesting when the zeros are at the RHP. Rational transfer functions are called *minimum phase* if their numerator polynomial is Hurwitz. Zeros at RHP are also called “nonminimum-phase zeros”. The major effect of nonminimum-phase zeros is the undershoot. The step response will start towards an opposite direction from the reference input. This may make the response slow.

Effects of Poles

Additional poles may increase the rise time.

Chapter 6

Design of Control Systems

The learning outcomes of this chapter are to

- Find steady-state error for a unity-feedback system;
- Be aware of the central role of error signals in design of control systems;
- Recognize the improvements afforded by feedback control in reducing system sensitivity to parameter changes, disturbance rejection, and measurement noise attenuation;
- Be familiar with the PID controller as a key element of many feedback systems.

This chapter focuses on the controller design in the closed-loop system. A good control design should include but not limited to the following desired characteristics:

- **Stability:** The control system must be stable all the time. It is a basic and absolute requirement.
- **Tracking:** The system output should be as close as possible to the desired input signals, including both the desired transient performance and a small steady-state tracking error.
- **Disturbance rejection and noise attenuation:** All systems have unwanted disturbance inputs, which are usually constant signals, such as bias. The system should not respond significantly to the disturbance input. The control design should have the ability to reject the effects

of a constant disturbance. Also the sensing noises are inevitable. A good control design should attenuate the effect of sensing noises.

- **Robustness:** It is almost impossible to obtain an accurate model of a plant. The characteristics of the control system should be insensitive to the parameter changes in the plant due to environmental changes, such as temperature, humidity, altitude. The control performance should not be compromised even if the design is based on inaccurate parameters in the plant.

Most developments in the following are based on the unity-feedback structure. Most simple feedback systems can be transformed into the unity feedback system as shown in Fig. 6.1. The plant $G(s)$ includes the physical process, power amplifiers, actuators, gears and so on. The variable $Y(s)$ is known as the controlled variable or the system output. It is the one we want to control. The variable $R(s)$ in the unity feedback is the desired value of the system output. The variable $U(s)$ is the control signal. The difference between them is the system error $E(s) = R(s) - Y(s)$. The unit feedback can be used when the sensor can be modeled as a pure gain.

6.1 What is a good control design?

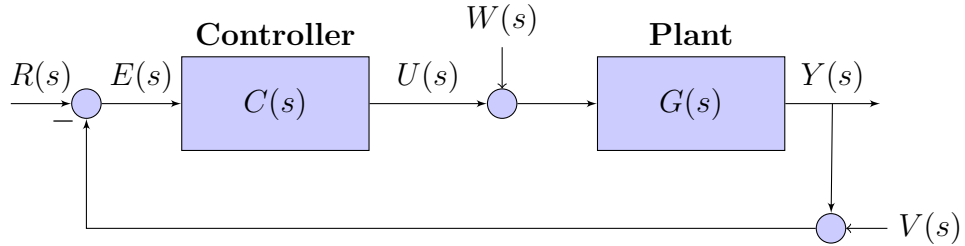


Figure 6.1: Unity Feedback

Let us first derive the basic transfer functions from various inputs to the output. Define the *sensitivity function* and *complementary sensitivity function* as

$$\mathcal{S} = \frac{1}{1 + C(s)G(s)}$$

and

$$\mathcal{T} = \frac{C(s)G(s)}{1 + C(s)G(s)},$$

respectively. It is easy to see that $\mathcal{S} + \mathcal{T} = 1$.

The transfer function from the reference input $R(s)$ to the output $Y(s)$ is

$$H_r(s) = \mathcal{T},$$

the transfer function from the disturbance input $W(s)$ to the output $Y(s)$ is

$$H_w(s) = G(s)\mathcal{S},$$

and the transfer function from the sensor noise $V(s)$ to the output $Y(s)$ is

$$H_v(s) = -\mathcal{T}.$$

Based on the principle of superposition,

$$Y(s) = \mathcal{T}R(s) + \mathcal{S}G(s)W(s) - \mathcal{T}V(s).$$

The error signal is

$$E(s) = R(s) - Y(s) = \mathcal{S}R(s) - \mathcal{S}G(s)W(s) + \mathcal{T}V(s).$$

6.1.1 Stability

Let us assume that the controller $C(s)$ and the plant $G(s)$ can be written as

$$C(s) = \frac{c(s)}{d(s)}$$

and

$$G(s) = \frac{b(s)}{a(s)},$$

respectively. The characteristic equation is

$$a(s)d(s) + b(s)c(s) = 0.$$

Even though the plant $G(s)$ may be unstable, we can design the controller by choosing $c(s)$ and $d(s)$ to make the closed-loop system stable.

6.1.2 Tracking

The tracking problem is to design the controller $C(s)$ so that the output $Y(s)$ follows the input $R(s)$ as close as possible. To make the steady-state tracking error zero when the input is a step signal, we can verify the DC gain to see if $\mathcal{T}(0) = 1$ or $\mathcal{S}(0) = 0$. For other types of inputs, more details will be given later.

6.1.3 Regulation

The regulation problem is to track a constant reference input as close as possible regardless of the disturbance inputs and sensing noises. From the basic equation, we know

$$E(s) = \mathcal{S}R(s) - \mathcal{S}G(s)W(s) + \mathcal{T}V(s).$$

To reduce the effects of disturbances and noises, we need the \mathcal{S} and \mathcal{T} to be “small” in some sense. It seems to be a conflicting requirement to make both the sensitivity and the complementary sensitivity function “small” at the same time. To make the sensitivity function “small”, we need to design a controller $C(s)$ so that $C(s)G(s)$ is “large”. However, when $C(s)G(s)$ is “large”, the complementary sensitivity function \mathcal{T} is close to 1. When $C(s)G(s)$ is “small”, the complementary sensitivity function \mathcal{T} is “small”. However, the sensitivity function \mathcal{S} is close to 1. The solution to this dilemma is to explore the frequency property of disturbances and noises. The plant disturbances occur at very low frequencies, such as a bias. But the sensing noises occur often at high frequencies. Therefore, we could design the controller $C(s)$ so that the magnitude $|C(j\omega)|$ is large at the low frequency and small at the high frequency.

Indeed, tragically, an experiment X-15 aircraft flying this control system crashed in 1966 owing partially to the saturation problem occurring in the high-gain scheme.

6.1.4 Robustness

It is important that the controller parameters are chosen in such a way that the closed-loop system is not sensitive to variations in process dynamics. Parameter changes due to heat or other causes should not appreciably affect the

performance of a system. The general topic of system characteristics changing with system parameter variations is called “sensitivity”. The sensitivity of a transfer function G to one of its parameters K is defined as the ratio of percentage change in G to percentage change in K as

$$\mathcal{S}_K^G = \frac{dG/G}{dK/K} = \frac{K}{G} \frac{dG}{dK}.$$

First, let us calculate the sensitivity of the closed-loop transfer function $T(s)$ to the plant transfer function $G(s)$. The closed-loop transfer function is

$$T(s) = \frac{C(s)G(s)}{1 + C(s)G(s)}.$$

According to the definition,

$$\mathcal{S}_G^T = \frac{G}{T} \frac{dT}{dG} = \frac{G}{\frac{CG}{1+CG}} \frac{(1+CG)C - C^2G}{(1+CG)^2} = \frac{1}{1 + G(s)C(s)}.$$

This is exactly how sensitivity function is defined, and it is the transfer function from the reference input to the system error.

Many plants by their very nature, have parameters that vary extensively during operation. An example is an aircraft, whose parameters vary over wide ranges with speed, altitude, and so on. Hence aircraft autopilot designs must satisfy rigorous specifications for sensitivity to plant variations.

6.2 Tracking Problems

6.2.1 Test Signals

The common test signals are step signals, ramp signals, and parabolic inputs shown in Fig. 6.2. The test signal can be mathematically written as polynomial inputs as

$$r(t) = \frac{t^k}{k!},$$

where k is the order of the polynomial input, t is the time variable, and $k!$ is the factorial of k .

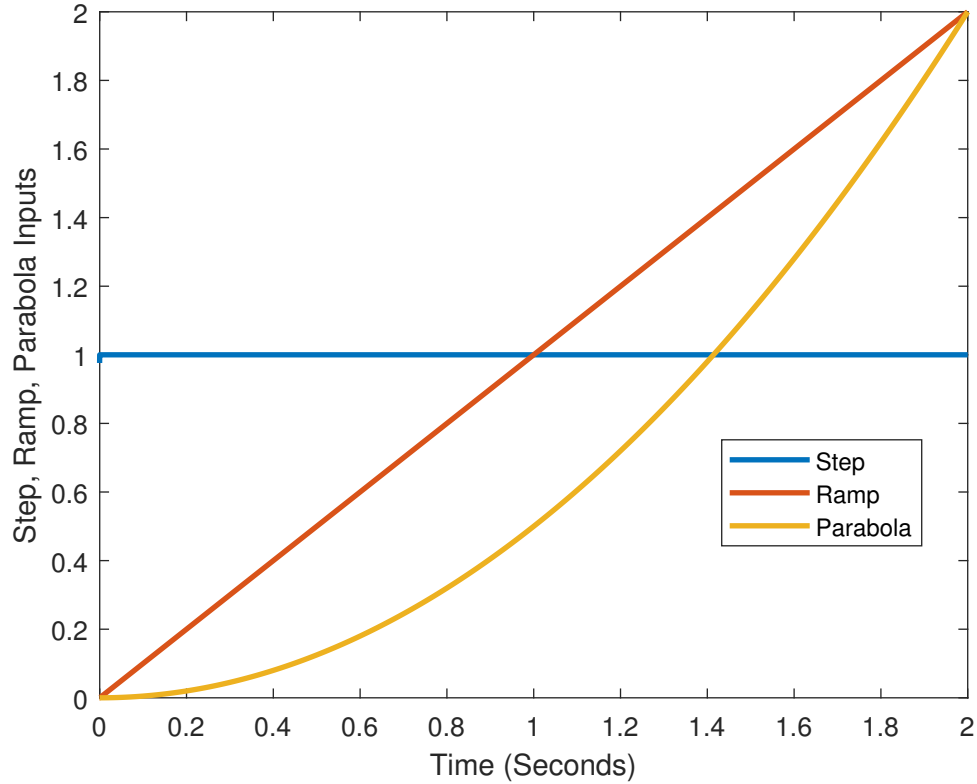


Figure 6.2: Test Signals

6.2.2 System Types for Reference Input

Control systems can be classified according to their ability to track polynomial inputs.

In general, the units of the input and the output are different for a non-unity feedback system. Typically, voltage is the unit of the physical system input at the summer. The output of a physical system could be position, temperature, and angles. A sensor will convert a physical variable into an electrical signal such that the input and the sensor output are in the same units. Therefore, the sensor may have its own dynamics instead of a pure gain.

According to the block diagram shown in Fig. 6.3, the tracking error can

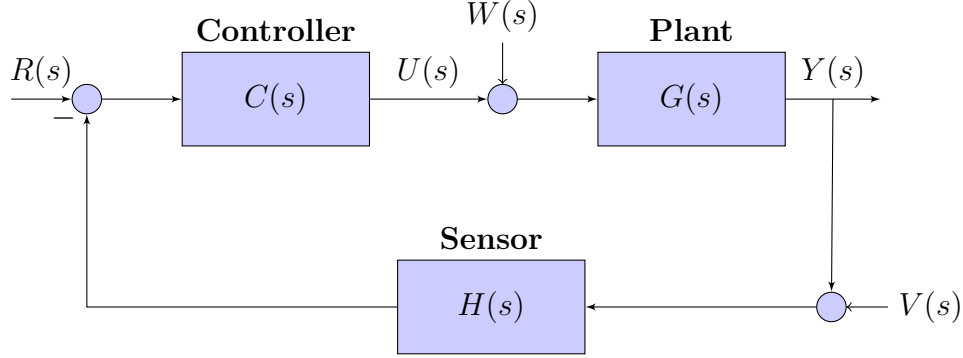


Figure 6.3: Feedback Control System with Sensor Dynamics

be written as

$$E(s) = R(s) - Y(s).$$

Let us assume that $W(s) = 0$ and $V(s) = 0$. The transfer function from $R(s)$ to $Y(s)$

$$T(s) = \frac{Y(s)}{R(s)} = \frac{CG}{1 + CGH}$$

implies that $E(s) = R(s) - H(s)R(s) = (1 - T(s))R(s)$. The Laplace transform of a polynomial with degree k , that is, $r(t) = t^k/k!$, is

$$R(s) = \frac{1}{s^{k+1}}.$$

The system is type n if it can track a polynomial of degree n with a constant error. It turns out that a system is type n if $1 - H(s)$ can be written as $1 - T(s) = s^n P(s)$ and $P(0) \neq 0$ is a finite constant. In other words, the number of zeros of the transfer function from the reference input to the system error determines the type of the system for reference input. To find the steady-state tracking error with the assumption that the closed-loop system is stable, we can use the final value theorem. The steady-state tracking error e_{ss} is

$$e_{ss} = \lim_{t \rightarrow \infty} e(t) = \lim_{s \rightarrow 0} s(1 - H(s)) \frac{1}{s^{k+1}} = \lim_{s \rightarrow 0} \frac{1 - T(s)}{s^k} = \lim_{s \rightarrow 0} \frac{s^n}{s^k} P(s).$$

When $n > k$, $e_{ss} = 0$. When $n < k$, $e_{ss} = \infty$. When $n = k$, $e_{ss} = P(0)$.

When the closed-loop control system is a unity feedback system, there is a simple way to determine the system type by counting the number of poles of the forward path gain $C(s)G(s)$. Let us write $C(s)G(s)$ as the form $\frac{Q(s)}{s^n}$ and $Q(0) \neq 0$ is a finite constant. The constant $Q(0)$ is called the error constant to quantify the tracking error. It turns out the system is type n . Let us assume the reference input is polynomial of degree n . Then the Laplace transform is $R(s) = 1/s^{n+1}$. The tracking error is

$$e_{ss} = \lim_{s \rightarrow 0} s \frac{1}{s^{n+1}} \frac{1}{1 + \frac{Q(s)}{s^n}} = \lim_{s \rightarrow 0} \frac{1}{s^n} \frac{s^n}{s^n + Q(s)}.$$

For the step response, $n = 0$,

$$e_{ss} = \frac{1}{1 + Q(0)},$$

and $Q(0)$ is called the position error. For the ramp input, $n = 1$,

$$e_{ss} = \frac{1}{Q(0)},$$

and $Q(0)$ is called the velocity constant. When $n = 2$,

$$e_{ss} = \frac{1}{Q(0)},$$

and $Q(0)$ is called the acceleration constant.

6.2.3 System Types for Disturbance Input

To study the disturbance rejection, we can assume $R(s) = 0$ because of the principle of superposition of linear systems. In this case,

$$E(s) = -Y(s) = \frac{G(s)}{1 + C(s)G(s)H(s)}W(s),$$

and the transfer function from $W(s)$ to $E(s)$ is

$$H_w(s) = -\frac{G(s)}{1 + C(s)G(s)H(s)}.$$

We can also classify control systems according to their abilities to reject disturbances. Let us assume that the disturbance is the form of polynomials, that is,

$$W(s) = \frac{t^k}{k!}.$$

A control system is called type k for the disturbance input if k is the highest order of the polynomials that the control system can reject with a constant non-zero steady-state error. We can use the final value theorem to calculate the steady-state error

$$e_{ss} = \lim_{t \rightarrow \infty} e(t) = \lim_{s \rightarrow 0} sH_w(s) \frac{1}{s^{k+1}} = \lim_{s \rightarrow 0} \frac{H_w(s)}{s^k}.$$

When the disturbance is a constant, then $k = 0$. In this case,

$$e_{ss} = \lim_{s \rightarrow 0} H_w(0).$$

Typical examples include constant wind against an antenna control system, or the constant heat loss in a temperature control system. The system is type 0 if $H_w(0) \neq 0$ is a constant. When the disturbance is a ramp function, then

$$e_{ss} = \lim_{s \rightarrow 0} \frac{H_w(s)}{s}.$$

The system is called type 1 when

$$\lim_{s \rightarrow 0} \frac{H_w(s)}{s}$$

is a nonzero constant. In general, a system type to disturbance input can be defined according to the number of zeros at $s = 0$ of the transfer function $T_w(s)$. Suppose that a system is type n . If $n > k$ with k being the degree of the disturbance input, then $e_{ss} = 0$; if $n < k$, the error is unbounded; when $n = k$, the error is a nonzero constant.

6.3 PID Controller

The PID controller has three terms in the form of

$$D(s) = K_P + \frac{K_I}{s} + K_D s,$$

where K_P is the proportional gain, K_I is the integral gain, and K_D is the derivative gain.

6.3.1 P Controller

The proportional term can be used alone, and the controller is called “P-Controller”. The output of the P-controller is proportional to the error signal $e(t) = r(t) - y(t)$, and the output is $u(t) = K_P e(t)$. The transfer function of the P-Controller is a gain K_P . Assume that the plant has the transfer function

$$G(s) = \frac{b(s)}{a(s)},$$

and the system is in a unity feedback structure. The characteristic equation is $a(s) + K_P b(s) = 0$. The P-Controller will not change the type of the system. Assume that the plant is type 0, then the steady-state tracking error is $e_{ss} = \frac{1}{1+K_P G(0)}$. It can be seen that the larger the value of K_P , the smaller the steady-state tracking error.

Consider the standard first order system

$$G(s) = \frac{K}{\tau s + 1}$$

and second order system

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}.$$

The type of the system are 0 and 1 for the first order system and the second order system, respectively. The steady-state tracking errors for a unit step signal are

$$e_{ss} = \frac{1}{1 + K K_P}$$

and

$$e_{ss} = \frac{1}{1 + K_P},$$

respectively. We can conclude that the larger the value of K_P , the smaller the steady-state tracking error. For the transient performance, the closed-loop transfer function for first order plant is

$$H(s) = \frac{K K_P}{\tau s + 1 + K K_P} = \frac{\frac{K K_P}{1 + K K_P}}{\frac{\tau}{K K_P} s + 1}.$$

We can see that the larger the value of K_P , the smaller the time constant, and the better the transient performance. For second-order systems, the characteristic equation is

$$s^2 + 2\zeta\omega_n s + \omega_n^2 + K_P\omega_n^2.$$

The natural frequency for the closed-loop system becomes $\sqrt{K_P(1 + \omega_n^2)}$ and the damping ratio becomes

$$\zeta = \frac{\omega_n}{\sqrt{K_P(1 + \omega_n^2)}}.$$

Even though a large K_P will reduce the steady-state tracking error, it also reduces the damping ratio. Therefore, the transient response will have a large overshoot and a reduced rise time for a large K_P .

The analysis for the disturbance input is similar. One way to reduce the steady-state tracking error without using a very large gain is to introduce the “I-controller”, which will be discussed below.

6.3.2 I Controller

The integral controller has the form

$$u(t) = K_I \int_0^t e(\tau) d\tau,$$

and the Laplace transform of the “I controller” is

$$D(s) = \frac{K_I}{s}.$$

The output of the I controller is linearly proportional to the integral of system errors. The I controller has its own dynamics with one integrator. The purpose of the I controller is to reduce the steady-state errors, and improve the steady-state output response to disturbances.

Assume that the plant has the transfer function

$$G(s) = \frac{b(s)}{a(s)},$$

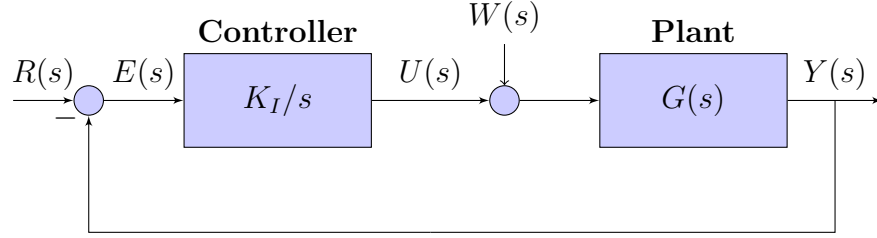


Figure 6.4: Unity Feedback

and the system is in a unity feedback structure. Then, the characteristic equation of the closed-loop system is $sa(s) + K_I b(s) = 0$. For first order systems, the characteristic equation is

$$\tau s^2 + s + K_I K = 0.$$

We can adjust K_I to improve the dynamic response. For second order systems, the characteristic equation is

$$s^3 + 2\zeta\omega_n s^2 + \omega_n^2 s + K_I \omega_n = 0.$$

Based on Routh array shown in Table 6.1, we have

Row 1	1	ω_n^2
Row 2	$2\zeta\omega_n$	$K_I \omega_n$
Row 3	b_1	0
Row 4	$K_I \omega_n$	

Table 6.1: Routh Array

$$b_1 = \frac{2\zeta\omega_n^2 - K_I \omega_n}{2\zeta\omega_n} = \frac{2\zeta\omega_n - K_I}{2\zeta}.$$

It is easy to see that the system may be unstable if K_I is too large.

The forward path transfer function is $G(s)K_I/s$. Therefore, the I-controller increases the system type by 1. For first order systems,

$$G(s) \frac{K_I}{s} = \frac{K K_I}{s(\tau s + 1)}$$

the closed-loop system is type 1 with respect to reference inputs. It can track step reference inputs with zero steady-state error. The same claim applies to standard second-order systems. For unit step response, the transfer function from the step reference input to the control input of the standard first and second order systems is

$$\frac{U(s)}{R(s)} = \frac{K_I/s}{1 + K_I G(s)/s} = \frac{K_I}{s + K_I G(s)}.$$

In the steady state,

$$u_{ss} = \lim_{s \rightarrow 0} \frac{K_I}{s + K_I G(s)} = \frac{1}{G(0)}.$$

It makes sense that the steady-state control is $1/G(0)$ since the product of the control signal and the DC gain of the process is 1, which is the output at the steady state.

The transfer function from the disturbance input to the error is

$$\frac{E(s)}{W(s)} = \frac{-G(s)}{1 + G(s)K_I/s} = -\frac{sG(s)}{s + G(s)K_I}.$$

For standard first and second order processes, the system is type 1 for the disturbance input. The system has zero steady-state output error for step disturbance inputs. The transfer function from the disturbance input to the control signal is

$$\frac{U(s)}{W(s)} = \frac{-G(s)K_I/s}{1 + G(s)K_I/s} = -\frac{G(s)K_I}{s + G(s)K_I}.$$

When the disturbance is a unit step signal, the steady-state control signal is

$$u_{ss} = \lim_{s \rightarrow 0} -\frac{G(s)K_I}{s + G(s)K_I} = -1.$$

It makes sense that the sum of the control signal and the disturbance is zero so that the disturbance has no influence on the output.

As a result, integral feedback in this case results in zero steady-state error for both reference and disturbance inputs. Furthermore, plant parameter changes can be tolerated; that is, the results above are independent of the plant parameter values. Also, regardless of the integral gain K_I , the asymptotical tracking, and disturbance rejection properties are preserved, provided that the closed-loop system remains stable. These properties of integral control are referred to as robust.

6.3.3 D Controller

The derivative feedback is also called rate feedback. It has the form

$$u(t) = K_D e^{(1)}(t)$$

in the time domain. The control signal is proportional to the derivative of the system error. The Laplace transform of the D controller is $U(s) = K_D s E(s)$. The D-control cannot be used alone since it does not provide the necessary control signal at the steady-state since the derivative of steady-state signal is always zero. The goal of the D controller is to improve closed-loop stability, speed up transient response, and reduce overshoot. The disadvantage of D controller is that it amplifies noise.

6.3.4 PID Controller

The PID control contains all three terms. In the time domain, the controller in the parallel form is

$$u(t) = K_P e(t) + K_I \int_{t_0}^t e(\tau) d\tau + K_D e^{(1)}(t).$$

By properly tuning all the three parameters, a PID controller has all the advantages of each individual term.

6.4 PID Controller Tuning

The PID controller can be written in two different forms: parallel form and standard form, as

$$D(s) = K_P + \frac{K_I}{s} + K_D s$$

and

$$D(s) = K_P \left(1 + \frac{1}{T_I s} + T_D s \right),$$

respectively.

6.4.1 The Step Response Method

The step response approach is to test a process using a step signal in the open loop. We assume no knowledge of the plant or process information. This method can be viewed as a traditional method based on modeling and control where a very simple process model is used. The step response is characterized by a first order plus time delay model in the form of

$$H(s) = \frac{A \exp(-\tau_d s)}{\tau s + 1},$$

as shown in Fig. 6.5, where τ_d is a time delay, A is the DC gain, and τ is the time constant. The DC gain A can be determined by the steady-state value

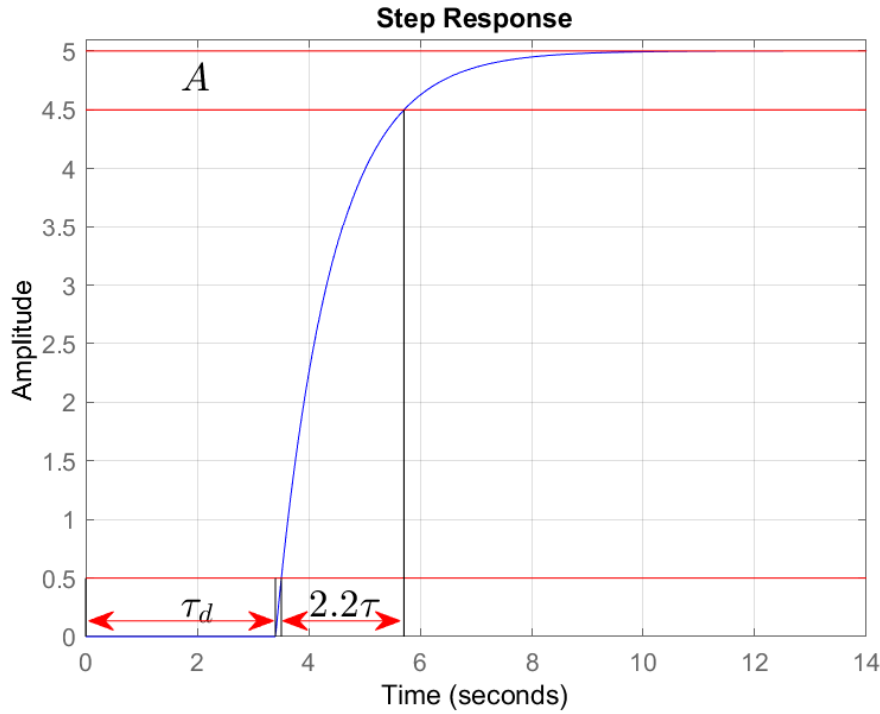


Figure 6.5: Ziegler-Nichols Tuning

of the unit step response. The time delay is from the time when the input is applied until the output starts to respond to the input. The time constant

can be calculated using the concept of rise time, 2.2τ , which is the time for the output to raise from 10% to 90% of its final value. Then, we calculate the reaction rate

$$R = \frac{A}{\tau}.$$

Based on the value of $L = \tau_d$ and R , the parameters of the PID controller is given in Table 6.2.

Controller Type	Controller Parameters
P	$K_P = \frac{1}{RL}$
PI	$K_P = \frac{0.9}{RL}, T_I = \frac{L}{0.3}$
PID	$K_P = \frac{1.2}{RL}, T_I = 2L, T_D = 0.5L$

Table 6.2: Ziegler-Nichols Tuning

6.4.2 The Frequency Response Method

First, we connect the controller with the plant in a unity feedback arrangement, and the proportional gain is set to be very small. Note that we have no knowledge of the transfer function of the plant $G(s)$. Then a short pulse is applied to test the system. We gradually increase the proportional gain K_P until the onset of instability, that is, the step response of the closed-loop system shows the behavior of undamped oscillation. The minimum gain K_P to produce the undamped oscillation behavior is called the ultimate gain denoted by K_u , and the period of oscillation is called the ultimate period denoted by P_u . The ultimate period can be obtained by the time between successive oscillation peaks. When performing such a test on a process loop, it is important to ensure the oscillation peaks do not reach the limits of the saturation of the actuator. In other words, in order for the oscillation to accurately reveal the process characteristics of the amplitude and frequency when the closed-loop system is neutrally stable, the oscillations must be naturally limited and not artificially limited by the saturation of the actuator.

The suggested values of the PID controller are given in Table 6.3.

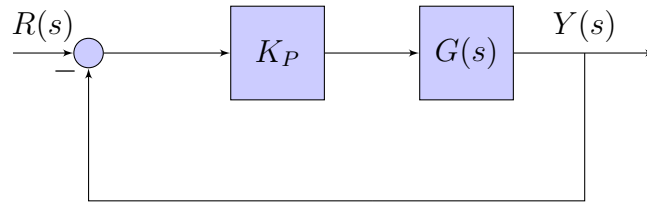


Figure 6.6: Determination of ultimate gain and period

Controller Type	Controller Parameters
P	$K_P = 0.5K_u$
PI	$K_P = 0.45K_u, T_I = \frac{P_u}{1.2}$
PID	$K_P = 1.6K_u, T_I = 0.5P_u, T_D = 0.125P_u$

Table 6.3: Ziegler-Nichols Tuning

Review Questions

- 6.1 What does the P, I and D stand for in PID Controller.
- 6.2 List advantages and disadvantages of feedback in control.
- 6.3 What makes a good control design?
- 6.4 What is the main objective of introducing integral control?

Part III

Computer Aid Tools

Chapter 7

Matlab

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python. By its nature, this note cannot cover all the background and details necessary for a complete understanding of MATLAB. It should give you enough information to apply MATLAB to the analysis and design of control systems. For further details, you are referred to other sources, including MATLAB reference manuals.

Free training courses on Matlab can be found at <https://matlabacademy.mathworks.com/>. When you're done, you can download a shareable certificate of completion. Please finish

1. *Getting Started*
 - MATLAB Onramp;
 - Simulink Onramp;
 - Control Design Onramp with Simulink.
2. *MATLAB*
 - MATLAB Fundamentals;
 - MATLAB Programming Techniques.
3. *Simulink*

- Simulink Fundamentals.

4. *Math and Optimization:*

- Introduction to Symbolic Math with MATLAB;
- Solving Nonlinear Equations with MATLAB;
- Introduction to Linear Algebra with MATLAB.

As you work in MATLAB, you issue commands that create variables and call functions. For example, create a matrix variable named *A* by typing this statement in at the command line:

```
1 A = [1 2; 4 6]
```

7.1 Control System Toolbox

Control System Toolbox provides algorithms and apps for systematically analyzing, designing, and tuning linear control systems. You can specify a system as a transfer function in zero-pole-gain representations. Apps and functions, such as step response plot, let you analyze and visualize system behavior in the time domain. You can tune PID controllers. You can validate your design by verifying rise time, overshoot, settling time and other requirements.

7.1.1 Test Signals

```
1 t = (-1:0.01:1)';  
2 unitstep = t ≥ 0;  
3 ramp = t.*unitstep;  
4 quad = t.^2.*unitstep/2;  
5 plot(t,[unitstep ramp quad])
```

7.1.2 Model Interconnections

The control system toolbox provides functions to construct the transfer function of a complex system by using the interconnection of simple subsystems. The concept is similar to the block diagram reduction.

```

1 H1 = tf(2,[1 3 0])
2 H2 = zpk([],-5,5)
3 % Series Connection
4 H = H2 * H1
5 % or equivalently
6 H = series(H1,H2);
7 % Parallel Connection
8 H = H1 + H2
9 % or equivalently
10 H = parallel(H1,H2);
11 % Feedback Connections
12 H = feedback(H1,H2)
13 %Note that feedback assumes negative feedback by default. ...
    To apply positive feedback, use the following syntax:
14 H = feedback(H1,H2,+1);

```

7.1.3 PID Tuner

```

1 s=tf('s');
2 sysp=1/(600*s^2+70*s+1);
3 systd = set(sysp,'InputDelay', 5);
4 pidTuner(systd,'P')

```

The command `pidTuner(sys,type)` launches the PID Tuner app and designs a controller of type **P** for plant **sys**. The PID Tuner app is shown in Fig. 7.1. In the app, the type of the controller can be changed. Sliding the two bars to adjust the response time from slower to faster and transient behavior from aggressive to robust.

7.2 Symbolic Math Toolbox

Symbolic Math Toolbox provides functions for calculus of polynomials since transfer functions can be written as the ratio of two polynomials.

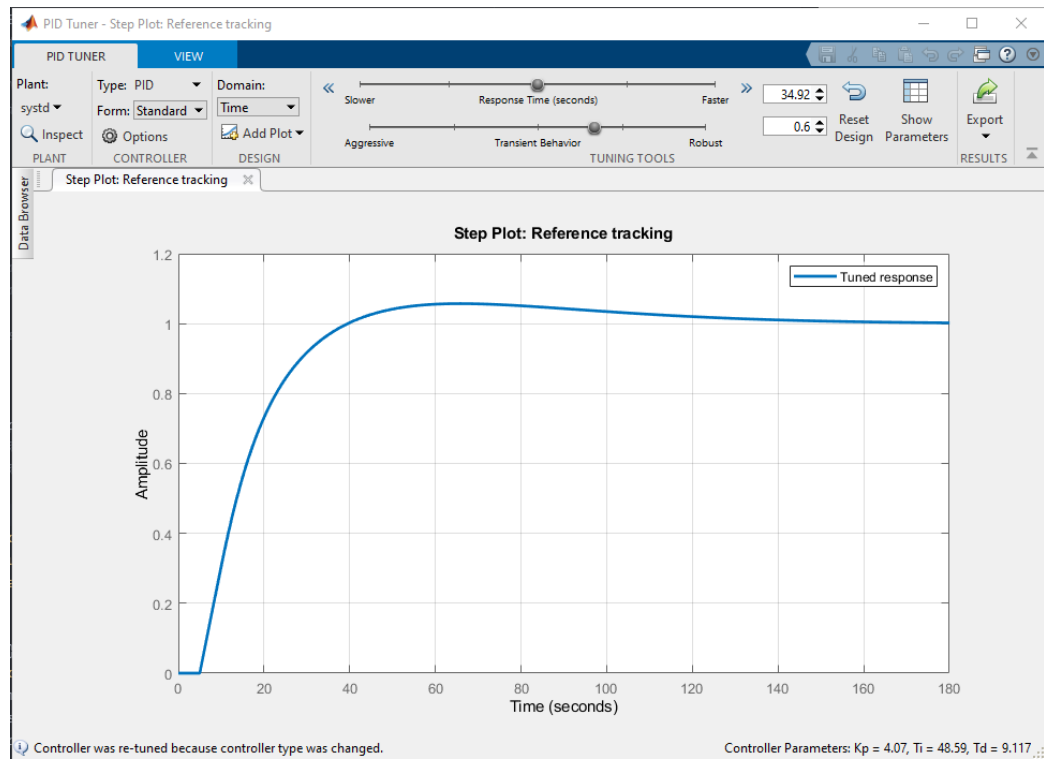


Figure 7.1: PID Tuner App

```

1 % Symbolic Toolbox
2 syms s m1 m2 b ks kw R
3 % Polynomials
4 f1 = s^2+s*b/m1+ks/m1+kw/m1;
5 f2 = -s*b/m1-ks/m1;
6 f3 = -s*b/m2-ks/m2;
7 f4 = s^2+ s*b/m2+ks/m2
8 f5 = kw*R/m1;
9
10 %Multiplication of polynomials
11 den = collect(f1*f4-f2*f3,s)
12 den = simplify(den)
13 den = collect(den,s)
14
15 num = collect(-f5*f3,s)

```

Symbolic Math Toolbox also provides functions for matrix calculus.

```

1 % Define real variables
2 syms h r real
3
4 % Define matrix Variables
5 P = sym('P', [3 3], 'real'); % symbolic real matrix
6
7 % Define symmetric matrix variables
8 P = triu(P,0) + triu(P,1).'; % make P symmetric
9
10 Q = sym('Q', [3 3], 'real');
11 S = sym('S', [3 1], 'real');
12
13 A = [1 h h^2/2; 0 1 h; 0 0 1];
14 Gamma = eye(3);
15 C = [1 0 0];
16
17 K = Gamma*S/r;
18
19 Pk_k_1 = (A-K*C)*P*(A-K*C)' + Gamma*Q*Gamma' - K*r*K';
20
21 a = expand(Pk_k_1(1,1));
22
23 pretty(a);

```

7.3 Simscape

Simscape enables you to rapidly create models of physical systems within the Simulink environment. With Simscape, you build physical component models based on physical connections that directly integrate with block diagrams and other modeling paradigms. You model systems such as electric motors, bridge rectifiers, hydraulic actuators, and refrigeration systems, by assembling fundamental components into a schematic. Simscape add-on products provide more complex components and analysis capabilities.

Simscape helps you develop control systems and test system-level performance. You can create custom component models using the MATLAB based Simscape language, which enables text-based authoring of physical modeling components, domains, and libraries. You can parameterize your models using MATLAB variables and expressions, and design control systems for your physical system in Simulink. To deploy your models to other simulation environments, including hardware-in-the-loop (HIL) systems, Simscape

supports C-code generation.

7.3.1 Simscape Electrical

Simscape Electrical provides component libraries for modeling and simulating electronic, mechatronic, and electrical power systems. It includes models of semiconductors, motors, and components for applications such as electromechanical actuation, smart grids, and renewable energy systems. You can use these components to evaluate analog circuit architectures, develop mechatronic systems with electric drives, and analyze the generation, conversion, transmission, and consumption of electrical power at the grid level.

Simscape Electrical helps you develop control systems and test system-level performance. You can parameterize your models using MATLAB variables and expressions, and design control systems for electrical systems in Simulink. You can integrate mechanical, hydraulic, thermal, and other physical systems into your model using components from the Simscape family of products. To deploy models to other simulation environments, including hardware-in-the-loop (HIL) systems, Simscape Electrical supports C-code generation.

Chapter 8

LabView

Laboratory Virtual Instrument Engineering Workbench (LabVIEW) is a system-design platform and development environment for a visual programming language from National Instruments. LabVIEW is systems engineering software for applications that require test, measurement, and control with rapid access to hardware and data insights. LabVIEW offers a graphical programming approach that helps you visualize every aspect of your application, including hardware configuration, measurement data, and debugging. This visualization makes it simple to integrate measurement hardware from any vendor, represent complex logic on the diagram, develop data analysis algorithms, and design custom engineering user interfaces.

8.1 Getting Started Modules

Begin here <http://www.learnni.com/getting-started/> to learn about the program and how to get started.

8.2 MathScript

Launch **NI LabVIEW** and click **Tools** → **MathScript Window...** as shown in Fig. 8.1.

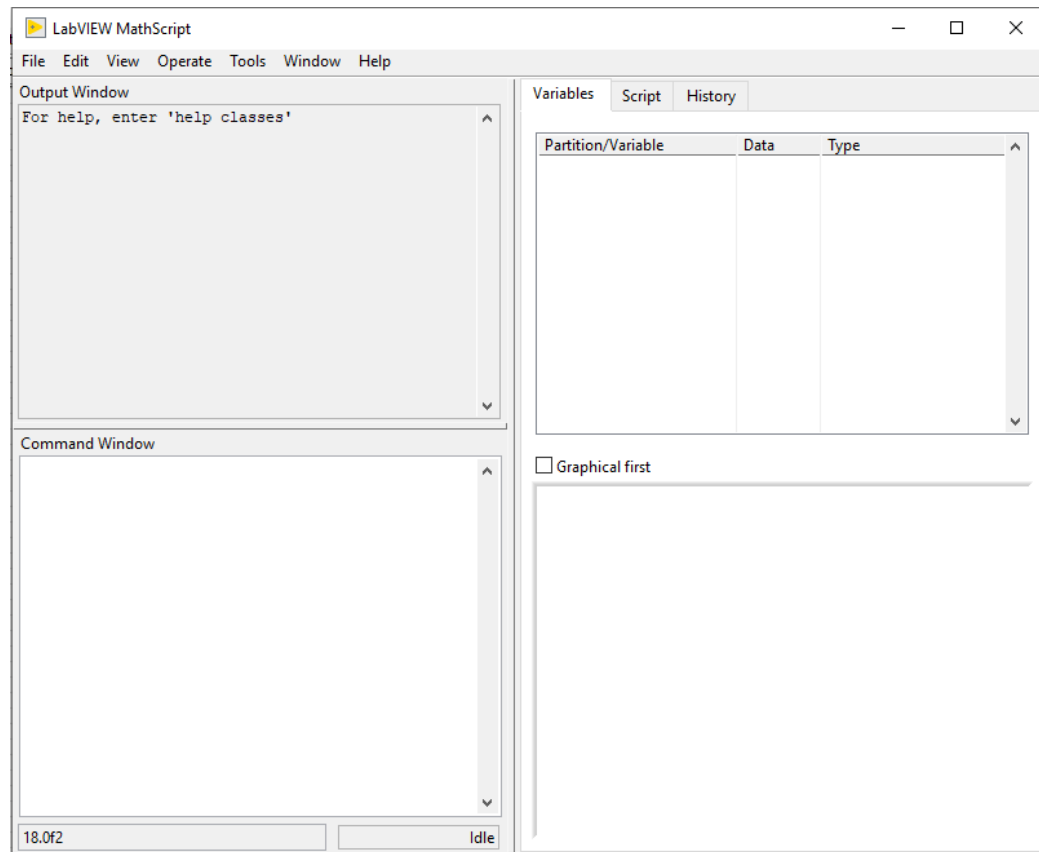


Figure 8.1: MathScript RT Module Functions

8.3 Mathematics

To access the **Functions** Palette, right click in any blank space of the **Block Diagram** window. Choose the category **Mathematics**, and select **Polynomial**. The **Mathematics** Palette is shown in Fig. 8.2.

8.4 Control and Simulation

To access the **Functions** Palette, right click in any blank space of the **Block Diagram** window. Choose the category **Control & Simulation**, and select **Control Design** or **Simulation**. The **Control & Simulation** Palette is shown in Fig. 8.3. Click **Control Design**, choose **Model Construction** to construct

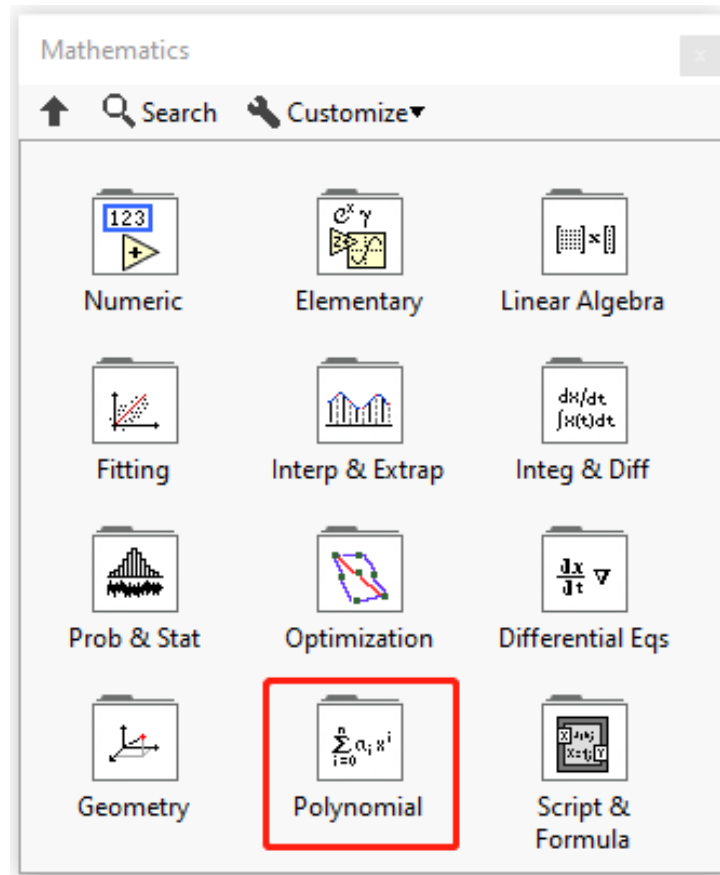


Figure 8.2: Mathematics Functions

transfer functions and [Model Interconnection](#) to display the results of the mathematical operations in Fig. 8.4.

8.5 Examples:

8.5.1 System Modeling

CDEx Nonlinear and Linear Simulation.vi

SimEx Nonlinear and Linear Pendulum Simulation.vi

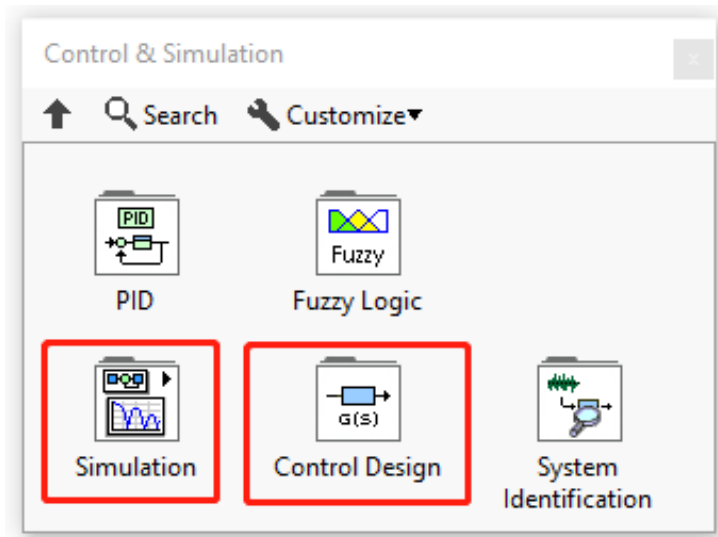


Figure 8.3: Control and Simulation Functions

8.5.2 Block Diagram Reduction

CDEx Block Diagram Reduction.vi

8.5.3 Dynamic Response

On-Off Controller.vi

Aircraft bank angle.vi

8.5.4 System Type

Objective

Get started with control system analysis and design, and get familiar with key control concepts, such as closed-loop control, step response, ramp response, transient response and steady-state error.

Software

LabVIEW and the LabVIEW Control Design and Simulation Module.

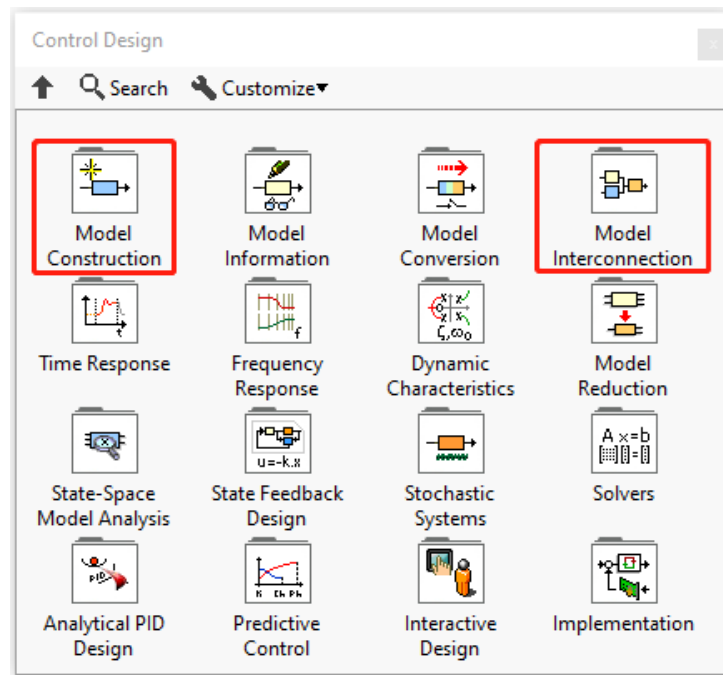


Figure 8.4: Control Design

Lab

1. Launch **NI LabVIEW** and click **Help** → **Find Examples...**
2. In the **NI Example Finder** window, double-click **Toolkits and Modules** ⇒ **Control and Simulation** ⇒ **Control Design** ⇒ **Time Analysis** ⇒ **CDEx Effect of Controller Type.vi**
3. Run the VI continuously.

Report

1. Observe the transient response and record **Steady-state Tracking Error** and **Steady-state Velocity Error** when the **Control Gain** is 0.838 and the **Controller Type** is **Type 1 - one integrator**.
2. Repeat Step 1 by changing the type of Controller to **Type 0 - No integrators** and **Type 2 - two integrators**.

3. Repeat Step 1 with high gains by moving the slider **Controller Gain** to the right and with low gains by moving the slider **Controller Gain** to the left.
4. Describe how the type of control and the Controller Gain affect the step (Steady-state Tracking Error) and the ramp (Steady-state Velocity Error) response.

8.5.5 PID Design

CDEx PID Design.vi

Appendices

Appendix A

Image Credits

Figure	License	Arthur	URL
1.3	CC BY 3.0	Egmason	Flush toilet
1.6	CC BY-SA 3.0	M.Minderhoud	Adaptive Cruise Control
3.2	CC BY-SA 3.0	Pflodo	Kirchhoff's circuit laws
3.3	CC BY-SA 3.0	Kwinkunks	Kirchhoff's circuit laws
3.4	CC BY-SA 3.0	Ong saluri	Operational amplifier
3.5	CC BY-SA 3.0	Ong saluri	Operational amplifier
4.2	CC BY-SA 3.0	UlrichHeither	Quadcopter

Bibliography

- [1] N. Nise, *Control Systems Engineering*. John Wiley & Sons, Limited, 2019.
- [2] G. Franklin, J. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*. Pearson Education, 2011.
- [3] C. L. Phillips and J. M. Parr, *Feedback Control Systems*. Pearson, 2011.
- [4] R. Dorf and R. Bishop, *Modern Control Systems*. Pearson Prentice Hall, 2008.
- [5] K. Ogata, *Modern Control Engineering*, ser. Instrumentation and controls series. Prentice Hall, 2010.