Chapter 2 Classical Analysis Methods

Classical Analysis Methods

2.1 Process Dynamic Responses

Plant Description

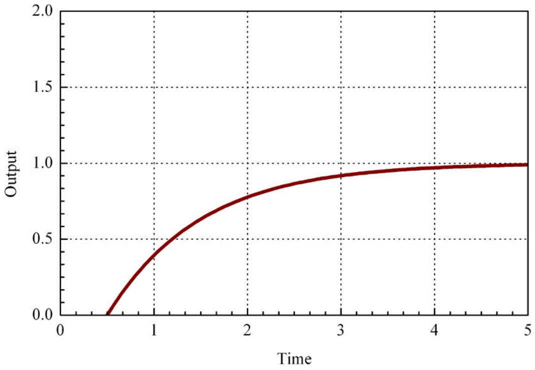
Different plants may be described by the same model:
Distillation column, paper-making machine, disk, maglev, ...
Description: Linear time-invariant causal model $G(t)$, where t is the continuous time variable. $G(s)$ denote its transfer function.
Causality: $G(t) = 0$ for $t < 0$. Implication: the output depends only on the current and the previous inputs
Properness:

1. Proper—$G(s)|_{s=j\infty}$ is finite (the degree of denominator is greater than or equals the degree of numerator)
2. Strictly proper—$G(s)|_{s=j\infty} = 0$ (the degree of denominator is greater than the degree of numerator)
3. Bi-proper—$G(s)|_{s=j\infty}$ is a nonzero constant (the degree of denominator equals the degree of numerator)
4. Improper—Not proper

Stable Plants with Time Delays

When the original mass or energy equilibrium is upset by a change at the input, the output will eventually reach a new equilibrium.

Feature: Do not have closed RHP poles



Models of Stable Plants

$$G(s) = \frac{K}{(\tau_1 s + 1)(\tau_2 s + 1)...(\tau_n s + 1)} e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
$\tau_i (i = 1, 2, ..., n)$—Have positive real parts and denote time constants

The first-order model frequently used in practice:

$$G(s) = \frac{K}{\tau s + 1} e^{-\theta s}$$

The model can well be illustrated by utilizing a shower

Models of Stable Plants

$$G(s) = \frac{K}{(\tau_1 s + 1)(\tau_2 s + 1)\ldots(\tau_n s + 1)} e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
$\tau_i (i = 1, 2, \ldots, n)$—Have positive real parts and denote time constants

The first-order model frequently used in practice:

$$G(s) = \frac{K}{\tau s + 1} e^{-\theta s}$$

The model can well be illustrated by utilizing a shower

The Model of a Shower

At t, turn up the valve of hot water by a small increment Δq

At t_1 the increment of the warm water temperature is Δc and does not increase anymore

The warmer water flows to the outlet at t_2

$$K = \Delta c / \Delta q$$
$$\tau = t_1 - t$$
$$\theta = t_3 - t_2.$$

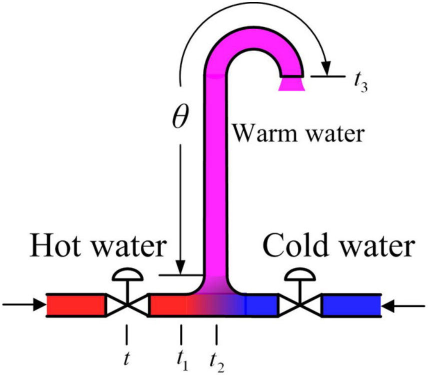
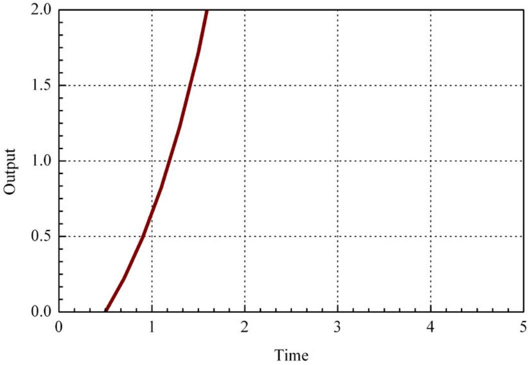


Figure: Shower

Unstable Plants with Time Delays

When the original mass or energy equilibrium is upset by a change at the input, the output will increase or decrease faster and faster until the physical limit is reached.

Feature: Have RHP poles



Models of Unstable Plants

$$G(s) = \frac{K}{(-\tau_1 s + 1)(-\tau_2 s + 1)...(-\tau_m s + 1)\times (\tau_{m+1} s + 1)...(\tau_n s + 1)} e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
$\tau_i (i = 1, 2, ..., m, ..., n)$—Have positive real parts and denote time constants

The first-order model frequently used in practice:

$$G(s) = \frac{K}{\tau s - 1} e^{-\theta s}$$

Models of Unstable Plants

$$G(s) = \frac{K}{(-\tau_1 s + 1)(-\tau_2 s + 1)...(-\tau_m s + 1) \times (\tau_{m+1} s + 1)...(\tau_n s + 1)} e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
$\tau_i (i = 1, 2, ..., m, ..., n)$—Have positive real parts and denote time constants

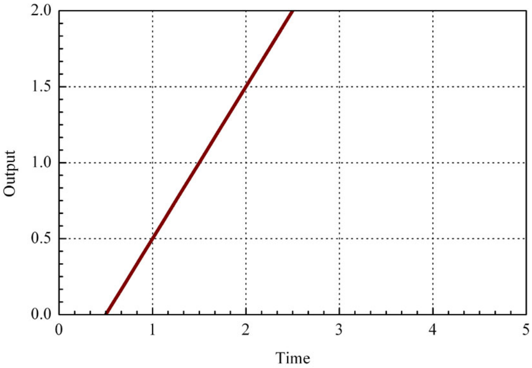
The first-order model frequently used in practice:

$$G(s) = \frac{K}{\tau s - 1} e^{-\theta s}$$

Integrating Plants with Time Delays

When the original mass or energy equilibrium is upset by a change at the input, the output will increase or decrease with a fixed speed until the physical limit is reached.

Feature: Have poles at the origin



Models of Integrating Plants

$$G(s) = \frac{K}{s^m(\tau_1 s + 1)(\tau_2 s + 1)...(\tau_n s + 1)} e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
$\tau_i (i = 1, 2, ..., n)$—Have positive real parts and denote time constants
m—Integer

The first-order model frequently used in practice:

$$G(s) = \frac{K}{s} e^{-\theta s}$$

Models of Integrating Plants

$$G(s) = \frac{K}{s^m(\tau_1 s + 1)(\tau_2 s + 1)...(\tau_n s + 1)} e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
$\tau_i (i = 1, 2, ..., n)$—Have positive real parts and denote time constants
m—Integer

The first-order model frequently used in practice:

$$G(s) = \frac{K}{s} e^{-\theta s}$$

Categories of Plants

$$\textbf{According to pole positions:} \begin{cases} \text{Stable plants} \\ \text{Integrating plants} \\ \text{Unstable plants} \end{cases}$$

$$\textbf{According to zero positions:} \begin{cases} \text{MP plants} \\ \text{NMP plants} \end{cases}$$

MP plants: Its transfer function does not contain zeros in the closed RHP or a time delay

NMP plants: All plants that are not MP

The General Plant

In theoretical study, the following model may be used:

$$G(s) = \frac{KN_+(s)N_-(s)}{M_+(s)M_-(s)}e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
"$+$"—Denote that all of the roots are in the closed RHP
"-"—Denote that all of the roots are in the open LHP

Assumption 1: $N_+(0) = N_-(0) = M_+(0) = M_-(0) = 1$, **which is made solely to simplify the statement**
Assumption 2: $\deg\{N_+\} + \deg\{N_-\} \leq \deg\{M_-\} + \deg\{M_+\}$, **with which the plant is proper**

The General Plant

In theoretical study, the following model may be used:

$$G(s) = \frac{KN_+(s)N_-(s)}{M_+(s)M_-(s)}e^{-\theta s}$$

K—Real constant denoting the static gain
θ—Positive real constant denoting the time delay
"$+$"—Denote that all of the roots are in the closed RHP
"-"—Denote that all of the roots are in the open LHP

Assumption 1: $N_+(0) = N_-(0) = M_+(0) = M_-(0) = 1$, **which is made solely to simplify the statement**
Assumption 2: $\deg\{N_+\} + \deg\{N_-\} \leq \deg\{M_-\} + \deg\{M_+\}$, **with which the plant is proper**

2.2 Rational Approximations for Time Delay

Why Use Rational Approximations

Main reasons:

- The time delay is an irrational function, which is of infinite dimension
- Most design methods developed so far are based on rational functions. They are only applicable to plants of finite dimension

One way to overcome the problem: Approximate the time delay by employing rational functions

2.2 Rational Approximations for Time Delay

Why Use Rational Approximations

Main reasons:

- The time delay is an irrational function, which is of infinite dimension
- Most design methods developed so far are based on rational functions. They are only applicable to plants of finite dimension

One way to overcome the problem: Approximate the time delay by employing rational functions

Rational Approximations

1. Apprximation with lags

$$e^{-\theta s} = \lim_{n\to\infty} \left(\frac{1}{1+\theta s/n} \right)^n$$

The first-order approximation:

$$e^{-\theta s} = \frac{1}{1+\theta s}$$

2. Taylor series expansion

$$e^{-\theta s} = \lim_{n\to\infty} 1 - \theta s + \theta^2 s^2/2! + ... + (-1)^n \theta^n s^n/n!$$

The first-order approximation:

$$e^{-\theta s} = 1 - \theta s$$

Rational Approximations

1. Apprximation with lags

$$e^{-\theta s} = \lim_{n\to\infty}\left(\frac{1}{1+\theta s/n}\right)^n$$

The first-order approximation:

$$e^{-\theta s} = \frac{1}{1+\theta s}$$

2. Taylor series expansion

$$e^{-\theta s} = \lim_{n\to\infty} 1 - \theta s + \theta^2 s^2/2! + \ldots + (-1)^n \theta^n s^n/n!$$

The first-order approximation:

$$e^{-\theta s} = 1 - \theta s$$

3. The Pade apprximation

Tools: Rational fraction expressions

Basic ideas: Make the power series expansion of **a rational function** match a **given** power series expansion with as many as possible terms

Assume that the formal power series expansion of $F(s)$ is

$$F(s) = c_0 + c_1 s + c_2 s^2 + \ldots.$$

Let m and n be nonnegative integers. The Pade approximant of $F(s)$ is a rational fraction given by

$$\frac{V_{mn}(s)}{P_{mn}(s)} = \frac{a_m s^m + a_{m-1} s^{m-1} + \ldots + a_0}{b_n s^n + b_{n-1} s^{n-1} + \ldots + b_0}$$

To obtain the unique solution, take $b_0 = 1$ ($m + n + 1$ unknowns)

3. The Pade apprximation

Tools: Rational fraction expressions

Basic ideas: Make the power series expansion of **a rational function** match a **given** power series expansion with as many as possible terms

Assume that the formal power series expansion of $F(s)$ is

$$F(s) = c_0 + c_1 s + c_2 s^2 + \ldots.$$

Let m and n be nonnegative integers. The Pade approximant of $F(s)$ is a rational fraction given by

$$\frac{V_{mn}(s)}{P_{mn}(s)} = \frac{a_m s^m + a_{m-1} s^{m-1} + \ldots + a_0}{b_n s^n + b_{n-1} s^{n-1} + \ldots + b_0}$$

To obtain the unique solution, take $b_0 = 1$ ($m + n + 1$ unknowns)

Let the Taylor series expansion of the Pade approximant match the first $m+n+1$ terms of the power series expansion of $F(s)$:

$$\begin{aligned} & (b_n s^n + b_{n-1}s^{n-1} + \ldots + b_0)(c_0 + c_1 s + c_2 s^2 + \ldots + c_{m+n}s^{m+n}) \\ = \; & a_m s^m + a_{m-1}s^{m-1} + \ldots + a_0 \end{aligned}$$

Compare the coefficients of $1, s, \ldots, s^{m+n}$ in the two sides of the equation. One obtains

$$\begin{bmatrix} a_0 \\ a_1 \\ \ldots \\ a_m \end{bmatrix} = \begin{bmatrix} c_0 & 0 & 0 & \ldots & 0 \\ c_1 & c_0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ c_m & c_{m-1} & c_{m-2} & \ldots & c_{m-n} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \ldots \\ b_n \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} c_{m+1} & c_m & \ldots & c_{m-n+1} \\ c_{m+2} & c_{m+1} & \ldots & c_{m-n+2} \\ \ldots & \ldots & \ldots & \ldots \\ c_{m+n} & c_{m+n-1} & \ldots & c_m \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \ldots \\ b_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix} \tag{2}$$

Here $c_i = 0$ when $i < 0$.

If the equations have a solution, the coefficients of $P_{mn}(s)$ can be obtained from (2), and the coefficients of $V_{mn}(s)$ can be obtained from (1).
For exponential functions, the Pade approximant has a more clear expression. The m/n Pade approximant of a time delay can be written as

$$\begin{aligned} e^{-\theta s} &\approx \frac{V_{mn}(\theta s)}{P_{mn}(\theta s)} \\ V_{mn}(\theta s) &= \sum_{j=0}^{m} \frac{(m+n-j)!m!}{(m+n)!j!(m-j)!}(-\theta s)^j \\ P_{mn}(\theta s) &= \sum_{j=0}^{n} \frac{(m+n-j)!n!}{(m+n)!j!(n-j)!}(\theta s)^j \end{aligned}$$

It can be verified that $P_{mn}(\theta s) = V_{nm}(-\theta s)$.

The first-order approximation:

$$e^{-\theta s} = \frac{1 - \theta s/2}{1 + \theta s/2}$$

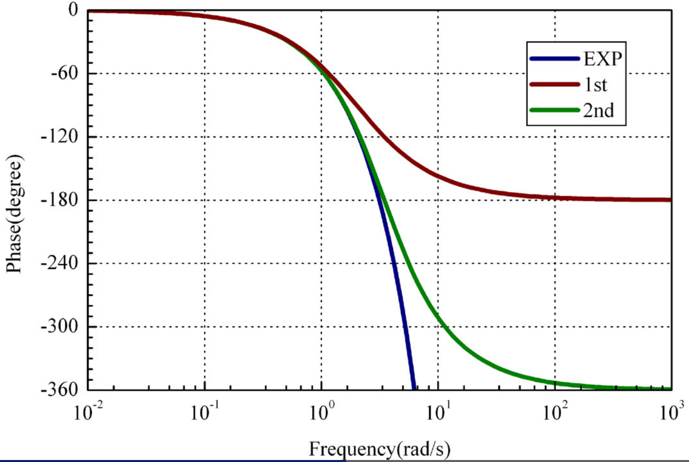
When $m = n$, the all-pass Pade approximant is obtained.

All-pass: For SISO systems, a transfer function is all-pass if its magnitude equals 1 at all points on the imaginary axis.
Features of all-pass functions: An all-pass transfer function passes without attenuation input sinusoids of all frequencies.
Zeros and poles of an all-pass Pade approximant:

- All zeros are in the open RHP
- All poles are in the open LHP
- The zeros and the poles are mirror-images of each other

Features of the all-pass Pade approximant:

1. Better precision can be obtained than the Taylor series of the same order.
2. The magnitude characteristic of time delay is preserved; the only difference is the phase



Limitation of Rational Approximations

Stability: A closed-loop system is stable if its characteristic equation has no roots in the **closed** RHP

Rational approximations were seldom used in classical control theory for analyzing the stability of the closed-loop system, because they **cannot** guarantee the correctness of the result.

Example

To see how the designer may be misled, consider the simple system shown in Figure. The characteristic equation of the closed-loop system is

$$1 + se^{s} = 0.$$

Limitation of Rational Approximations

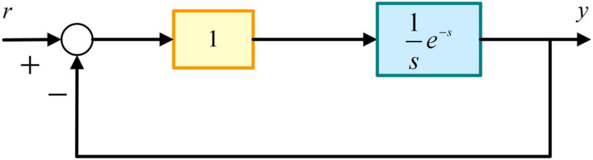
Stability: A closed-loop system is stable if its characteristic equation has no roots in the **closed** RHP

Rational approximations were seldom used in classical control theory for analyzing the stability of the closed-loop system, because they **cannot** guarantee the correctness of the result.

Example

To see how the designer may be misled, consider the simple system shown in Figure. The characteristic equation of the closed-loop system is

$$1 + se^{s} = 0.$$



Example (ctd.1)

It is easy to verify that the closed-loop system is stable. With the Taylor series expansion of different orders, different roots for the characteristic equation can be obtained :

$$
\begin{aligned}
1+s=0 &\qquad \text{The root is } s=-1\\
1+s+s^2=0 &\qquad \text{The roots are } s=-0.5\pm j0.8660\\
1+s+s^2+s^3/2=0 &\qquad \text{The roots are } s=-1.5437,\\
&\qquad\qquad\qquad\qquad -0.2282\pm j1.1151
\end{aligned}
$$

Example (ctd.2)

According to these results, the closed-loop system should be stable. However, a higher order Taylor series expansion gives the following equation:

$$1 + s + s^2 + \frac{s^3}{2} + \frac{s^4}{3!} + \frac{s^5}{4!} + \frac{s^6}{5!} = 0,$$

which has two roots in the RHP:

$$s = 0.1041 \pm j3.0815.$$

Nevertheless, this does not imply that rational approximations cannot be used for the analysis and design of a control system.
Key: Choosing proper methods

Example (ctd.2)

According to these results, the closed-loop system should be stable. However, a higher order Taylor series expansion gives the following equation:

$$1 + s + s^2 + \frac{s^3}{2} + \frac{s^4}{3!} + \frac{s^5}{4!} + \frac{s^6}{5!} = 0,$$

which has two roots in the RHP:

$$s = 0.1041 \pm j3.0815.$$

Nevertheless, this does not imply that rational approximations cannot be used for the analysis and design of a control system.
Key: Choosing proper methods

2.3 Time Domain Performance Indices

Control Structures in This Book

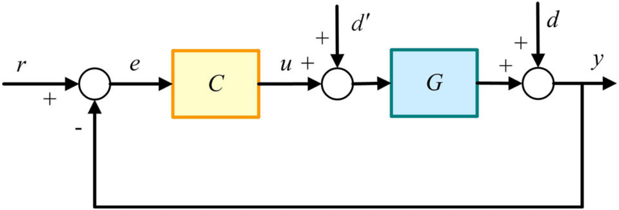
The most elementary feedback control system has two components:

- A plant to be controlled
- A controller to generate the input to the plant

The main control structure discussed in this book:
The unity feedback control loop
The unity feedback control loop has played a vital role in the study of control theory:

- It is the most widely used structure
- Most of the non-unity feedback control loops can easily be converted into the unity feedback control loop

The Unity Feedback Loop



$G(s)$—Plant
$C(s)$—Controller
$r(s)$—Reference
$e(s)$—Error
$u(s)$—Controller output (control variable/manipulated variable)

$y(s)$—System output (controlled variable)
$d'(s)$—Input disturbance
$d(s)$—Output disturbance
Measurement noise—Very small and thus neglected

Test Signals

The response of a system depends on not only the **model**, but also the **input** and the **initial condition**

Without loss of generality, it is a common practice to use the standard initial condition; that is, the system is at rest initially with its output and all time derivatives thereof being zero

In the analysis and design of control systems, performances of various systems should be compared on the same basis. This can be achieved by specifying a particular test signal for the input and then comparing responses of different systems to the test signal

Test Signals

The response of a system depends on not only the **model**, but also the **input** and the **initial condition**

Without loss of generality, it is a common practice to use the standard initial condition; that is, the system is at rest initially with its output and all time derivatives thereof being zero

In the analysis and design of control systems, performances of various systems should be compared on the same basis. This can be achieved by specifying a particular test signal for the input and then comparing responses of different systems to the test signal

Frequently used Test Signals

Impulse:

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ \infty & t = 0 \end{cases},$$

$$\int_0^\infty \delta(t)\, dt = 1,$$

No ideal impulse exists in the real world. It can be approximated by a rectangular pulse. The Laplace transform is 1

Step:

$$r(t) = \begin{cases} 0 & t < 0 \\ A & t \geq 0 \end{cases},$$

where A is a constant. The step signal with $A = 1$ is called the unit step. The Laplace transform is $1/s$

Frequently used Test Signals

Impulse:

$$\delta(t) = \begin{cases} 0 & t \neq 0 \\ \infty & t = 0 \end{cases},$$

$$\int_0^{\infty} \delta(t)\, dt = 1,$$

No ideal impulse exists in the real world. It can be approximated by a rectangular pulse. The Laplace transform is 1

Step:

$$r(t) = \begin{cases} 0 & t < 0 \\ A & t \geq 0 \end{cases},$$

where A is a constant. The step signal with $A = 1$ is called the unit step. The Laplace transform is $1/s$

Ramp:

$$r(t) = \begin{cases} 0 & t < 0 \\ At & t \geq 0 \end{cases},$$

where A is a constant. The Laplace transform is $1/s^2$

Sinusoidal:

$$r(t) = \begin{cases} 0 & t < 0 \\ A \sin \omega t & t \geq 0 \end{cases}.$$

Its Laplace transform is $A\omega/(s^2 + \omega^2)$

It is impossible to design a controller working well for all inputs. Most controllers are designed for **steps**

Ramp:

$$r(t) = \begin{cases} 0 & t < 0 \\ At & t \geq 0 \end{cases},$$

where A is a constant. The Laplace transform is $1/s^2$

Sinusoidal:

$$r(t) = \begin{cases} 0 & t < 0 \\ A\sin\omega t & t \geq 0 \end{cases}.$$

Its Laplace transform is $A\omega/(s^2+\omega^2)$

It is impossible to design a controller working well for all inputs. Most controllers are designed for **steps**

Ramp:

$$r(t) = \begin{cases} 0 & t < 0 \\ At & t \geq 0 \end{cases},$$

where A is a constant. The Laplace transform is $1/s^2$

Sinusoidal:

$$r(t) = \begin{cases} 0 & t < 0 \\ A \sin \omega t & t \geq 0 \end{cases}.$$

Its Laplace transform is $A\omega/(s^2 + \omega^2)$

It is impossible to design a controller working well for all inputs. Most controllers are designed for **steps**

Time Domain Indices

In classical control theory, the performance of a control system is usually characterized in terms of the transient response and the steady-state response:

- Transient response—The response that goes from the initial state to the steady state
- Steady-state response—The manner in which the system output behaves as the time approaches infinity

People usually pay more attention to the transient response, because

- It is difficult to obtain the required transient response
- The system almost always goes from one state to another

Time Domain Indices

In classical control theory, the performance of a control system is usually characterized in terms of the transient response and the steady-state response:

- Transient response—The response that goes from the initial state to the steady state
- Steady-state response—The manner in which the system output behaves as the time approaches infinity

People usually pay more attention to the transient response, because

- It is difficult to obtain the required transient response
- The system almost always goes from one state to another

Performance indices of the **transient response** are generally defined by utilizing the unit step response

Overshoot σ

The maximum value of the unit step response, measured from unity. It is common to use the percentage overshoot

Rise time t_r

Overdamped systems: The time required for the unit step response to rise from 10% to 90% of its steady-state value
Underdamped systems: The time from zero to that the unit step response reaches the steady-state value for the first time is usually used in the description of the rise time

Settling time t_s

The time required for the unit step response to reach and stay within a given error band (usually 5% of the steady-state value)

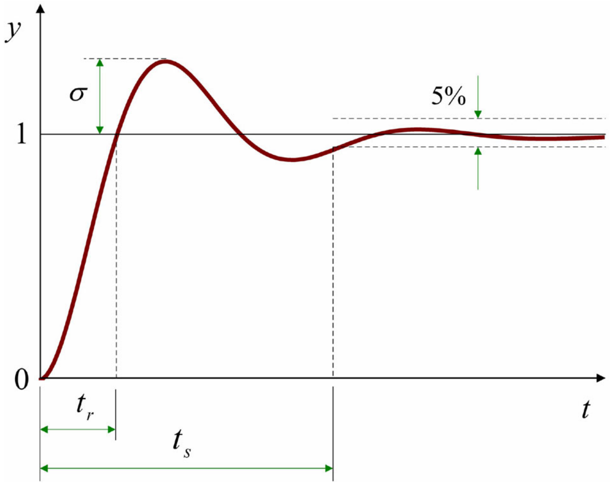


Figure: Step response curve for time-domain performance indices

One major objective of control systems: To reject the effect of disturbances and keep the system output as close as possible to the reference

To provide the information for disturbance responses, two transient performance indices are defined for the unit step disturbance at the plant **input**:

Perturbation peak ρ

The maximum value of the disturbance response, measured from the steady-state value

Recovery time t_{rs}

The time required for the disturbance response to reach and stay within a given error band (usually 5% of the steady-state value)

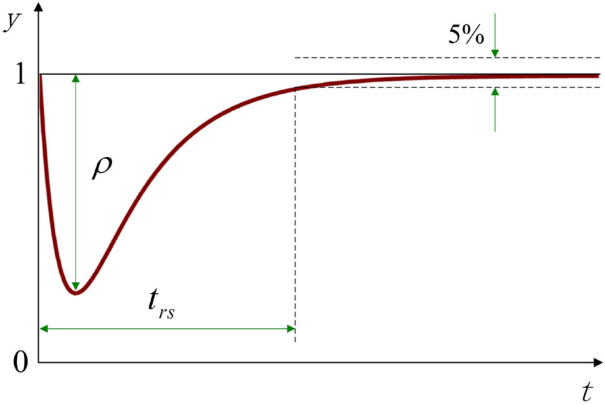


Figure: Disturbance response curve for time-domain performance indices

Measure for steady-state performance: Steady-state error

Steady-state error: The difference between the desired output and the real output of the system as time goes to infinity

It is desirable that the steady-state error gradually vanishes

Let $L(s) = G(s)C(s)$ be the open-loop transfer function. From Final Value Theorem it is known that the steady-state error of a stable system is

$$\lim_{t \to \infty} e(t) = \lim_{s \to 0} \frac{sr(s)}{1 + L(s)}$$

The steady-state error depends on not only the **input** but also the **open-loop transfer function**. Whether or not a system exhibits a steady-state error is determined by the **system type**

Definition

The system is said to be of Type m if $L(s)$ has m poles at the origin

The steady-state error of the system for a **step** reference is

$$\lim_{t\to\infty} e(t) = \lim_{s\to 0} \frac{1}{1+L(s)} = \frac{1}{1+\lim_{s\to 0} L(s)}$$

It is a constant for a Type 0 system. If the zero steady-state error is required, the system type must be at least one

The steady-state error of the system for a **ramp** reference is

$$\lim_{t\to\infty} e(t) = \lim_{s\to 0} \frac{1}{s+sL(s)} = \frac{1}{\lim_{s\to 0}[sL(s)]}$$

The steady-state error of a Type 0 system is infinite. A Type 1 system can follow a ramp with a finite steady-state error. A Type 2 or higher system can follow a ramp with zero steady-state error

The Problem of the classical performance indices: Difficult to express them in mathematical forms

An alternative for controller design: Integral performance indices

Merits: Can be optimized with mathematical methods

Integral Absolute Error (IAE):

$$\mathrm{IAE} = \int_0^\infty |e(t)|dt.$$

Integral Square Error (ISE):

$$\mathrm{ISE} = \int_0^\infty e^2(t)dt.$$

Integral of Time Multiply by Absolute Error (ITAE):

$$\mathrm{ITAE} = \int_0^\infty t|e(t)|dt.$$

The Problem of the classical performance indices: Difficult to express them in mathematical forms

An alternative for controller design: Integral performance indices

Merits: Can be optimized with mathematical methods

Integral Absolute Error (IAE):

$$\mathrm{IAE} = \int_0^\infty |e(t)|dt.$$

Integral Square Error (ISE):

$$\mathrm{ISE} = \int_0^\infty e^2(t)dt.$$

Integral of Time Multiply by Absolute Error (ITAE):

$$\mathrm{ITAE} = \int_0^\infty t|e(t)|dt.$$

The ISE index is the only one that is widely applied. This is because

- It is easier to perform the optimization procedure with the ISE index than with other indices (**Mathematical convenience!**)
- Other indices can not provide superior performance to the ISE

ISE-based indices: LQ, LQG, H_2, etc.

Basic design objective of the H_2 control: Search for a controller such that the ISE of the system is minimized for the **impulse** (or equivalently, the white noise with zero-mean and unit variance)

Basic design objective of the H_∞ control: Minimize the worst ISE resulted from all energy-bounded inputs:

$$\sup_{r(t)} \int_0^\infty e^2(t)dt$$

Limitation: All these indices do not relate to classical indices

The ISE index is the only one that is widely applied. This is because

- It is easier to perform the optimization procedure with the ISE index than with other indices (**Mathematical convenience!**)
- Other indices can not provide superior performance to the ISE

ISE-based indices: LQ, LQG, H_2, etc.
Basic design objective of the H_2 control: Search for a controller such that the ISE of the system is minimized for the **impulse** (or equivalently, the white noise with zero-mean and unit variance)
Basic design objective of the H_∞ control: Minimize the worst ISE resulted from all energy-bounded inputs:

$$\sup_{r(t)} \int_0^\infty e^2(t)dt$$

Limitation: All these indices do not relate to classical indices

2.4 Frequency Response Analysis

Frequency Response

Basis of frequency response analysis: When a linear system is subject to a sinusoidal input, its ultimate response is also a sustained sinusoidal wave
The closed-loop transfer function of the unity feedback loop is

$$T(s) = \frac{L(s)}{1 + L(s)}$$

The frequency response of the system is

$$T(j\omega) = \frac{L(j\omega)}{1 + L(j\omega)}$$

which can be expressed with magnitude and phase as

$$T(j\omega) = |T(j\omega)|\angle T(j\omega)$$

In frequency response analysis some points with special characteristics are selected to define the performance indices

Resonance peak T_p

The maximum value of the magnitude of the closed-loop frequency response

Resonance frequency ω_r

The frequency at which the resonance peak occurs

Bandwidth BW

The range of frequencies beyond which the magnitude of a signal drops down by more than 3dB

Feature of frequency response analysis: The stability, as well as the performance, of the **closed-loop** system can be determined from the characteristics of the **open-loop** response

In frequency response analysis some points with special characteristics are selected to define the performance indices

Resonance peak T_p

The maximum value of the magnitude of the closed-loop frequency response

Resonance frequency ω_r

The frequency at which the resonance peak occurs

Bandwidth BW

The range of frequencies beyond which the magnitude of a signal drops down by more than 3dB

Feature of frequency response analysis: The stability, as well as the performance, of the **closed-loop** system can be determined from the characteristics of the **open-loop** response

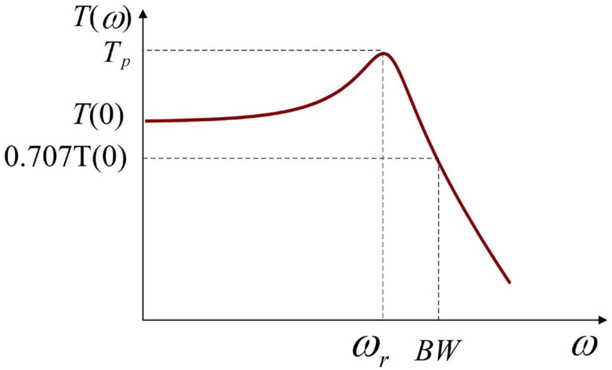


Figure: Magnitude curve for frequency response analysis

Nyquist Stability Criterion

Nyquist path: It starts at the origin, goes up the imaginary axis, turns into the RHP following a semicircle of infinity radius, and comes up the negative imaginary axis to the origin again.

Nyquist plot: As a point $s = j\omega$ makes one circuit around this curve, the point $L(j\omega)$ traces out a curve called the Nyquist plot of the transfer function $L(s)$

Theorem (Nyquist Stability Criterion)

Let n denote the total number of poles of $L(s)$ in the RHP, then the closed-loop system is stable if and only if the Nyquist plot of $L(s)$ does not pass through the $(-1, j0)$ point and encircles it n times counterclockwise

Nyquist Stability Criterion

Nyquist path: It starts at the origin, goes up the imaginary axis, turns into the RHP following a semicircle of infinity radius, and comes up the negative imaginary axis to the origin again.

Nyquist plot: As a point $s = j\omega$ makes one circuit around this curve, the point $L(j\omega)$ traces out a curve called the Nyquist plot of the transfer function $L(s)$

Theorem (Nyquist Stability Criterion)

Let n denote the total number of poles of $L(s)$ in the RHP, then the closed-loop system is stable if and only if the Nyquist plot of $L(s)$ does not pass through the $(-1, j0)$ point and encircles it n times counterclockwise

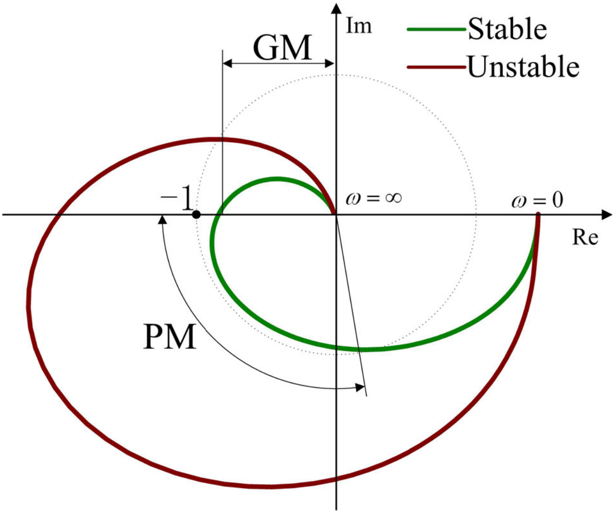


Figure: Nyquist plot of a stable open-loop system

Bode Stability Criterion

Bode plot: Consists of two graphs: One is a graph of the logarithm of the magnitude, $|L(j\omega)|$; the other is a graph of the phase angle, $\angle L(j\omega)$ As a point $s = j\omega$ makes one circuit around this curve, the point $L(j\omega)$ traces out a curve called the Nyquist plot of the transfer function $L(s)$

Theorem (Bode Stability Criterion)

A closed-loop system is unstable if the frequency response of $L(s)$ has a magnitude greater than or equal to 1 at the frequency where the phase angle is -180°

Bode Stability Criterion

Bode plot: Consists of two graphs: One is a graph of the logarithm of the magnitude, $|L(j\omega)|$; the other is a graph of the phase angle, $\angle L(j\omega)$ As a point $s = j\omega$ makes one circuit around this curve, the point $L(j\omega)$ traces out a curve called the Nyquist plot of the transfer function $L(s)$

Theorem (Bode Stability Criterion)

A closed-loop system is unstable if the frequency response of $L(s)$ has a magnitude greater than or equal to 1 at the frequency where the phase angle is -180°

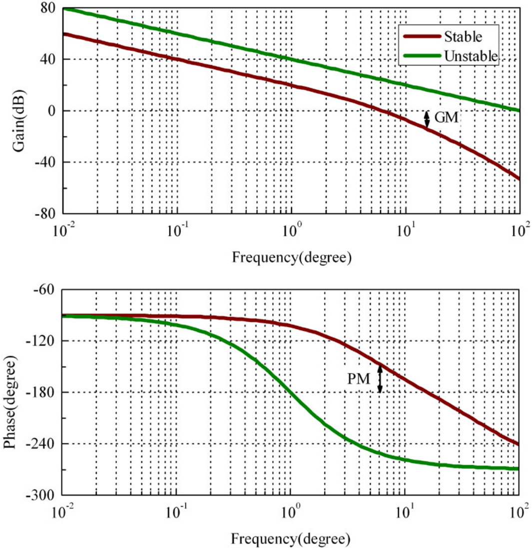


Figure: Bode plot

Stability Margins

Not only can the Bode plot be used to test the stability of the closed-loop system, but also to describe the relative stability:

Gain Margin (GM)

Let ω_c be the frequency where the phase angle is -180°. The gain margin is the reciprocal of the magnitude at the frequency.

$$\text{GM} = |L(j\omega_c)|^{-1}, \angle L(j\omega_c) = -180^\circ$$

Phase Margin (PM)

Let ω_g be the frequency where the magnitude is 1. The phase margin is the additional phase lag needed to distabilize the system.

$$\text{PM} = 180^\circ + \angle L(j\omega_g), |L(j\omega_g)| = 1$$

2.5 Transformation of Two Commonly Used Models

Experimental Modeling methods

Two typical experimental methods:

- Step response method—Based on the information from the open-loop step response. The model is usually a transfer function in the form of the first-order plant with time delay
- Ultimate gain method—Tune a **proportional controller** K_C to build a sustained oscillation, which gives the ultimate gain K_u and the ultimate period T_u

Typical design methods based on the step response model: e.g. Cohen-Coon (C-C) method
Typical design methods based on the the ultimate gain model: e.g. Ziegler-Nichols (Z-N) method

2.5 Transformation of Two Commonly Used Models

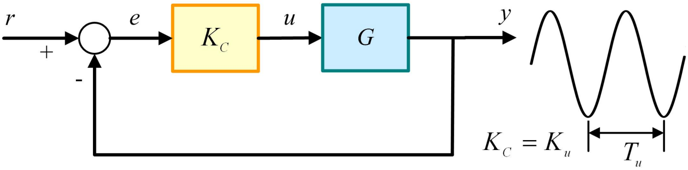
Experimental Modeling methods

Two typical experimental methods:

- Step response method—Based on the information from the open-loop step response. The model is usually a transfer function in the form of the first-order plant with time delay
- Ultimate gain method—Tune a **proportional controller** K_C to build a sustained oscillation, which gives the ultimate gain K_u and the ultimate period T_u

Typical design methods based on the step response model:
e.g. Cohen-Coon (C-C) method
Typical design methods based on the the ultimate gain model: e.g. Ziegler-Nichols (Z-N) method

The Ultimate Cycle Model



Explanation: An increase of K_C will make all the points on the Nyquist plot move radially outward from the origin. When K_C increases to the extent that the Nyquist plot passes through $(-1, j0)$, the stability limit is reached and a sustained oscillation occurs

Ultimate Cycle Model =>Step Response Model

Historically, the two models are developed independently

Idea: The plants they describe are the same one, they must be internally equivalent

Goal of this section: The quantitative relationship between them

Assume that the plant is stable and given by

$$G(s) = \frac{K}{\tau s + 1} e^{-\theta s}$$

The three parameters can be calculated by utilizing the obtained ultimate cycle model:

$$\begin{aligned} \angle K_u G(j\omega_u) &= -\pi \\ |K_u G(j\omega_u)| &= 1 \end{aligned}$$

Ultimate Cycle Model =>Step Response Model

Historically, the two models are developed independently

Idea: The plants they describe are the same one, they must be internally equivalent

Goal of this section: The quantitative relationship between them

Assume that the plant is stable and given by

$$G(s) = \frac{K}{\tau s + 1} e^{-\theta s}$$

The three parameters can be calculated by utilizing the obtained ultimate cycle model:

$$\begin{aligned} \angle K_u G(j\omega_u) &= -\pi \\ |K_u G(j\omega_u)| &= 1 \end{aligned}$$

Since

$$K_u G(j\omega_u) = \frac{K_u K e^{-j\theta\omega_u}}{j\tau\omega_u + 1}$$

The relationship between the step response model and the ultimate cycle model can be expressed as

$$-\theta\omega_u - \arctan(\tau\omega_u) = -\pi, \frac{K_u K}{\sqrt{(\tau\omega_u)^2 + 1}} = 1$$

When the plant gain K is known (Otherwise one can carry out the procedure twice), the time constant of the step response model is

$$\tau = \frac{\sqrt{(K_u K)^2 - 1}}{\omega_u}$$

The time delay of the step response model is

$$\theta = \frac{\pi - \arctan(\sqrt{(K_u K)^2 - 1})}{\omega_u}$$

Step Response Model => Ultimate Cycle Model

It is a challenge to obtain analytical expressions by solving the foregoing equations. Here they are derived by time domain analysis

Without loss of generality, assume that a unit step reference is used in the ultimate cycle method. The system output is

$$y(s) = \frac{1}{s}\frac{K_C K e^{-\theta s}}{\tau s + 1 + K_C K e^{-\theta s}}$$

Let

$$f(s) = \frac{K_C K e^{-\theta s}}{\tau s + 1 + K_C K e^{-\theta s}}$$

Then

$$y(s) = f(s)/s$$

where $f(s)$ is an irrational function

To derive the ultimate cycle model, $f(s)$ is expanded first by using a rational approximation. The Pade approximation is employed. Assume that the Maclaurin series extension of $f(s)$ is

$$f(s) = f(0) + f'(0)s + \frac{f''(0)}{2!}s^2 + \frac{f^{(3)}(0)}{3!}s^3 + \frac{f^{(4)}(0)}{4!}s^4 + \ldots$$

where

$$\begin{aligned}
f(0) &= \frac{K_C K}{1 + K_C K} \\
f'(0) &= -\frac{K_C K(\theta + \tau)}{(1 + K_C K)^2} \\
f''(0) &= -\frac{K_C K(K_C K\theta^2 - \theta^2 - 2\theta\tau - 2\tau^2 + 2K_C K\theta\tau)}{(1 + K_C K)^3} \\
f^{(3)}(0) &= -\frac{\begin{array}{c} K_C K(\theta^3 + 3K_c^2 K^2\theta^2\tau - 12K_C K\theta^2\tau + 6\tau^3 \\ -12K_C K\theta\tau^2 + 6\theta\tau^2 + K_c^2 K^2\theta^3 + 3\theta^2\tau - 4K_C K\theta^3) \end{array}}{(1 + K_C K)^4}
\end{aligned}$$

$$f^{(4)}(0) = -\frac{\begin{array}{c} K_C K(K_c^3 K^3 \theta^4 + 4K_c^3 K^3 \theta^3 \tau - 11K_c^2 K^2 \theta^4 \\ -44K_c^2 K^2 \theta^3 \tau - 48K_c^2 K^2 \theta^2 \tau^2 + 11K_C K \theta^4 \\ +44K_C K \theta^3 \tau + 84K_C K \theta^2 \tau^2 + 72K_C K \theta \tau^3 \\ -\theta^4 - 4\theta^3 \tau - 12\theta^2 \tau^2 - 24\theta \tau^3 - 24\tau^4) \end{array}}{(1 + K_C K)^5}.$$

Let the Pade approximation of $f(s)$ be

$$f(s) = \frac{a_2 s^2 + a_1 s + a_0}{b_2 s^2 + b_1 s + 1}$$

Then

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} f(0) & 0 & 0 \\ f'(0) & f(0) & 0 \\ f''(0)/2! & f'(0) & f(0) \end{bmatrix} \begin{bmatrix} 1 \\ b_1 \\ b_2 \end{bmatrix}$$

$$\begin{bmatrix} f''(0)/2! & f'(0) \\ f^{(3)}(0)/3! & f''(0)/2! \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = -\begin{bmatrix} f^{(3)}(0)/3! \\ f^{(4)}(0)/4! \end{bmatrix}$$

This leads to

$$
\begin{aligned}
a_0 &= f(0), \\
a_1 &= b_1 f(0) + f'(0), \\
a_2 &= b_2 f(0) + b_1 f'(0) + f''(0)/2!, \\
b_1 &= -\frac{f''(0)f^{(3)}(0)/12 - f'(0)f^{(4)}(0)/24}{f''(0)^2/4 - f'(0)f^{(3)}(0)/6}, \\
b_2 &= -\frac{-f^{(3)}(0)^2/36 + f''(0)f^{(4)}(0)/48}{f''(0)^2/4 - f'(0)f^{(3)}(0)/6}.
\end{aligned}
$$

The system output is

$$
\begin{aligned}
y(s) &= \frac{a_2 s^2 + a_1 s + a_0}{s(b_2 s^2 + b_1 s + 1)} \\
&= \frac{a_2(s^2 + p_1 s + p_0)}{b_2 s[(s+q)^2 + \omega^2]}
\end{aligned}
$$

where

$$p_0 = \frac{a_0}{a_2},\ p_1 = \frac{a_1}{a_2},\ q = \frac{b_1}{2b_2},\ \omega^2 = \frac{1}{b_2} - \frac{b_1^2}{4b_2^2}.$$

Evidently, to obtain a sustained oscillation q should be zero, or equivalently, $b_1 = 0$. This implies that the ultimate frequency is

$$\omega_u = \frac{1}{b_2}$$

Then, the ultimate period is

$$\frac{1}{\omega_u} = \sqrt{\frac{\theta(KK_C\theta^3 + 6KK_C\theta^2\tau + 12KK_C\theta\tau^2 + \theta^3 + 6\theta^2\tau + 18\theta\tau^2 + 24\tau^3)}{12(KK_C\theta^2 + 4KK_C\theta\tau + 6KK_C\tau^2 + \theta^2 + 4\theta\tau + 6\tau^2)}}$$

Since

$$b_1 = -\frac{KK_C\theta^3 + 5KK_C\theta^2\tau + 8KK_C\theta\tau^2 - \theta^3 - 5\theta^2\tau - 12\theta\tau^2 - 12\tau^3}{2(KK_C\theta^2 + 4KK_C\theta\tau + 6KK_C\tau^2 + \theta^2 + 4\theta\tau + 6\tau^2)}$$

$b_1 = 0$ (the K_C at this moment is K_u) gives

$$KK_u = \frac{\theta^3 + 5\theta^2\tau + 12\theta\tau^2 + 12\tau^3}{\theta^3 + 5\theta^2\tau + 8\theta\tau^2}$$

The ultimate gain is

$$K_u = \frac{\theta^3 + 5\theta^2\tau + 12\theta\tau^2 + 12\tau^3}{K(\theta^3 + 5\theta^2\tau + 8\theta\tau^2)}$$

It can be seen that the two models can be directly converted into each other. One may use a controller based on the step response model when the ultimate gain and the ultimate frequency are obtained, and vice versa.

Since

$$b_1 = -\frac{KK_C\theta^3 + 5KK_C\theta^2\tau + 8KK_C\theta\tau^2 - \theta^3 - 5\theta^2\tau - 12\theta\tau^2 - 12\tau^3}{2(KK_C\theta^2 + 4KK_C\theta\tau + 6KK_C\tau^2 + \theta^2 + 4\theta\tau + 6\tau^2)}$$

$b_1 = 0$ (the K_C at this moment is K_u) gives

$$KK_u = \frac{\theta^3 + 5\theta^2\tau + 12\theta\tau^2 + 12\tau^3}{\theta^3 + 5\theta^2\tau + 8\theta\tau^2}$$

The ultimate gain is

$$K_u = \frac{\theta^3 + 5\theta^2\tau + 12\theta\tau^2 + 12\tau^3}{K(\theta^3 + 5\theta^2\tau + 8\theta\tau^2)}$$

It can be seen that the two models can be directly converted into each other. One may use a controller based on the step response model when the ultimate gain and the ultimate frequency are obtained, and vice versa.

2.6 Design Requirements and Performance Comparison

Design Requirements

The goal of this section is to clarify some plausible notions in control system design and comparison

Overshoot: For stable plants, the overshoot larger than 50% is rarely accepted. How large the overshoot should be depends on the plant:

- Some plants have a strict limitation on the overshoot
- For some other plants, there is no strict limitation on the overshoot. However, an excessive overshoot implies a large error, which may cause the saturation of the actuator

Rise time: It is desirable that the rise time is fast. However, in the system with an overshoot, the rise time and the overshoot usually conflict. To obtain a proper overshoot, the response speed may have to be sacrificed

Settling time: It is desirable that the settling time is as short as possible. The settling time relates the overshoot and the rise time. Normally, in a system with an overshoot the larger the overshoot, the faster the rise time and the longer the settling time; in a system without an overshoot the faster the rise time, the shorter the settling time

Disturbance response: In most cases, for the system with overshoot the smaller the perturbation peak, the larger the overshoot

Conclusions

The relationship among overshoot, rise time, settling time, integral indices, the reference response, and the disturbance response implies that they are not independent. Once the requirement on one index is given, a restriction is imposed at the same time to other indices.

A desired controller should provide fast and steady response. With respect to this objective, one can take an overshoot of 5%-20% if no other requirements are given. Such an overshoot can, normally, provide a good trade-off between performance and robustness.

Conclusions

The relationship among overshoot, rise time, settling time, integral indices, the reference response, and the disturbance response implies that they are not independent. Once the requirement on one index is given, a restriction is imposed at the same time to other indices.

A desired controller should provide fast and steady response. With respect to this objective, one can take an overshoot of 5%-20% if no other requirements are given. Such an overshoot can, normally, provide a good trade-off between performance and robustness.

Controller Comparison

Purpose of the comparison: Illustrate the advantage of the new method

Existing problem

No unified standard. The comparison is usually unfair

Controller order: If the order of the plant is high or the plant contains time delay, the order of the obtained controller is high. One may have to use model reduction techniques. The order of the controller is interrelated with the closed-loop response. It is certainly easier for a higher-order controller than for a lower-order controller to achieve good performance. In this case, comparison of controllers with different orders may be unfair

Overshoot:

- The controller with good performance, in many cases, has a large overshoot. The reason is that the controller cannot exactly cancel the zeros and poles of the plant when the plant is of high-order or has time delay.
- The performance may not be improved if the overshoot is reduced. A reduced overshoot can usually be obtained by sacrificing the performance (Figure)

For a fair comparison, a common basis is necessary. For example, when the overshoots of different controllers are compared, they should have the same rise time

Robustness consideration: A main problem in robust control is to tradeoff between the nominal performance and the robust stability

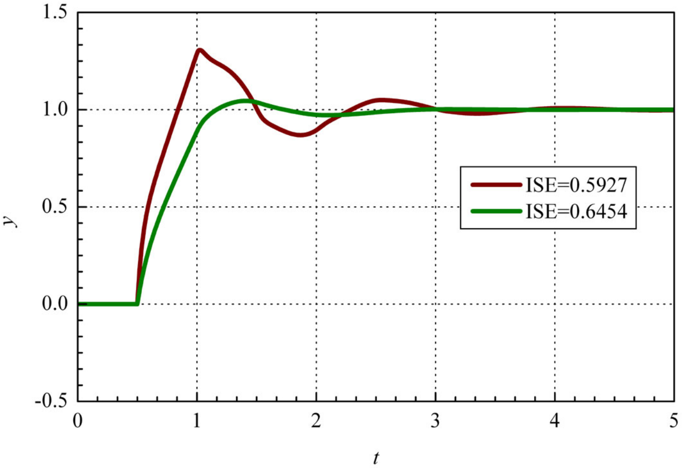


Figure: Conclusions

Nominal performance: The performance of the system with an exact model

Robust performance: The performance of the uncertain system

When the robustness of different systems are compared, it should be considered whether they have the same nominal performance. Good robustness may be obtained by sacrificing the nominal performance

In addition to performance comparison, an important aspect in evaluating a design method is the practicability

Practicability: While the performance specification is reached,

- Whether the design method is easy to use
- whether the design procedure and result is simple enough

Nominal performance: The performance of the system with an exact model

Robust performance: The performance of the uncertain system

When the robustness of different systems are compared, it should be considered whether they have the same nominal performance. Good robustness may be obtained by sacrificing the nominal performance

In addition to performance comparison, an important aspect in evaluating a design method is the practicability

Practicability: While the performance specification is reached,

- Whether the design method is easy to use
- whether the design procedure and result is simple enough

In many applications, the problem of practicability is even more important than the performance problem

An example: In an ultra supercritical power plant there are more than 100 loops. The time for configuration and tuning is very limited. In this case, it is impossible to use complicated methods.

Evaluate the Design Method in This Book

The best way to evaluate the design method in this book may be to apply other design methods to the example given in this book and answer the following questions:

1. With the same plant, which method gives better performance (for example, ISE)?
2. In a real situation, which method is easier to understand and use?
3. Which method can reach practical performance specifications more easily?

End of Chapter 2