University of Washington
Lecture Notes for ME547
Linear Systems

Professor Xu Chen
Bryan T. McMinn Endowed Research Professorship
Associate Professor
Department of Mechanical Engineering
University of Washington

Winter 2025

Abstract: ME547 is a first-year graduate course on modern control systems focusing on: state-space description of dynamic systems, linear algebra for controls, solutions of state-space systems, discrete-time models, stability, controllability and observability, state-feedback control, observers, observer state feedback controls, and when time allows, linear quadratic optimal controls. ME547 is a prerequisite to most advanced graduate control courses in the UW ME department.

"Xu Chen, Lecture Notes for University of Washington Linear Systems (ME547), Available at `https://faculty.washington.edu/chx/teaching/me547/`."

or if you use LaTeX:

```
@Misc{xchen21,
key = {controls,systems,lecture notes},
author = {Xu Chen},
title = {Lecture Notes for University of Washington Linear Systems (ME547)},
howpublished = {Available at \url{https://faculty.washington.edu/chx/teaching/me547/}},
month = January,
year = 2025
}
```

Contents

ME547: Linear Systems

Introduction

Xu Chen

University of Washington

The power of controls

- Our internal body temperature is regulated around 37° C or 98.6° F, whether in a sauna room or outside at the north pole.
- The power of *feedback controls*: it allows us to make a precision device out of a crude one that works well even in changing environments.
- We also use *prediction and feedforward controls*: as kids, we had learned to wear T-shirts in summer, long sleeves and coats in winter. With such predictive and feedforward controls, the burden of feedback control is greatly lifted.

Analysis and control of linear dynamic systems

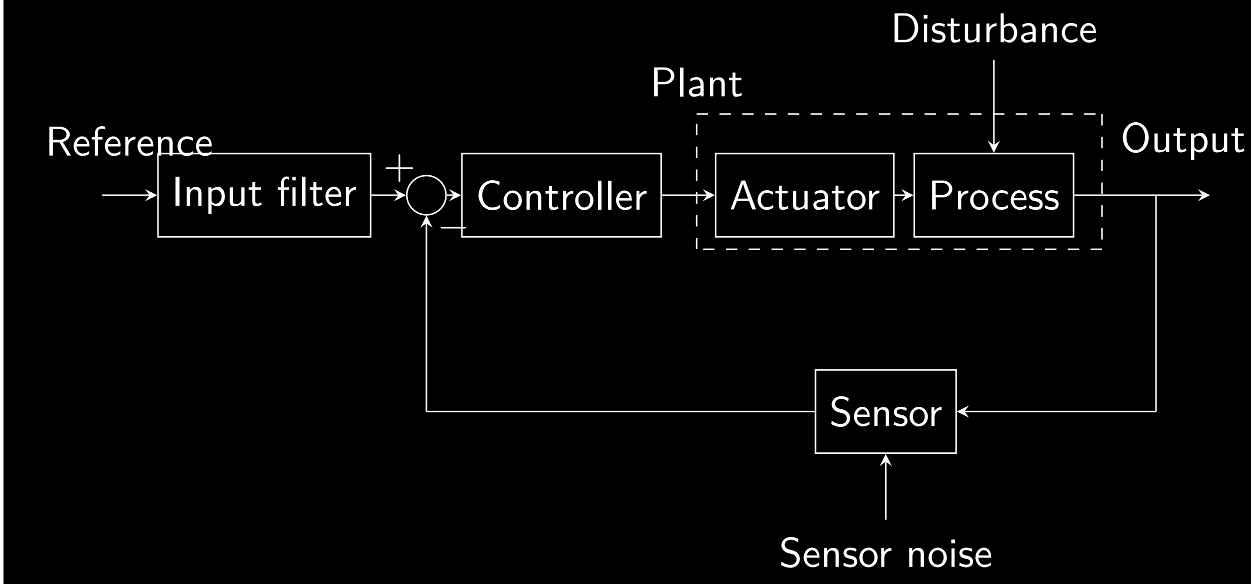
- System: an interconnection of elements and devices for a desired purpose
- Control System: an interconnection of components forming a system configuration that will provide a desired response
- Feedback: the use of information of the past or the present to influence behaviors of a system

Why automatic control?

A system can be either manually or automatically controlled. Why automatic control?

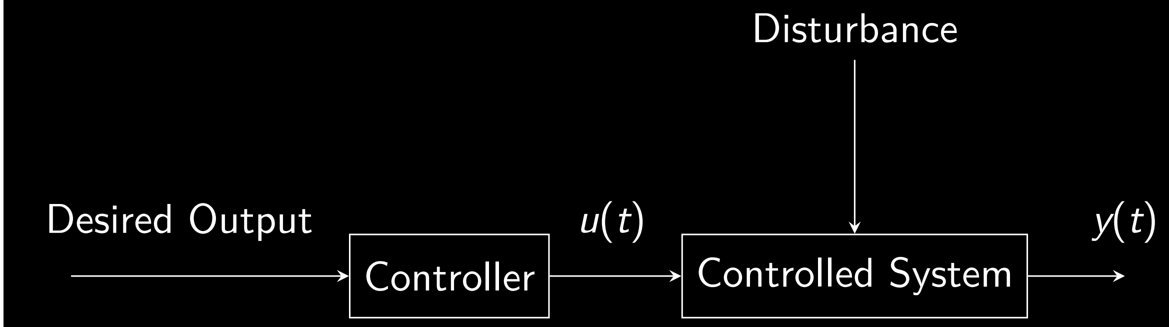
- Stability/Safety: difficult/impossible for humans to control the process or would expose humans to risk
- Performance: cannot be done “as well” by humans
- Cost: Humans are more expensive and can get bored
- Robustness: can deliver the requisite performance even if process behaves slightly differently

Terminologies



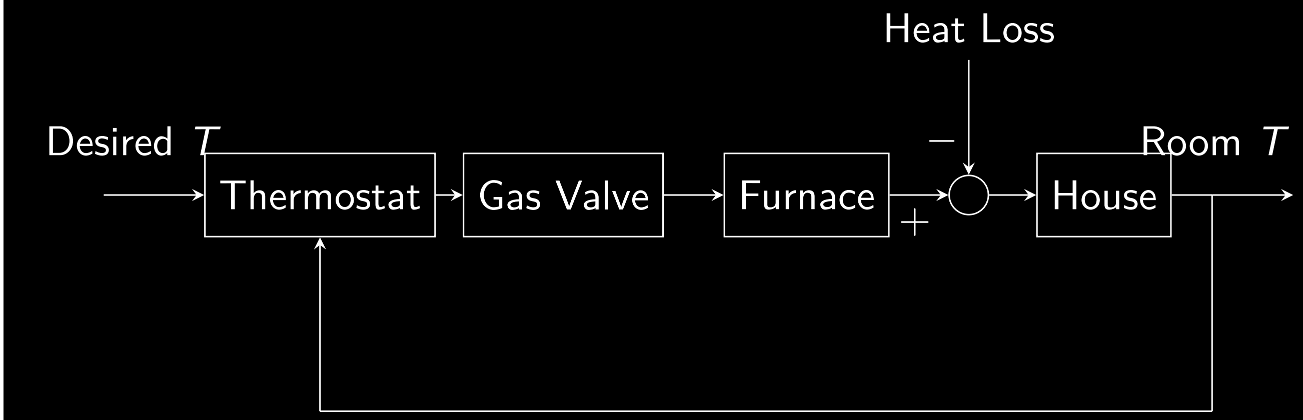
- Process: whose output(s) is/are to be controlled
- Actuator: device to influence the controlled variable of the process
- Plant: process + actuator
- Block diagram: visualizes system structure and the flow information in control systems

Open-loop control v.s. closed-loop control



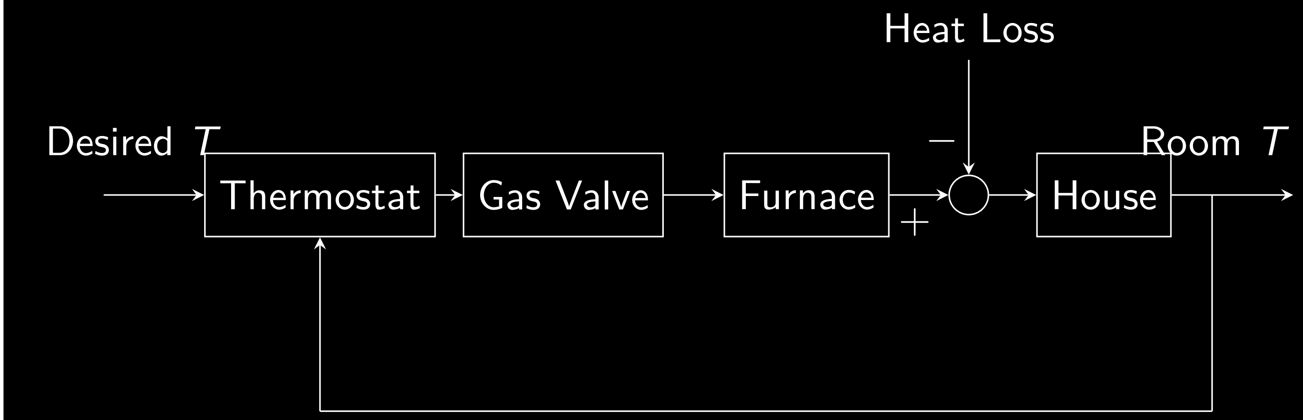
- the output of the plant does not influence the input to the controller
- input and output as *signals*: functions of time, e.g., speed of a car, temperature in a room, voltage applied to a motor, price of a stock, electrical-cardiograph, all as functions of time.

Open-loop control v.s. closed-loop control

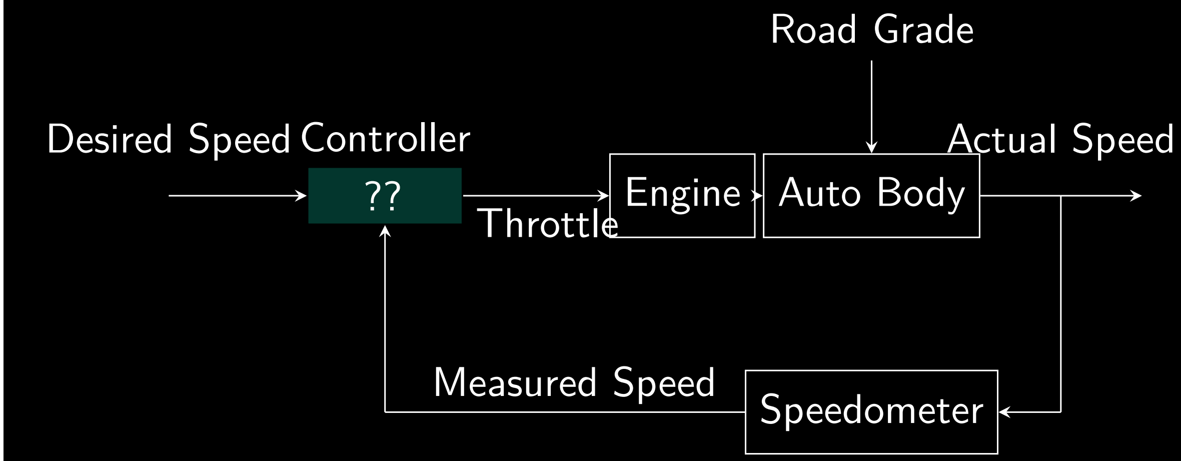


- multiple components (plant, controller, etc) have a closed interconnection
- there is always feedback in a closed-loop system

Closed-loop control: regulation example



Regulation control example: automobile cruise control



- What is the control objective?
- What are the process, process output, actuator, sensor, reference, and disturbance?

Control objectives

- Better stability
- Improved response characteristics
- *Regulation* of output in the presence of disturbances and noises
- Robustness to plant uncertainties
- *Tracking* time varying desired output

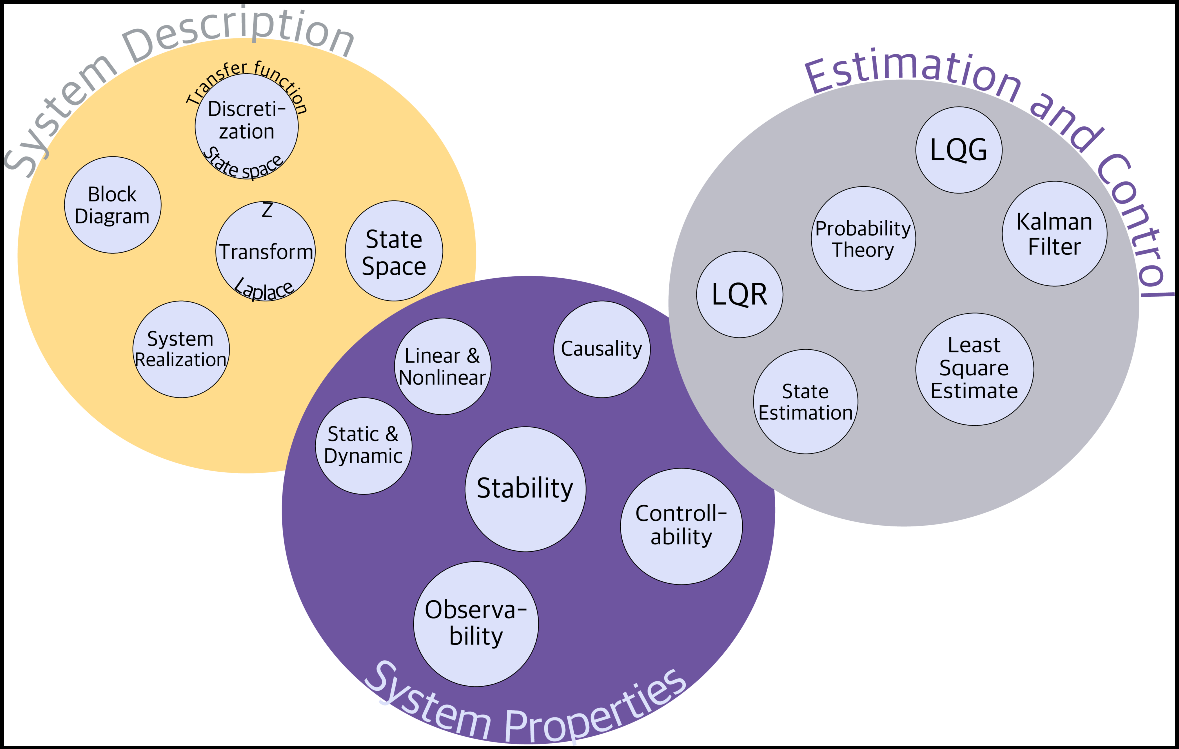
There are some aspects of control objectives that are universal. For example, we would always want our control system to result in closed-loop dynamics that are insensitive to disturbances. This is the disturbance rejection problem. Also, as pointed out previously, we would want the controller to be robust to plant modeling errors.

Means to achieve the control objectives

- Model the controlled plant
- Analyze the characteristics of the plant
- Design control algorithms (controllers)
- Analyze performance and robustness of the control system
- Implement the controller

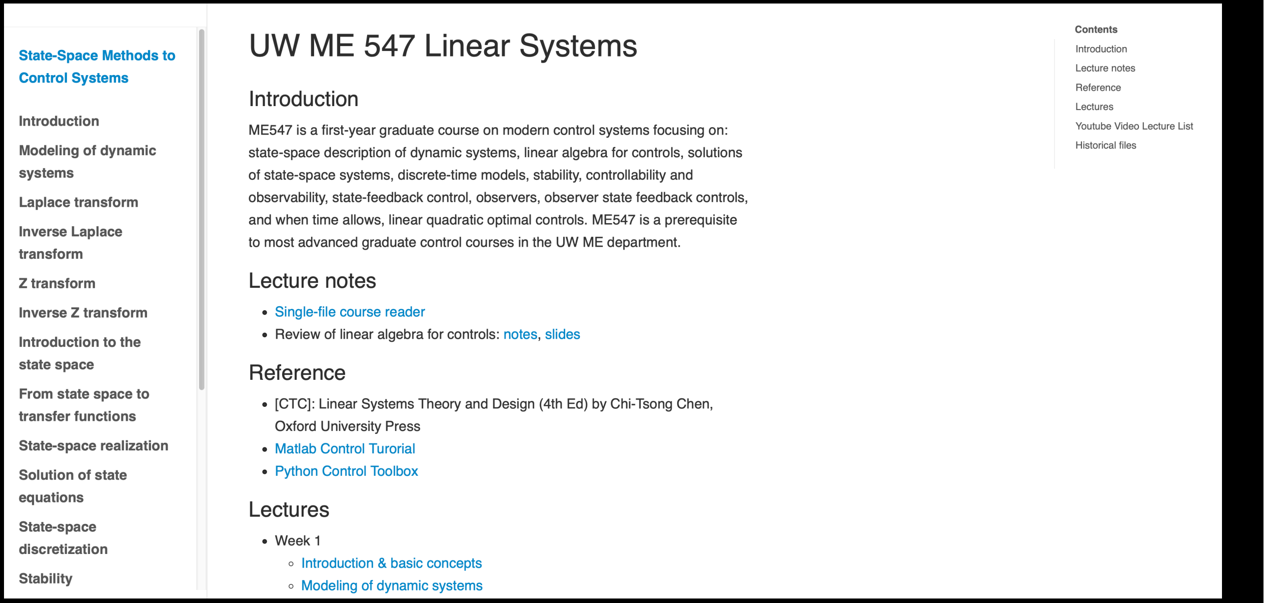
About this course

- a first-year graduate course on modern control systems



Website

- lectures: https://faculty.washington.edu/chx/teaching/me547/



State-Space Methods to Control Systems

Introduction

Modeling of dynamic systems

Laplace transform

Inverse Laplace transform

Z transform

Inverse Z transform

Introduction to the state space

From state space to transfer functions

State-space realization

Solution of state equations

State-space discretization

Stability

UW ME 547 Linear Systems

Introduction

ME547 is a first-year graduate course on modern control systems focusing on: state-space description of dynamic systems, linear algebra for controls, solutions of state-space systems, discrete-time models, stability, controllability and observability, state-feedback control, observers, observer state feedback controls, and when time allows, linear quadratic optimal controls. ME547 is a prerequisite to most advanced graduate control courses in the UW ME department.

Lecture notes

- Single-file course reader
- Review of linear algebra for controls: notes, slides

Reference

- [CTC]: Linear Systems Theory and Design (4th Ed) by Chi-Tsong Chen, Oxford University Press
- Matlab Control Turorial
- Python Control Toolbox

Lectures

- Week 1
  - Introduction & basic concepts
  - Modeling of dynamic systems

Contents

Introduction

Lecture notes

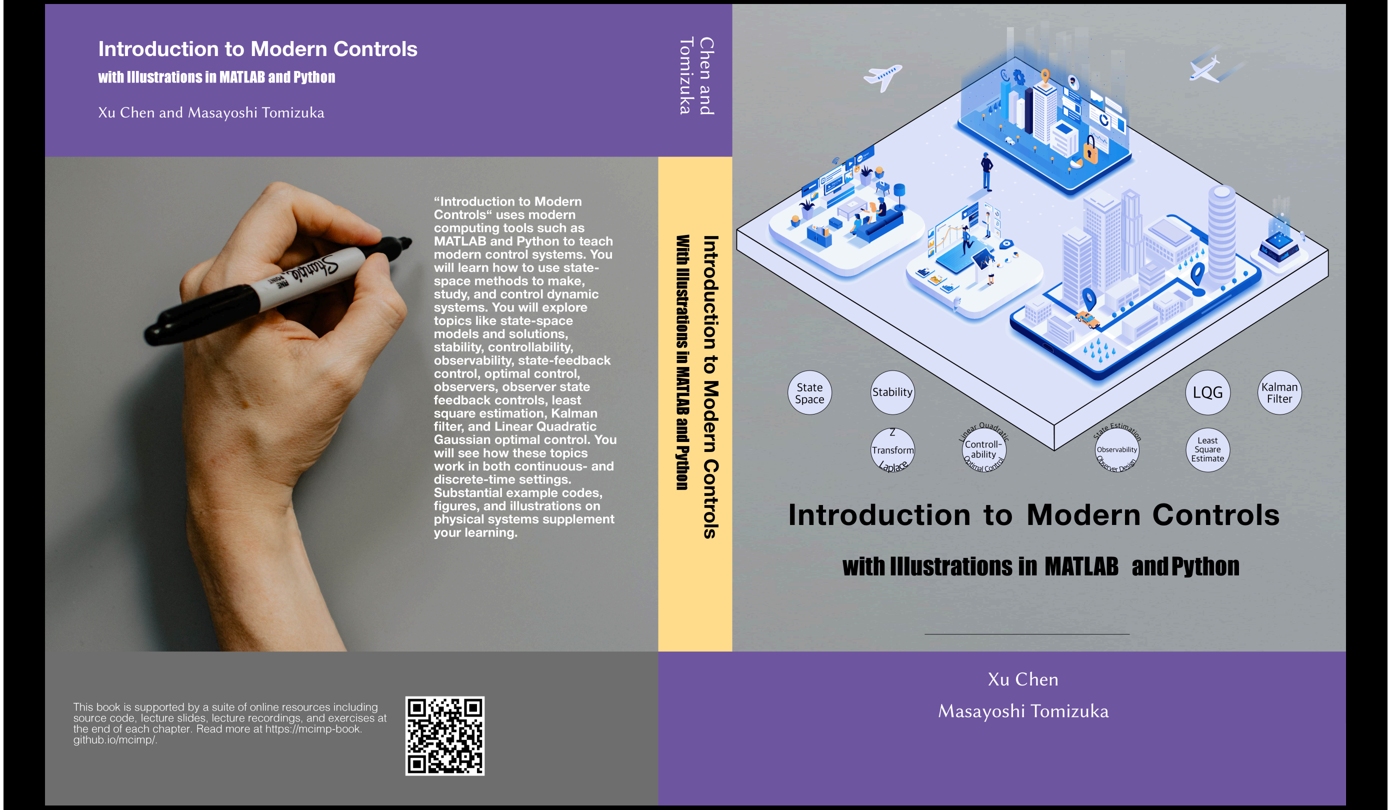
Reference

Lectures

Youtube Video Lecture List

Historical files

Textbook



Introduction to Modern Controls

with Illustrations in MATLAB and Python

Xu Chen and Masayoshi Tomizuka

"Introduction to Modern Controls" uses modern computing tools such as MATLAB and Python to teach modern control systems. You will learn how to use state-space methods to make, study, and control dynamic systems. You will explore topics like state-space models and solutions, stability, controllability, observability, state-feedback control, optimal control, observers, observer state feedback controls, least square estimation, Kalman filter, and Linear Quadratic Gaussian optimal control. You will see how these topics work in both continuous- and discrete-time settings. Substantial example codes, figures, and illustrations on physical systems supplement your learning.

This book is supported by a suite of online resources including source code, lecture slides, lecture recordings, and exercises at the end of each chapter. Read more at https://mcimp-book.github.io/mcimp/.

Chen and Tomizuka

Introduction to Modern Controls

With Illustrations in MATLAB and Python

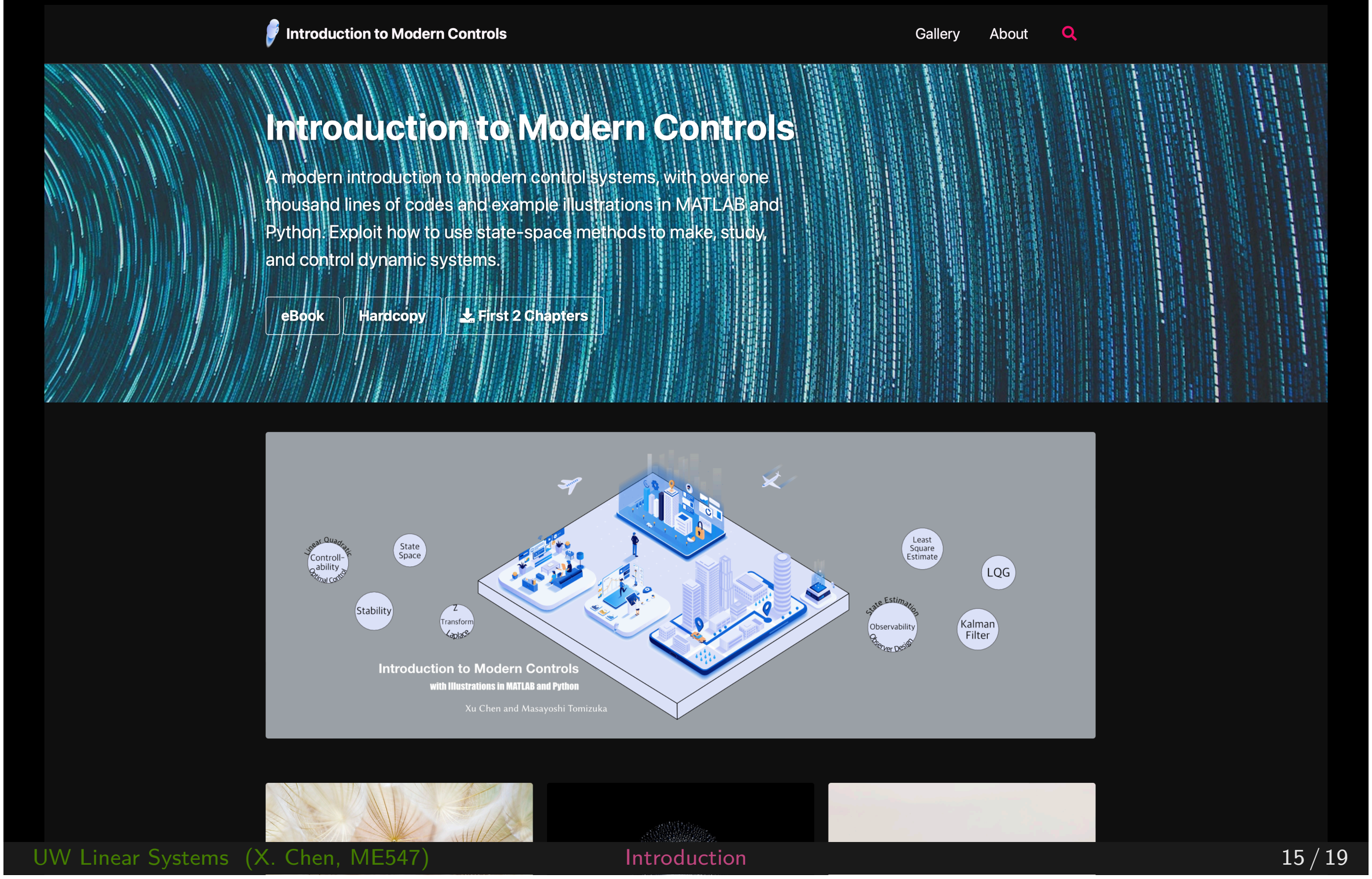
Introduction to Modern Controls

with Illustrations in MATLAB and Python

Xu Chen

Masayoshi Tomizuka

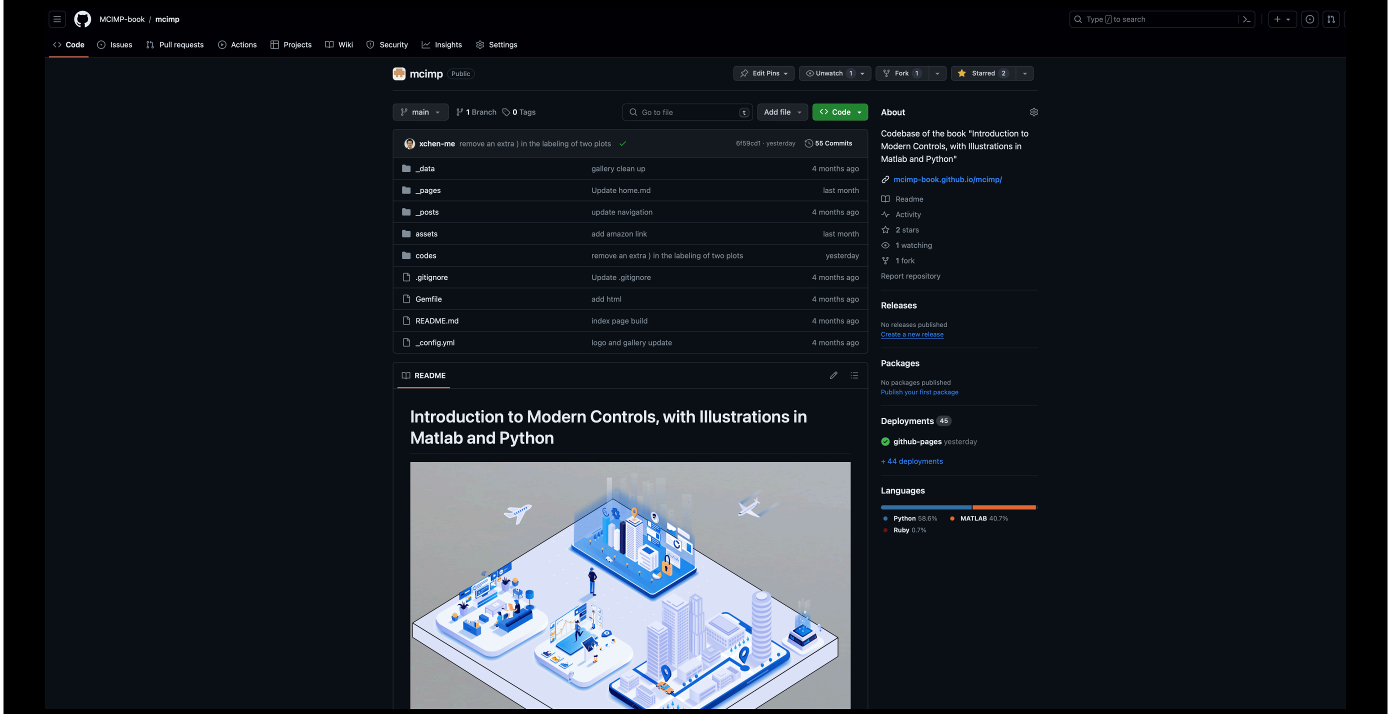
Textbook



Introduction to Modern Controls

A modern introduction to modern control systems, with over one thousand lines of codes and example illustrations in MATLAB and Python. Exploit how to use state-space methods to make, study, and control dynamic systems.

Textbook



xchen-me remove an extra) in the labeling of two plots ✓ 6f59cd1 · yesterday 55 Commits

_data	gallery clean up	4 months ago
_pages	Update home.md	last month
_posts	update navigation	4 months ago
assets	add amazon link	last month
codes	remove an extra) in the labeling of two plots	yesterday
.gitignore	Update .gitignore	4 months ago
Gemfile	add html	4 months ago
README.md	index page build	4 months ago
_config.yml	logo and gallery update	4 months ago

README

Introduction to Modern Controls, with Illustrations in Matlab and Python

About

Codebase of the book "Introduction to Modern Controls, with Illustrations in Matlab and Python"

mcimp-book.github.io/mcimp/

Readme
Activity
2 stars
1 watching
1 fork
Report repository

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Deployments 45

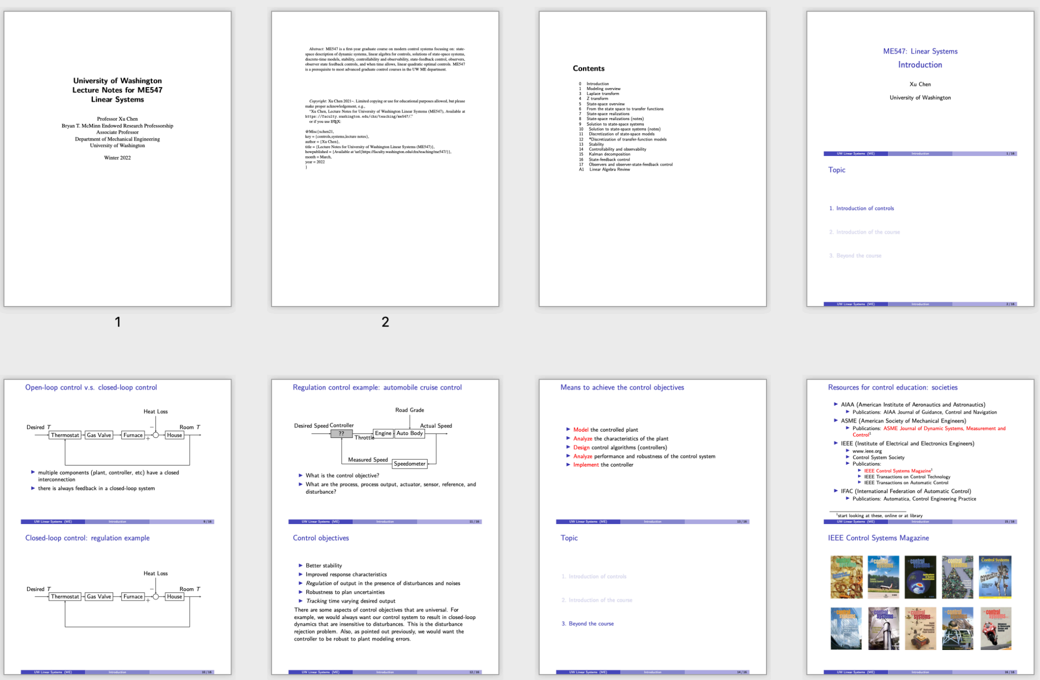
github-pages yesterday
+ 44 deployments

Languages

Python 58.6% MATLAB 40.7%
Ruby 0.7%

Written materials

- Open-source Course Notes



University of Washington
Lecture Notes for ME547
Linear Systems

Professor Xu Chen
Bryan T. McMinn Endowed Research Professorship
Associate Professor
Department of Mechanical Engineering
University of Washington

Winter 2022

1

Abstract: ME547 is a first-year graduate course on modern control systems focusing on: state-space description of dynamic systems, linear algebra for controls, solutions of state-space systems, discrete-time models, stability, controllability and observability, state-feedback control, observers, observer state feedback controls, and when time allows, linear quadratic optimal controls. ME547 is a prerequisite to most advanced graduate control courses in the UW ME department.

Copyright: Xu Chen 2021~. Limited copying or use for educational purposes allowed, but please make proper acknowledgement, e.g.,
"Xu Chen, Lecture Notes for University of Washington Linear Systems (ME547), Available at `https://faculty.washington.edu/chx/teaching/me547/`."
or if you use LaTeX:

```
@Misc{xchen21,
key = {controls,systems,lecture notes},
author = {Xu Chen},
title = {Lecture Notes for University of Washington Linear Systems (ME547)},
howpublished = {Available at \url{https://faculty.washington.edu/chx/teaching/me547/}},
month = March,
year = 2022
}
```

2

Contents

ME547: Linear Systems
Introduction

Xu Chen

University of Washington

UW Linear Systems (ME) Introduction 1 / 16

Topic

1. Introduction of controls
2. Introduction of the course
3. Beyond the course

UW Linear Systems (ME) Introduction 2 / 16

Open-loop control v.s. closed-loop control

- multiple components (plant, controller, etc) have a closed interconnection
- there is always feedback in a closed-loop system

UW Linear Systems (ME) Introduction 9 / 16

Closed-loop control: regulation example

UW Linear Systems (ME) Introduction 10 / 16

Regulation control example: automobile cruise control

- What is the control objective?
- What are the process, process output, actuator, sensor, reference, and disturbance?

UW Linear Systems (ME) Introduction 11 / 16

Control objectives

- Better stability
- Improved response characteristics
- *Regulation* of output in the presence of disturbances and noises
- Robustness to plan uncertainties
- *Tracking* time varying desired output

There are some aspects of control objectives that are universal. For example, we would always want our control system to result in closed-loop dynamics that are insensitive to disturbances. This is the disturbance rejection problem. Also, as pointed out previously, we would want the controller to be robust to plant modeling errors.

UW Linear Systems (ME) Introduction 12 / 16

Means to achieve the control objectives

- Model the controlled plant
- Analyze the characteristics of the plant
- Design control algorithms (controllers)
- Analyze performance and robustness of the control system
- Implement the controller

UW Linear Systems (ME) Introduction 13 / 16

Topic

1. Introduction of controls
2. Introduction of the course
3. Beyond the course

UW Linear Systems (ME) Introduction 14 / 16

Resources for control education: societies

- AIAA (American Institute of Aeronautics and Astronautics)
  - Publications: AIAA Journal of Guidance, Control and Navigation
- ASME (American Society of Mechanical Engineers)
  - Publications: ASME Journal of Dynamic Systems, Measurement and Control[1]
- IEEE (Institute of Electrical and Electronics Engineers)
  - www.ieee.org
  - Control System Society
  - Publications:
    - IEEE Control Systems Magazine[1]
    - IEEE Transactions on Control Technology
    - IEEE Transactions on Automatic Control
- IFAC (International Federation of Automatic Control)
  - Publications: Automatica, Control Engineering Practice

[1] start looking at these, online or at library

UW Linear Systems (ME) Introduction 15 / 16

IEEE Control Systems Magazine

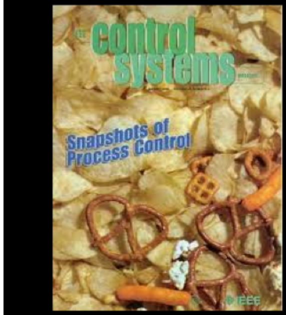
UW Linear Systems (ME) Introduction 16 / 16

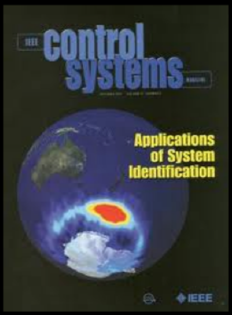
Resources for control education: societies

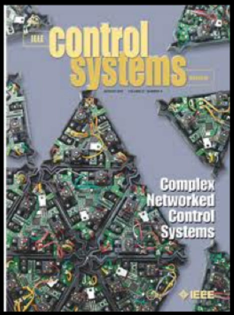
- AIAA (American Institute of Aeronautics and Astronautics)
  - Publications: AIAA Journal of Guidance, Control and Navigation
- ASME (American Society of Mechanical Engineers)
  - Publications: ASME Journal of Dynamic Systems, Measurement and Control[1]
- IEEE (Institute of Electrical and Electronics Engineers)
  - www.ieee.org
  - Control System Society
  - Publications:
    - IEEE Control Systems Magazine[1]
    - IEEE Transactions on Control Technology
    - IEEE Transactions on Automatic Control
- IFAC (International Federation of Automatic Control)
  - Publications: Automatica, Control Engineering Practice

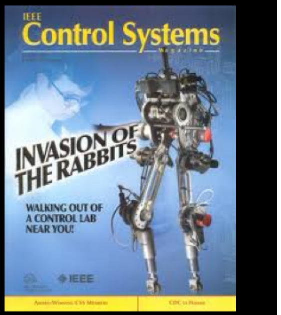
[1] start looking at these, online or at library

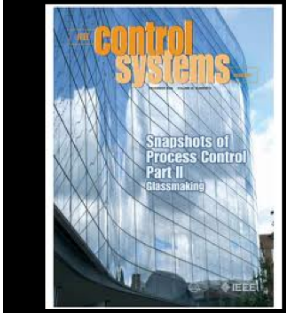
IEEE Control Systems Magazine

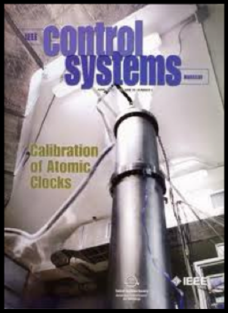


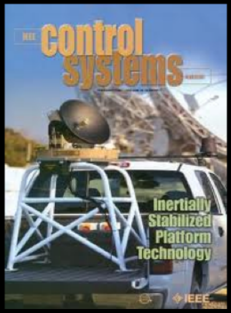


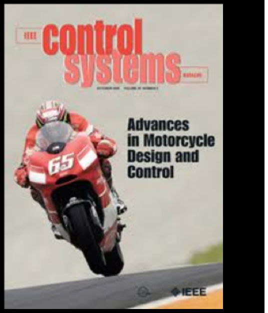












ME547: Linear Systems

Modeling of Dynamic Systems

Xu Chen

University of Washington

Why modeling?

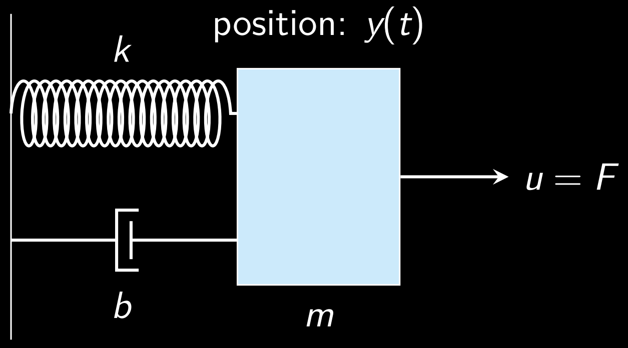
Modeling of physical systems:

- a vital component of modern engineering
- often consists of complex coupled differential equations
- only when we have good understanding of a system can we optimally control it:
  - can simulate and predict actual system response, and
  - design model-based controllers

Two general approaches of modeling

- based on physics:
  - using fundamental engineering principles such as Newton's laws, energy conservation, etc
- based on measurement data:
  - using input-output response of the system
  - a field itself known as system identification
- often the tools are combined in practice

Example: Mass spring damper

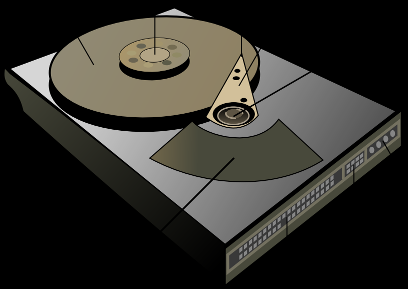


Newton's second law gives

$$m\ddot{y}(t) + b\dot{y}(t) + ky(t) = u(t)\,,\ \ y(0) = y_0,\ \ \dot{y}(0) = \dot{y}_0$$

- modeled as a second-order ODE with input $u(t)$ and output $y(t)$

Example: HDD



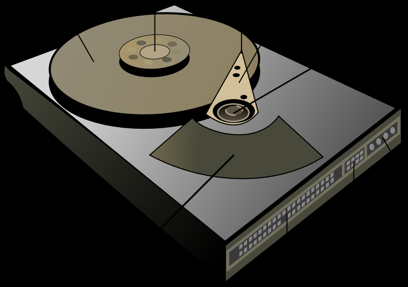
- Newton's second law for rotation

$$\underbrace{\sum_i \tau_i}_{\text{net torque}} = \underbrace{J}_{\text{moment of inertia}} \underbrace{\alpha}_{\text{angular acceleration}}$$

- letting θ :=output and τ :=input yields

$$\ddot{\theta} = \alpha = \frac{1}{J}\tau$$

Example: HDD



$$\ddot{\theta} = \alpha = \frac{1}{J}\tau \Leftrightarrow \Theta(s) = \frac{1}{Js^2}\mathrm{T}(s)$$

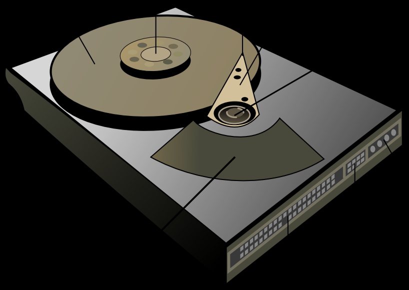
- with damping:

$$\ddot{\theta} + 2\zeta\omega_n\dot{\theta} + \omega_n^2\theta = \kappa\tau \Leftrightarrow \Theta(s) = \frac{\kappa}{s^2 + 2\zeta\omega_n s + \omega_n^2}\mathrm{T}(s)$$

- with multiple modes:

$$\ddot{\theta}_i + 2\zeta_i\omega_i\dot{\theta}_i + \omega_i^2\theta_i = \kappa_i\tau \Leftrightarrow \Theta_i(s) = \frac{\kappa_i}{s^2 + 2\zeta_i\omega_i s + \omega_i^2}\mathrm{T}(s)$$

Example: HDD

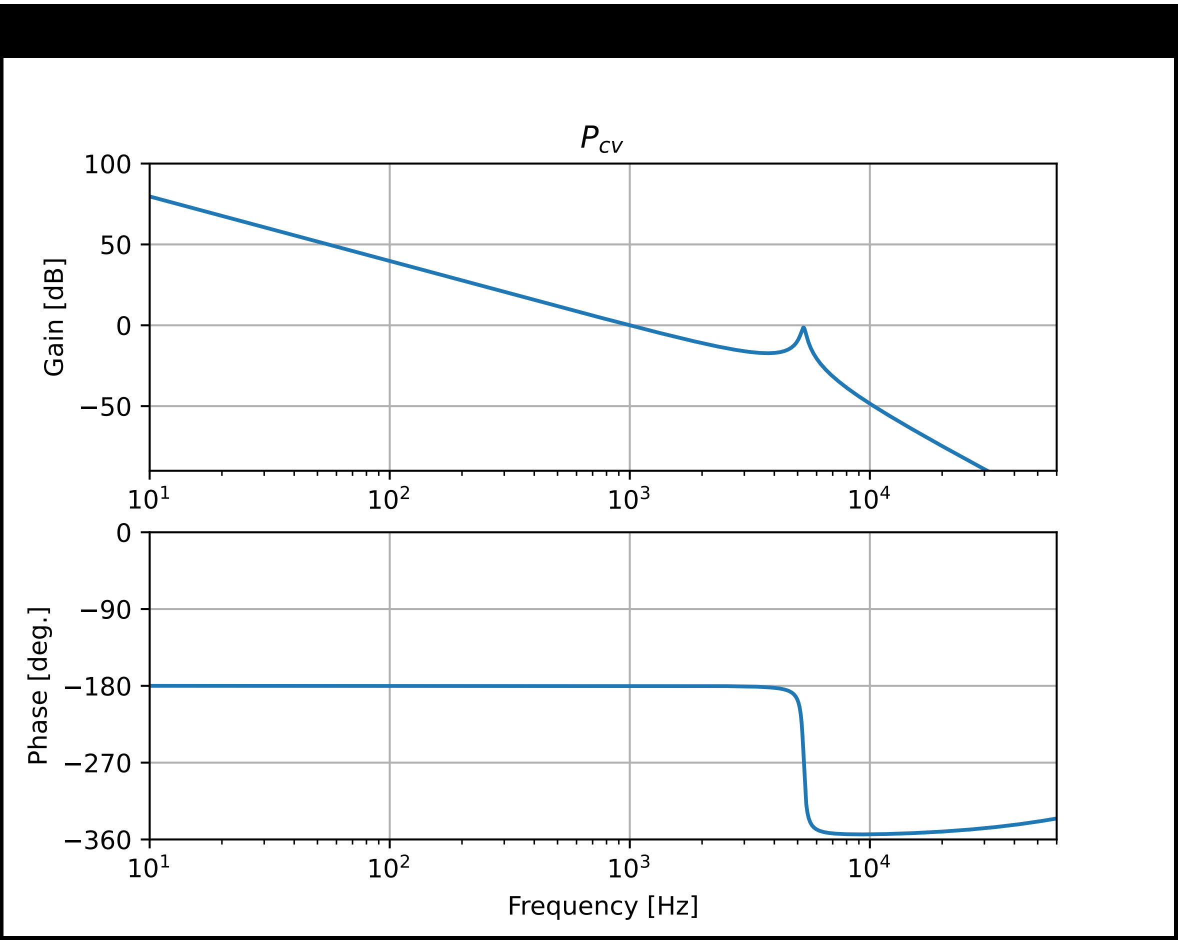


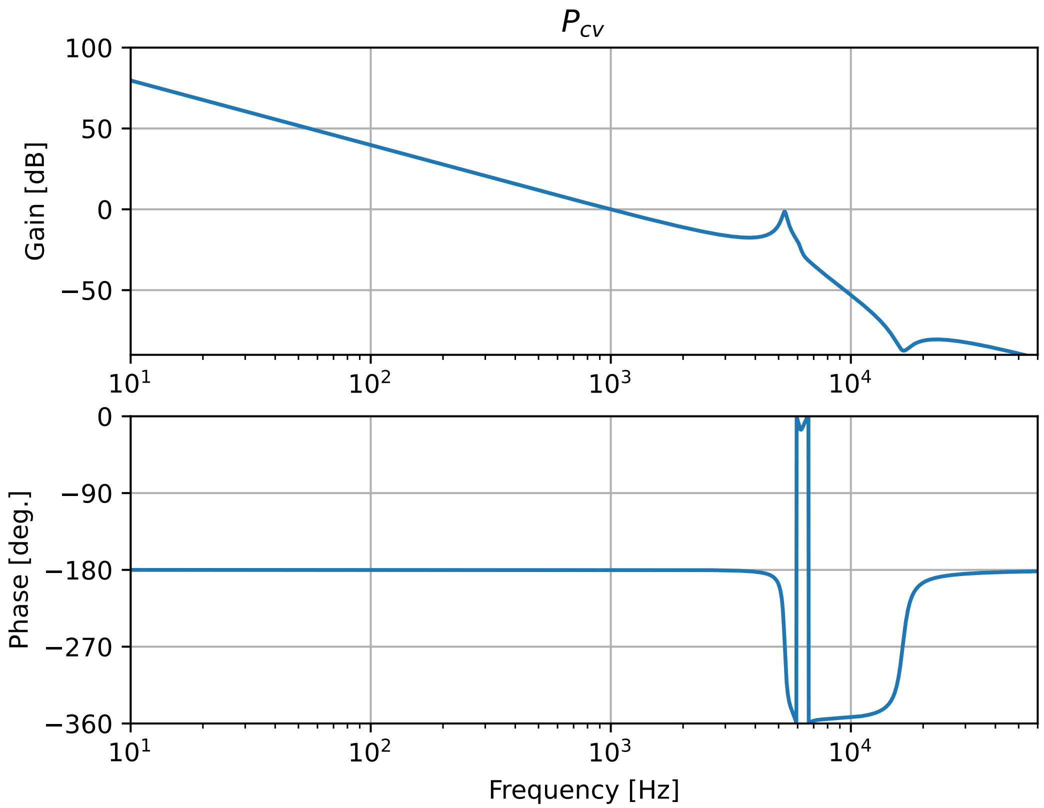
$$\ddot{\theta}_i + 2\zeta_i\omega_i\dot{\theta}_i + \omega_i^2\theta_i = \kappa_i\tau \Leftrightarrow \Theta_i(s) = \frac{\kappa_i}{s^2 + 2\zeta_i\omega_i s + \omega_i^2}\mathrm{T}(s)$$

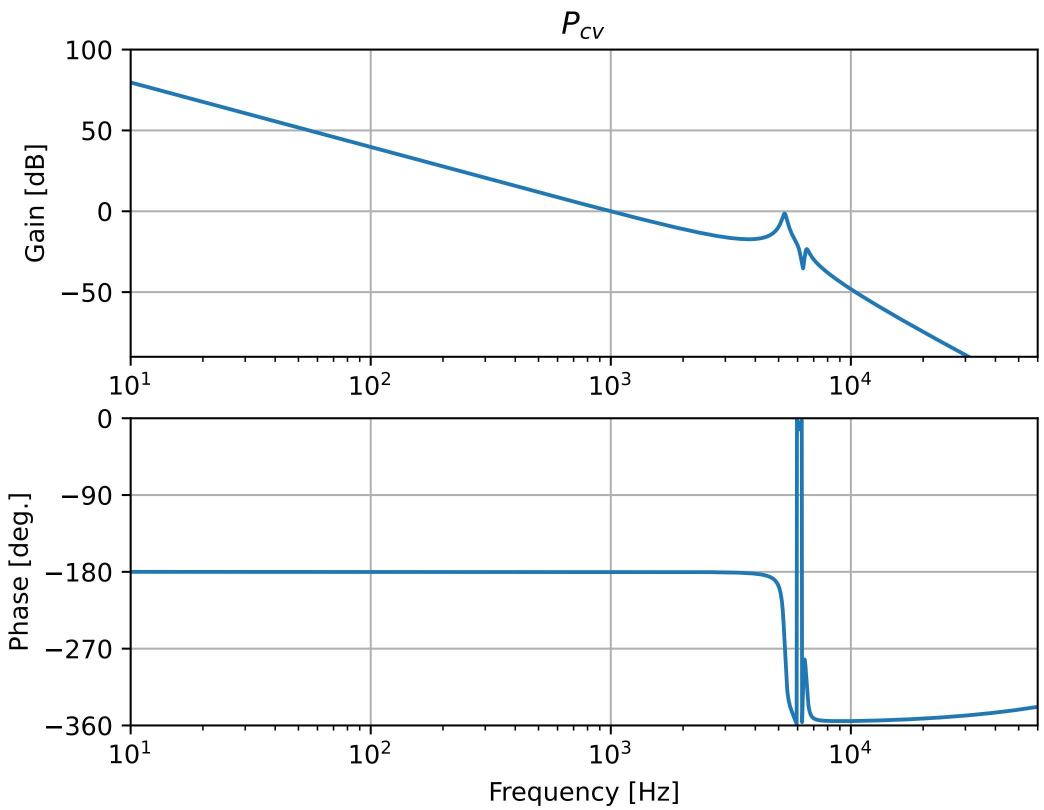
- final model:

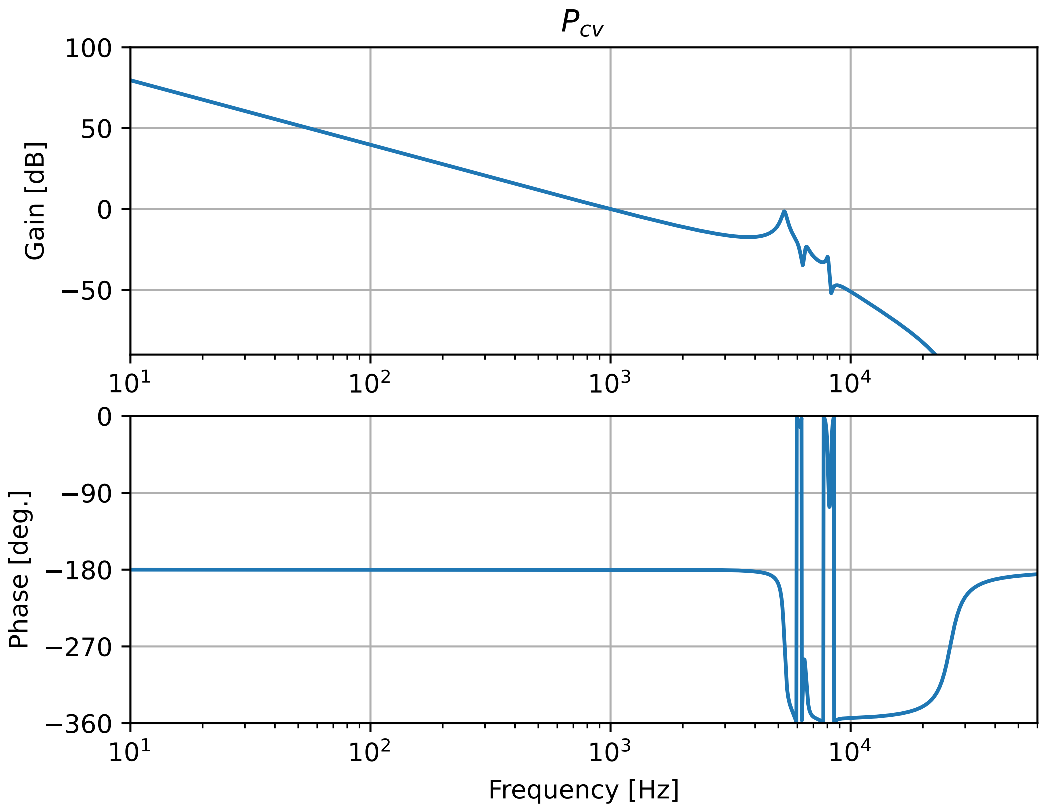
$$\Theta(s) = \sum_{i=1}^{n} \frac{\kappa_i}{s^2 + 2\zeta_i\omega_i s + \omega_i^2}\mathrm{T}(s)$$

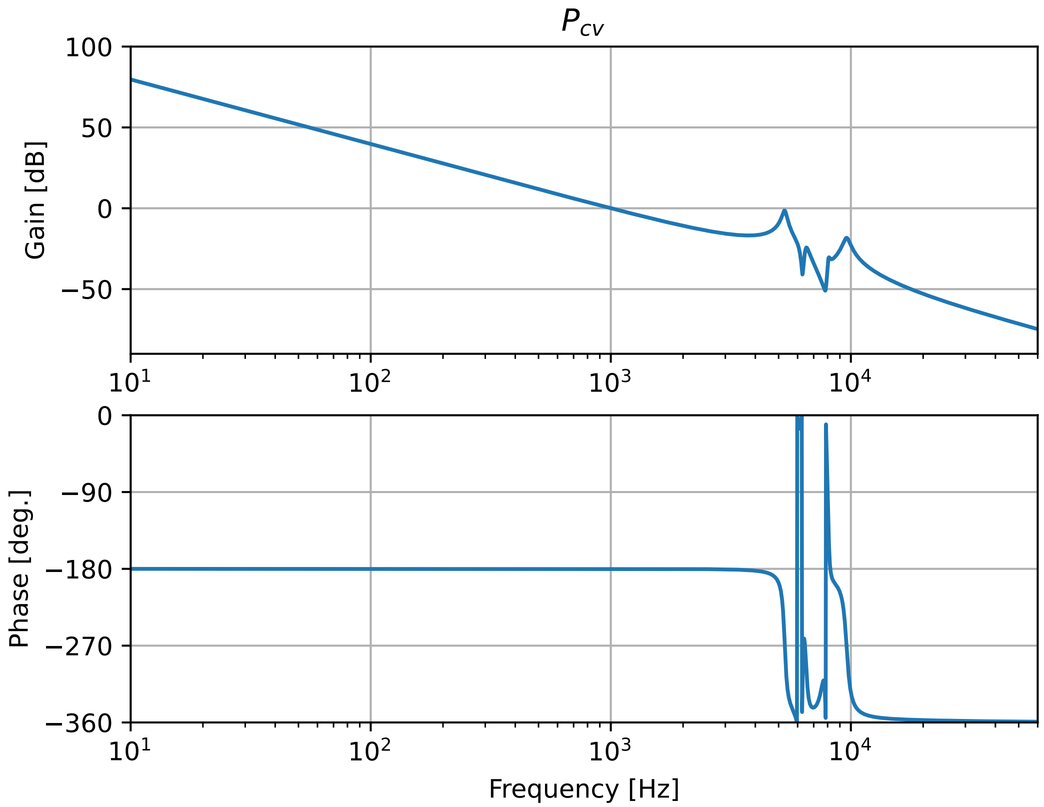
```
import numpy as np
import matplotlib.pyplot as plt
from scipy import signal
import control as ct
num_sector = 420  # Number of sector
num_rpm = 7200  # Number of RPM
Kp_vcm = 3.7976e+07  # VCM gain
omega_vcm = np.array([0, 5300, 6100, 6500, 8050, 9600, 14800, 17400,
                     21000, 26000, 26600, 29000, 32200, 38300, 43300,
                     ↪  44800]) * 2 * np.pi
kappa_vcm = np.array([1, -1.0, +0.1, -0.1, 0.04, -0.7, -
                     0.2, -1.0, +3.0, -3.2, 2.1, -1.5, +2.0, -0.2,
                     ↪  +0.3, -0.5])
zeta_vcm = np.array([0, 0.02, 0.04, 0.02, 0.01, 0.03, 0.01,
                    0.02, 0.02, 0.012, 0.007, 0.01, 0.03, 0.01, 0.01,
                    ↪  0.01])
Sys_Pc_vcm_c1 = ct.TransferFunction([], [1])  # Create an empty
↪  transfer function
for i in range(len(omega_vcm)):
    Sys_Pc_vcm_c1 = Sys_Pc_vcm_c1 + ct.TransferFunction(np.array(
        [0, 0, kappa_vcm[i]]) * Kp_vcm, np.array([1, 2 * zeta_vcm[i] *
        ↪  omega_vcm[i], (omega_vcm[i]) ** 2]))
```

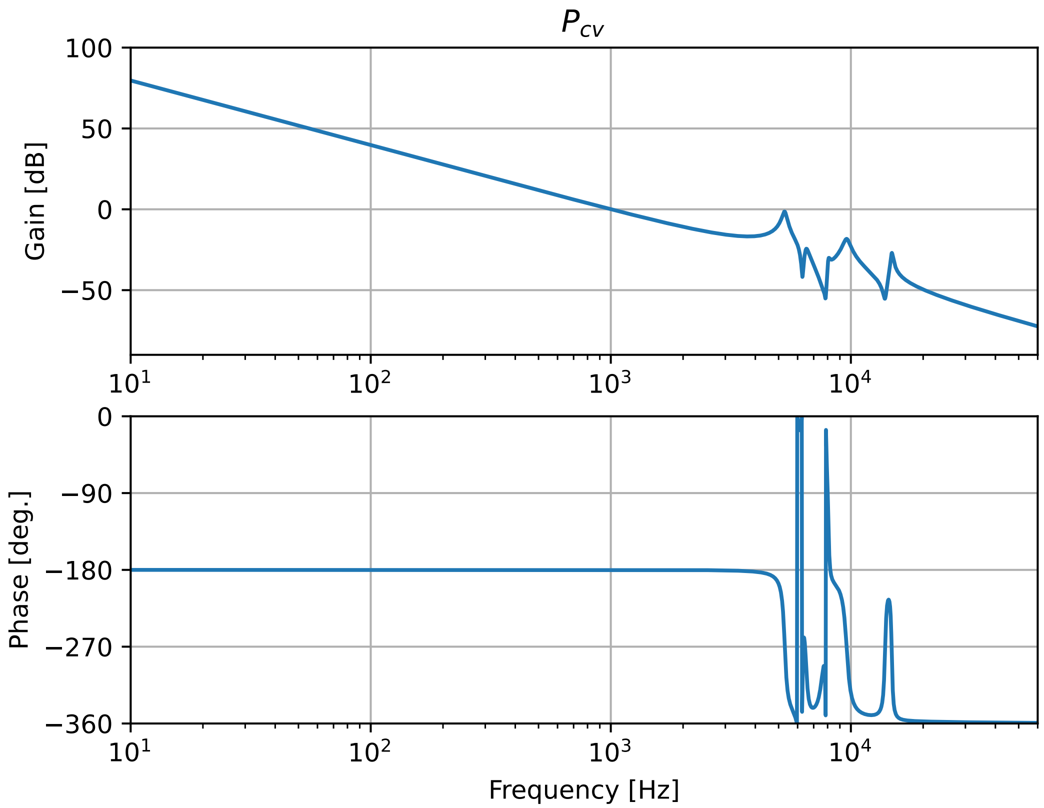

P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]

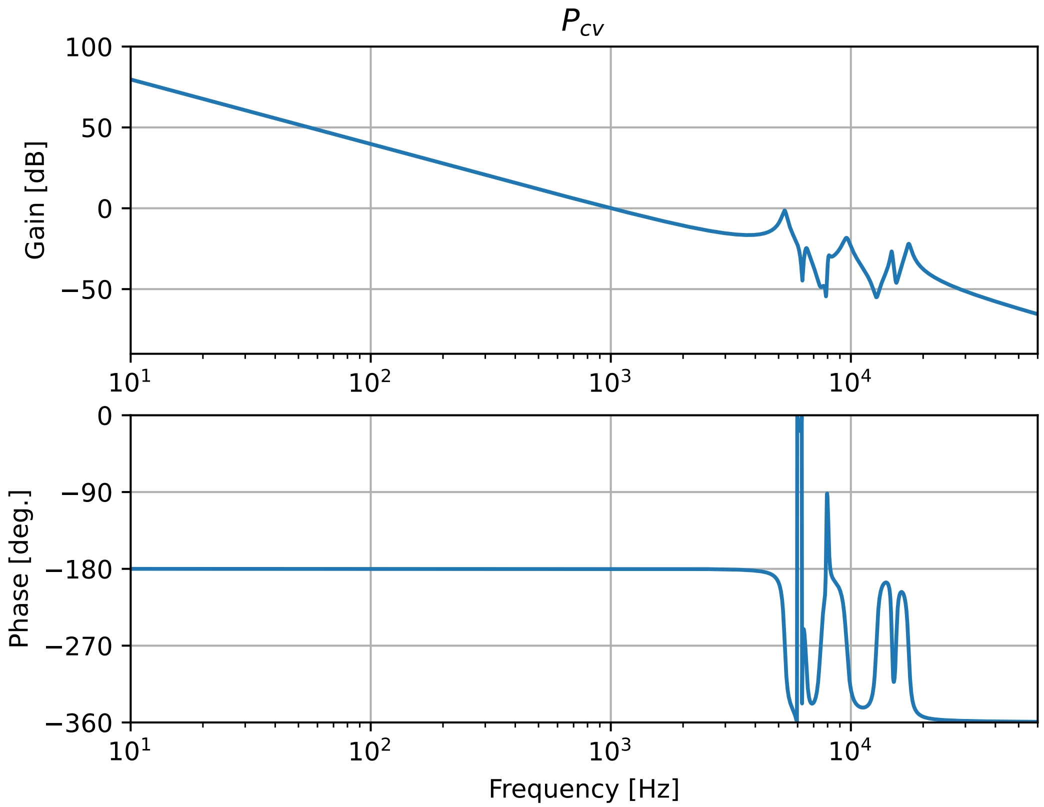

P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]

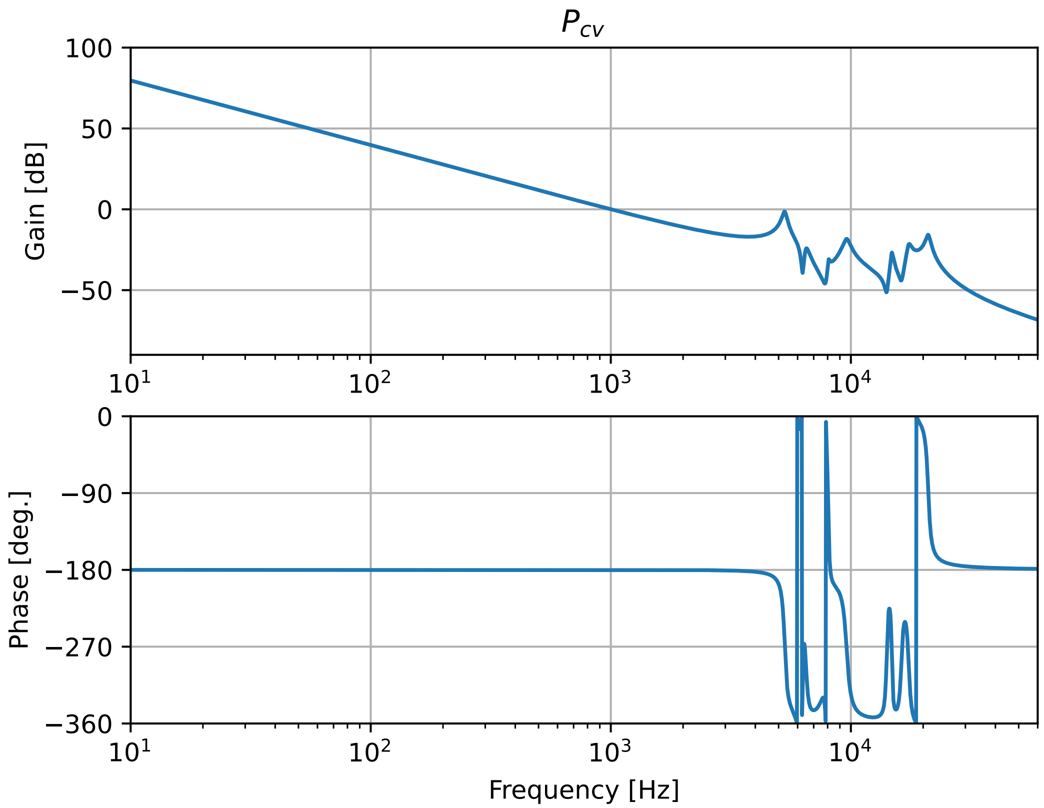

P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]

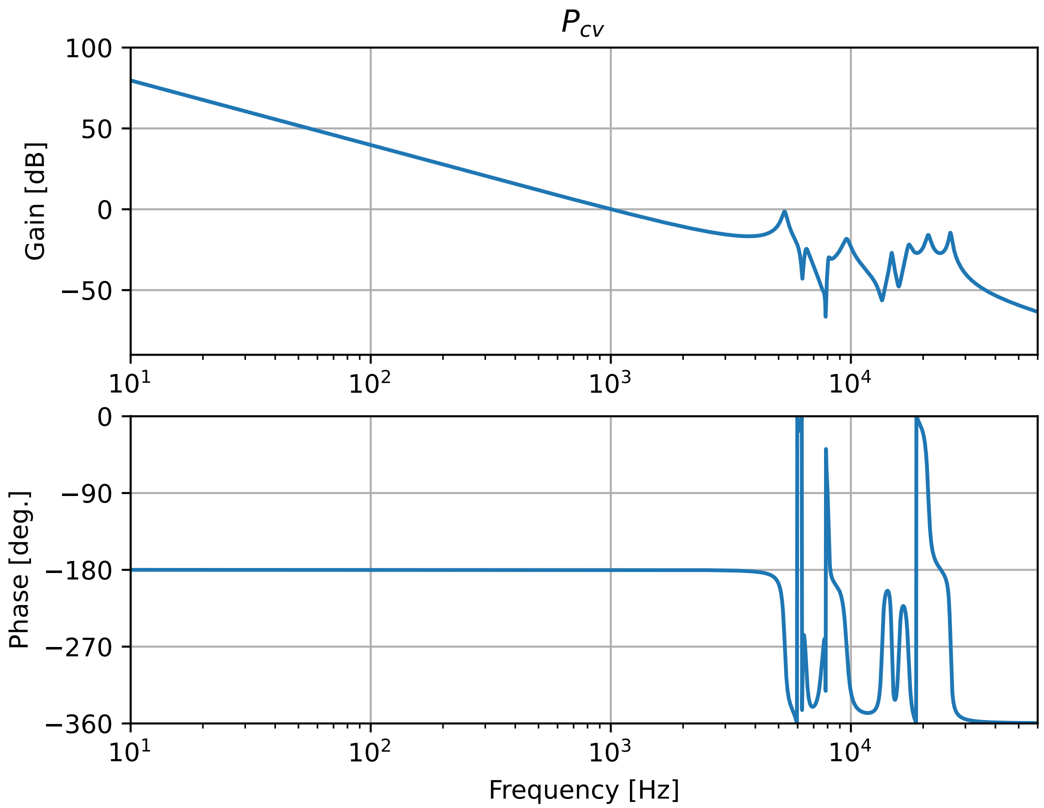

P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]

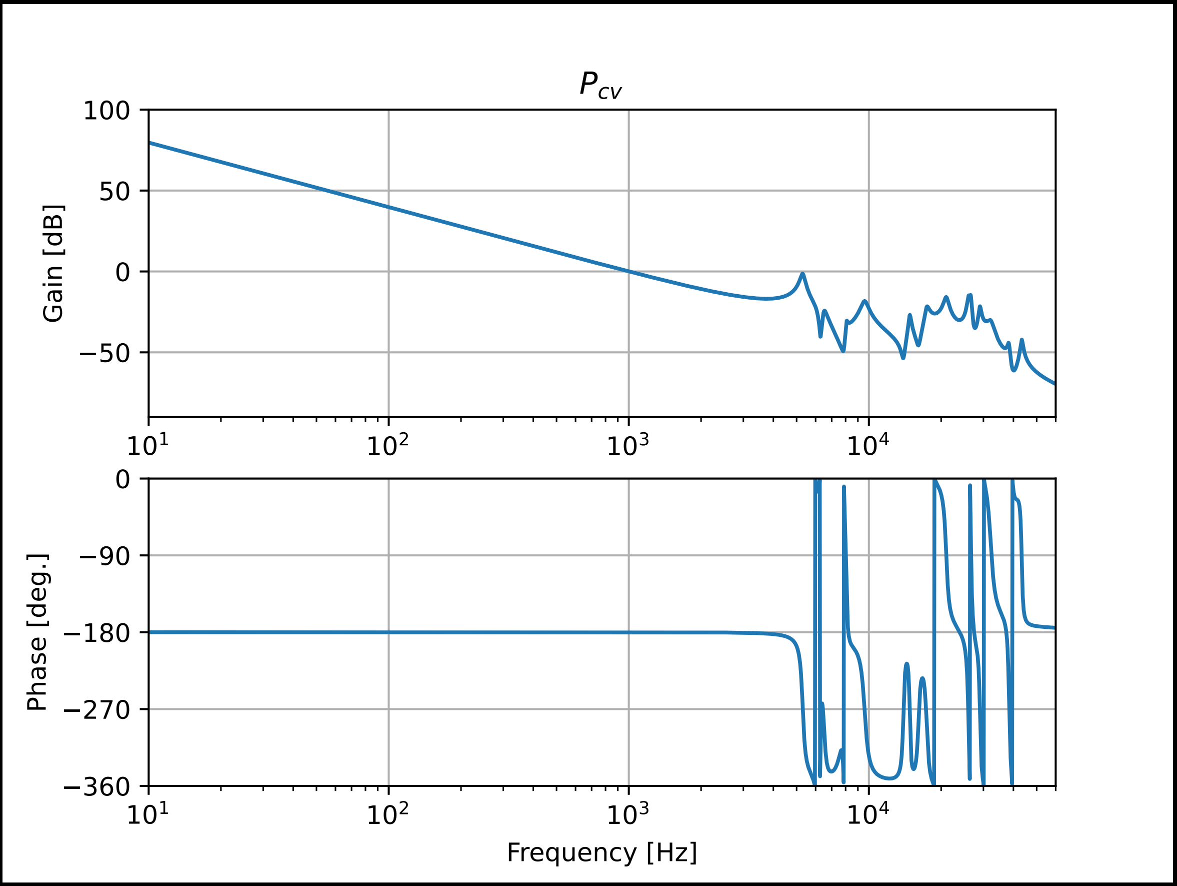

P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]


P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]


P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]

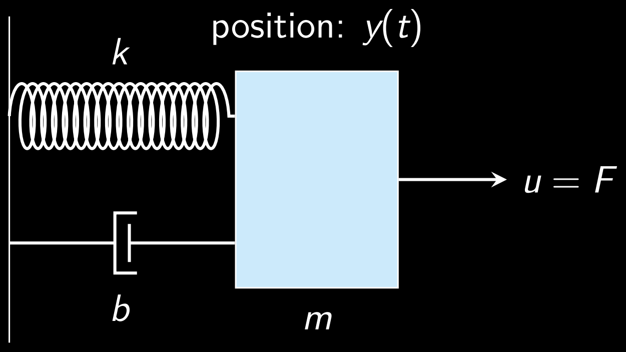

P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]


P_{cv}
Gain [dB]
100
50
0
−50
10^1
10^2
10^3
10^4
Phase [deg.]
0
−90
−180
−270
−360
10^1
10^2
10^3
10^4
Frequency [Hz]

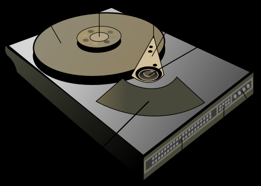


Models of continuous-time systems

- modeled as differential equations:



$$m\ddot{y}(t) + b\dot{y}(t) + ky(t) = u(t)\,,\ y(0) = y_0,\ \dot{y}(0) = \dot{y}_0$$



$$\ddot{\theta}_i + 2\zeta_i\omega_i\dot{\theta}_i + \omega_i^2\theta_i = \kappa_i\tau \Leftrightarrow \Theta_i(s) = \frac{\kappa_i}{s^2 + 2\zeta_i\omega_i s + \omega_i^2}\mathrm{T}(s)$$

Models of continuous-time systems

General continuous-time systems:

$$\frac{d^n y(t)}{dt^n} + a_{n-1}\frac{d^{n-1}y(t)}{dt^{n-1}} + \cdots + a_0 y(t) = b_m \frac{d^m u(t)}{dt^m} + b_{m-1}\frac{d^{m-1}u(t)}{dt^{m-1}} + \cdots + b_0 u(t)$$

with the initial conditions $y(0) = y_0, \ \ldots, y^{(n)}(0) = y_0^{(n)}$.

Models of discrete-time systems

General discrete-time systems

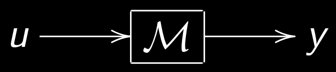
- inputs and outputs defined at discrete time instances $k = 1, 2, \ldots$
- described by ordinary difference equations in the form of

$$y(k)+a_{n-1}y(k-1)+\cdots+a_0 y(k-n) = b_m u(k+m-n)+\cdots+b_0 u(k-n)$$

Example: bank statements

- $x(k+1) = (1+\rho)x(k) + u(k), x(0) = x_0$
- k – month counter; ρ – interest rate; $x(k)$ – wealth at the beginning of month k; $u(k)$ – money saved at the end of month k; x_0 – initial wealth in account

Model properties: static v.s. dynamic, causal v.s. acausal


u $\mathcal{M}$ y

Model $\mathcal{M}$ is said to be

- *memoryless* or *static* if $y(t)$ depends only on $u(t)$
- *dynamic* (has memory) if y at time t depends on input values at other times
- e.g.: $y(t) = \mathcal{M}(u(t)) = \gamma u(t)$, $y(t) = \int_0^t u(\tau)d\tau$, $y(k) = \sum_{i=0}^k u(i)$
- *causal* if $y(t)$ depends on $u(\tau)$ for $\tau \leq t$
- *strictly causal* if $y(t)$ depends on $u(\tau)$ for $\tau < t$, e.g.: $y(t) = u(t-10)$

Linearity and time-invariance

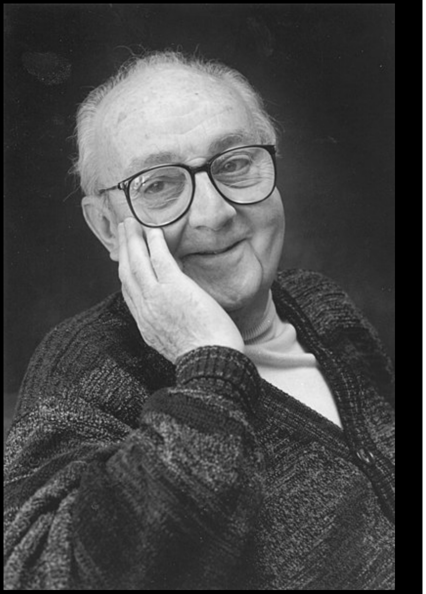
The system $\mathcal{M}$ is called

- *linear* if satisfying the *superposition* property:

$$\mathcal{M}(\alpha_1 u_1(t) + \alpha_2 u_2(t)) = \alpha_1 \mathcal{M}(u_1(t)) + \alpha_2 \mathcal{M}(u_2(t))$$

  for any input signals $u_1(t)$ and $u_2(t)$, and any real numbers α_1 and α_2
- *time-invariant* if its properties do not change with respect to time
- e.g., $\dot{y}(t) = Ay(t) + Bu(t)$ is linear and time-invariant
- $\dot{y}(t) = 2y(t) - \sin(y(t))u(t)$ is nonlinear, yet time-invariant
- $\dot{y}(t) = 2y(t) - t\sin(y(t))u(t)$ is time-varying
- assuming the same initial conditions, if we shift $u(t)$ by a constant time interval, i.e., consider $\mathcal{M}(u(t+\tau_0))$, then $\mathcal{M}$ is time-invariant if the output $\mathcal{M}(u(t+\tau_0)) = y(t+\tau_0)$

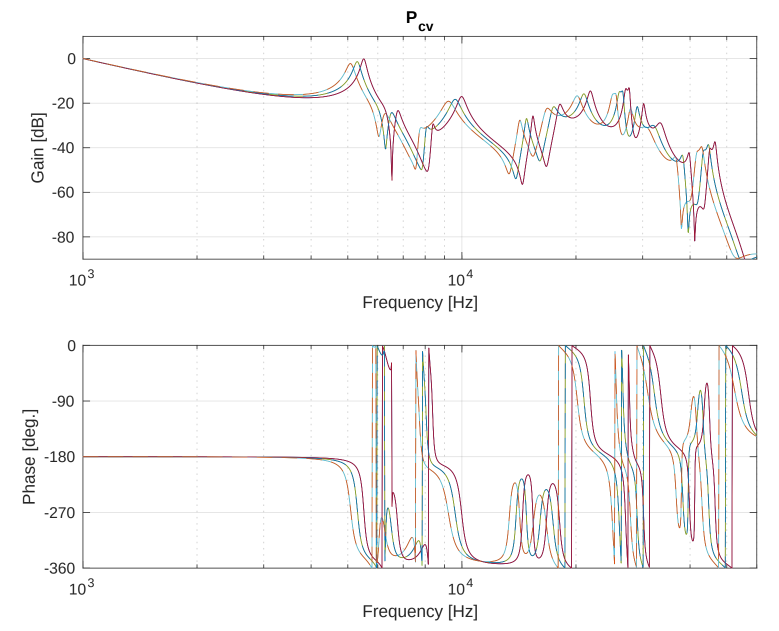
George Box



"All Models are Wrong, but Some are Useful"

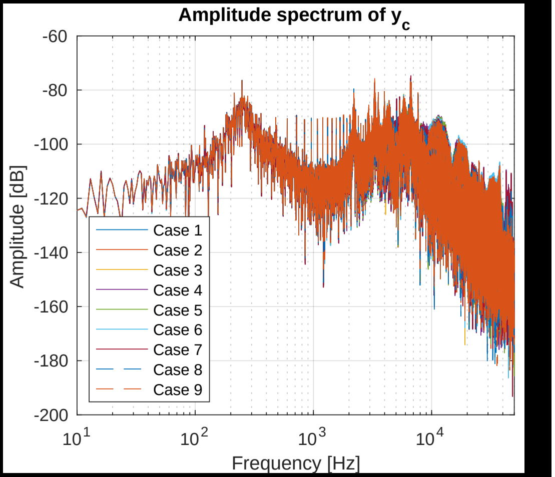
- statistical models always fall short of the complexities of reality but can still be useful nonetheless
- a dynamic system may simply be too complex (consider the neural system of human brains)
- or there are inevitable hardware uncertainties such as the fatigue of gears or bearings in a car

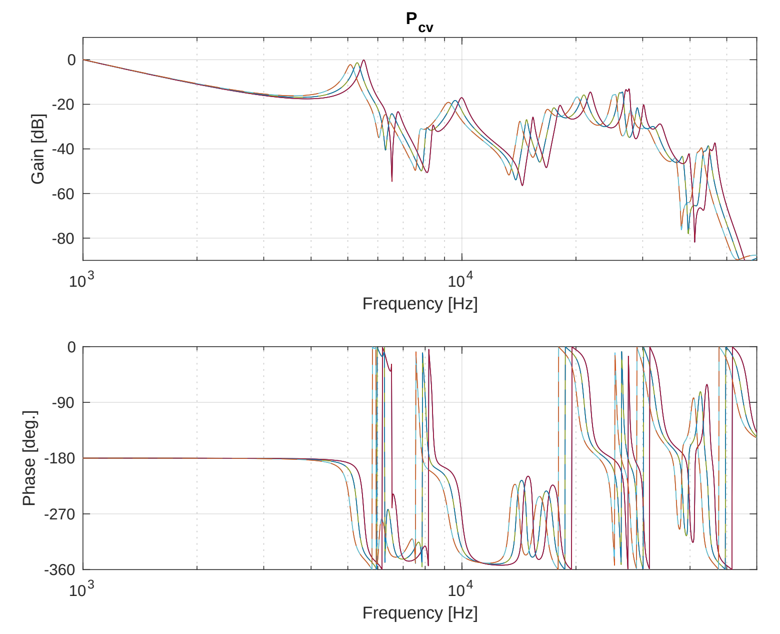
Example (HDDs under perturbation)



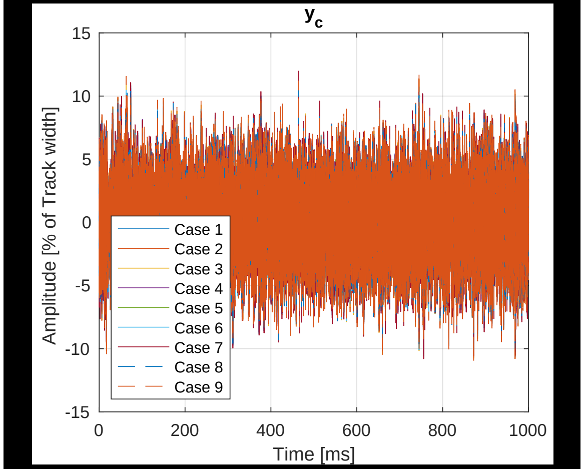
- temperature influence
- manufacturing variations
- but, control works!

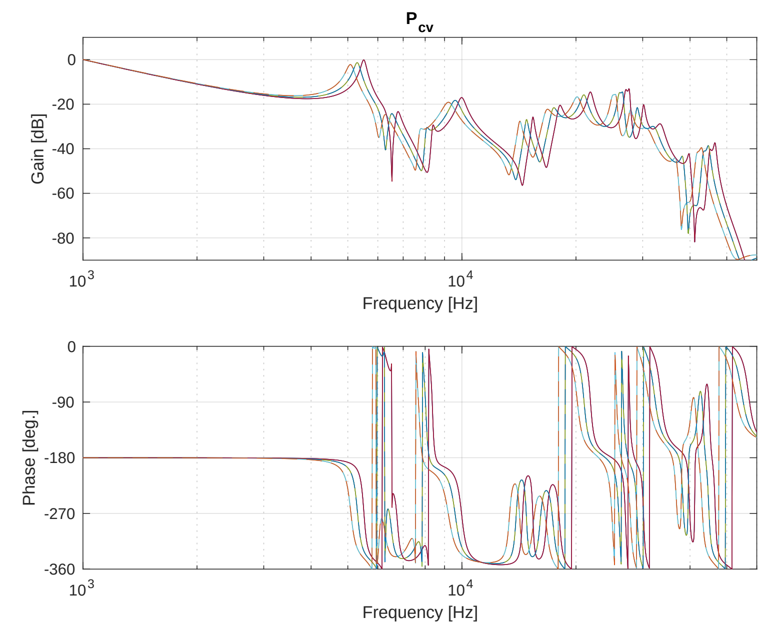
Example (HDDs under perturbation)





Example (HDDs under perturbation)





Example (HDDs under perturbation)

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 8790–8796

Benchmark Problem for Magnetic-Head Positioning Control System in HDDs

Takenori Atsumi [*]

[*] *Chiba Institute of Technology, Narashino, Chiba, 275-0016 Japan (e-mail: takenori.atsumi@p.chibakoudai.jp)*

Abstract: In the "Cloud Era", the data capacity of the hard disk drive (HDD) must grow to develop the cloud. As a result, we must improve the positioning accuracy of the magnetic head in the HDD. To encourage research about magnetic-head positioning control, we release a benchmark problem that works on MATLAB. This benchmark problem enables us to simulate the magnetic-head positioning control system for the latest HDDs with our designed controller. In this paper, a control design method with the decoupling filter is also presented for this benchmark problem.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Precision control, Data storage, Positioning systems, Actuators, Servo.

1. INTRODUCTION

According to a major data-storage device manufacturer, Western Digital, the future of the cloud service is dependent on the hard disk drive (HDD) capacity growth because demands for the data capacity in the cloud service are rapidly increasing. To solve this issue, we are going to improve the accuracy of a magnetic-head positioning control system so that size of bits for data stored on a disk decreases (Yamaguchi and Atsumi (2008); Abrahamson and Fu-ying (2015); Atsumi (2016); Nordrum (2019)).

In order to encourage research about the magnetic-head positioning control, a technical committee consisting of representatives of major universities and an HDD manufacturer with HDD servo research in Japan has developed an open-source HDD benchmark problem and released it on the MathWorks File Exchange (Atsumi et al. (2022)). This benchmark problem enables us to simulate the magnetic-head positioning control system for the latest HDDs used in the cloud storage with our designed controller.

This paper presents the details of the benchmark problems and a control design method with the decoupling filter.

2. HARD DISK DRIVE

Figure 1 shows a picture of the HDD with the cover opened. The HDD consists of a voice coil motor (VCM), PZT actuators, a head-stack assembly (HSA), magnetic heads, disks, and a spindle motor. Most of the latest HDDs for cloud storage employ helium-sealed technology (Aoyagi et al. (2022)). This means that flow-induced vibrations caused by high-speed spinning disks in the HDDs don't matter anymore. As a result, the major causes of tracking errors for the magnetic-head positioning system are external vibrations in storage boxes used in data centers. To compensate for the external vibrations, the latest HDDs employ the dual-stage actuator system that consists of the VCM and the PZT actuators. Figure 2 illustrates the magnetic-head positioning control system with the dual-stage actuator system.

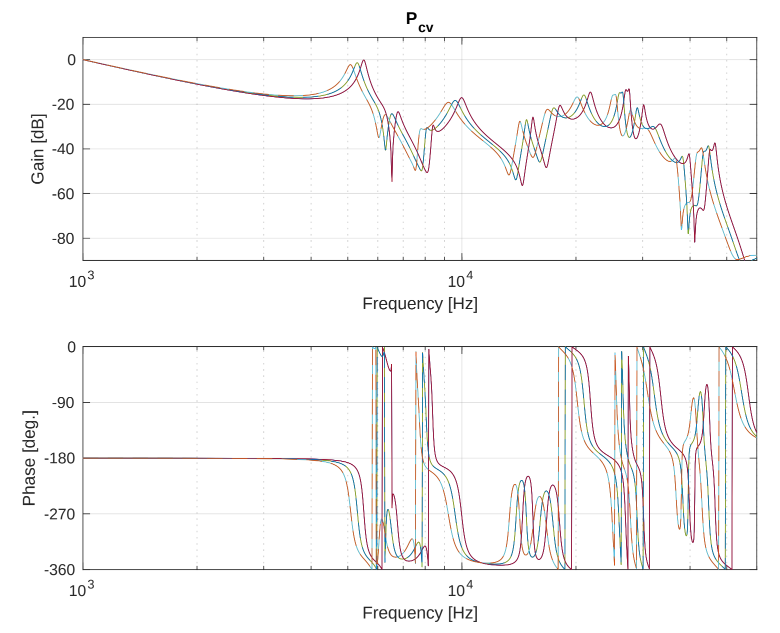
In the magnetic-head positioning system, the controlled variable is the magnetic-head position signal, which is generated from embedded information in servo sectors located at regular intervals on the disks, as shown in Fig. 3. Thus, the magnetic-head position signal is only available as a discrete-time signal. This means that the

Fig. 1. Hard disk drive.

Fig. 2. Magnetic-head positioning system.

Fig. 3. Sectored servo system.

2405-8963 Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2023.10.065

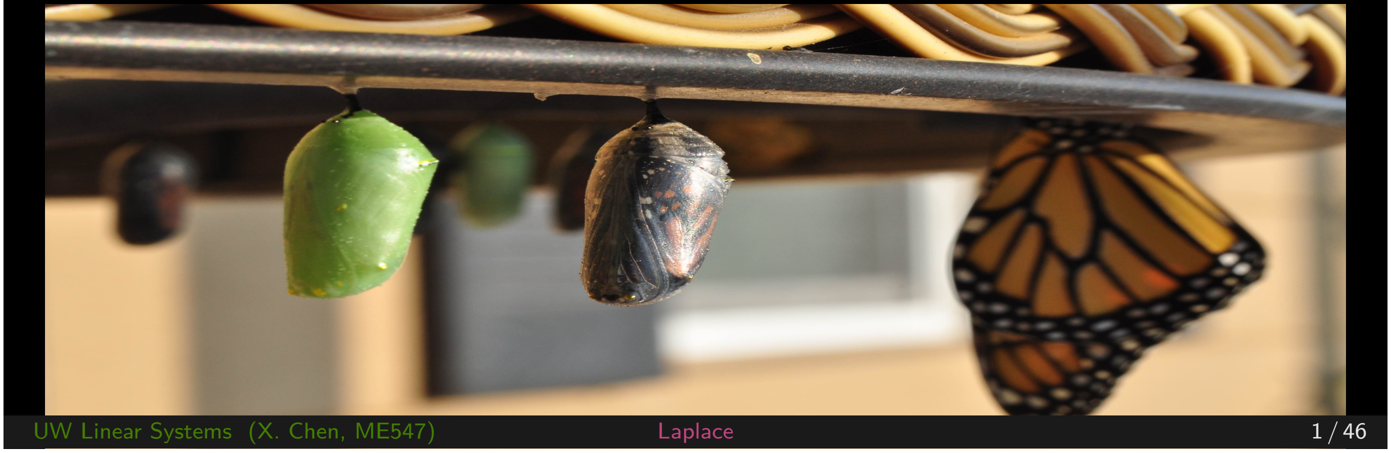


ME547: Linear Systems

Modeling: Review of Laplace Transform

Xu Chen

University of Washington



From infinite series to Laplace

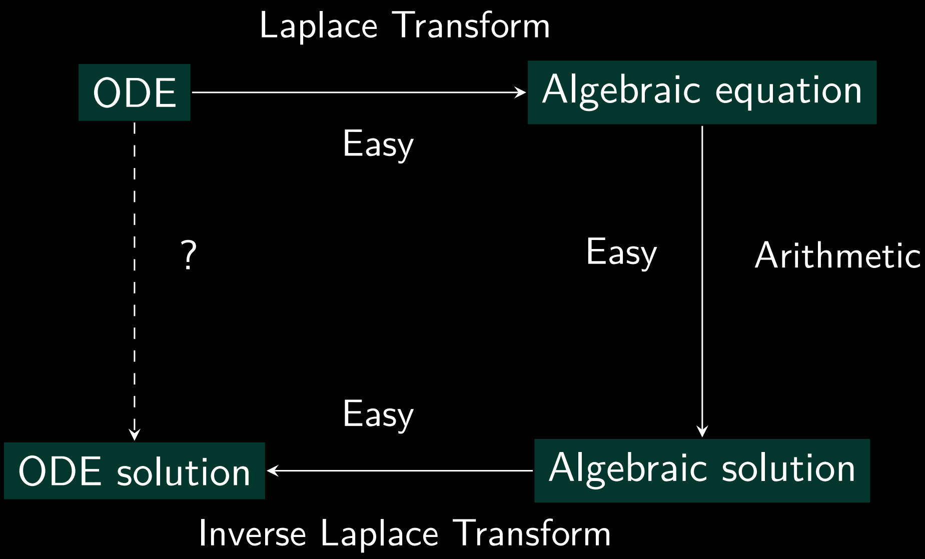
- $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = ?$
- how does it relate to the Laplace transform?

Introduction

Pierre-Simon Laplace (1749-1827)

- "the French Newton" or "Newton of France"
- 13 years younger than Lagrange
- studied under Jean le Rond d'Alembert (co-discovered fundamental theorem of algebra, aka d'Alembert/Gauss theorem)

The Laplace approach to ODEs



Laplace Transform

ODE

Algebraic equation

Easy

?

Easy

Arithmetic

Easy

ODE solution

Algebraic solution

Inverse Laplace Transform

Sets of numbers and the relevant domains

- *set*: a well-defined collection of distinct objects, e.g., $\{1, 2, 3\}$
- $\mathbb{R}$: the set of real numbers
- $\mathbb{C}$: the set of complex numbers
- $\in$: belong to, e.g., $1 \in \mathbb{R}$
- $\mathbb{R}_+$: the set of positive real numbers
- $\triangleq$: defined as, e.g., $y(t) \triangleq 3x(t) + 1$

Continuous-time functions

Formal notation:

$$f : \mathbb{R}_+ \to \mathbb{R}$$

where the domain of f is in $\mathbb{R}_+$, and the value of f is in $\mathbb{R}$

- we use $f(t)$ to denote a continuous-time function
- assume that $f(t) = 0$ for all $t < 0$

Laplace transform definition

For a continuous-time function

$$f:\ \mathbb{R}_+ \to \mathbb{R}$$

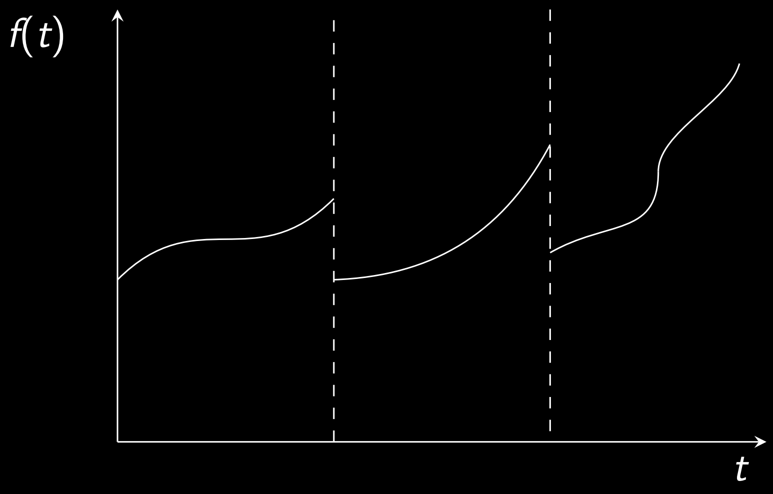
define Laplace Transform:

$$\boxed{F(s) = \mathcal{L}\{f(t)\} \triangleq \int_0^\infty f(t)e^{-st}dt}$$

$s \in \mathbb{C}$

Existence: Sufficient condition 1

- $f(t)$ is piecewise continuous

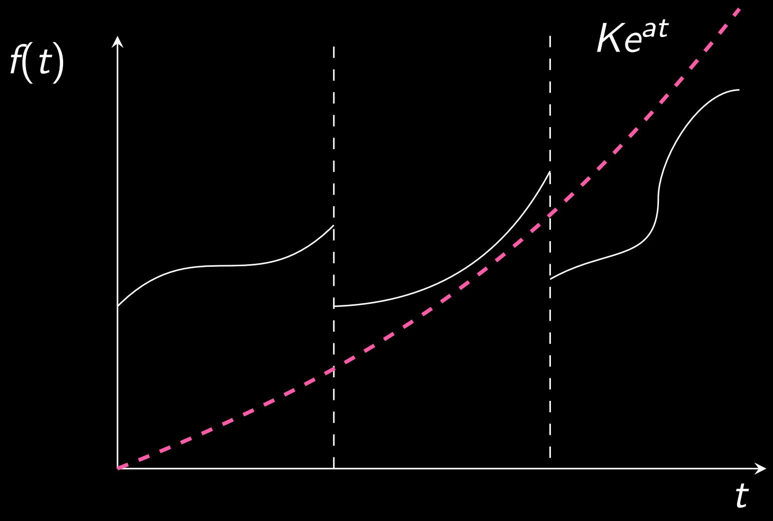


Existence: Sufficient condition 2

- $f(t)$ does not grow faster than an exponential as $t \to \infty$:

$$|f(t)| < ke^{\alpha t}, \text{ for all } t \geq t_0$$

for some constants: $k, \ \alpha, \ t_0 \in \mathbb{R}_+$.



Examples: Exponential

- $f(t) = e^{-at}, \ a \in \mathbb{C}$
- $F(s) = \frac{1}{s+a}$

Examples: Exponential

- $f(t) = 1(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$
- $F(s) = \frac{1}{s}$

Laplace transform and infinite series

- $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = ?$

Examples: Sine

- $f(t) = \sin(\omega t)$
- $F(s) = \frac{\omega}{s^2+\omega^2}$
- Use: $\sin(\omega t) = \frac{e^{j\omega t}-e^{-j\omega t}}{2j}$, $\mathcal{L}\{e^{j\omega t}\} = \frac{1}{s-j\omega}$

Recall: Euler formula

$$e^{ja} = \cos a + j \sin a$$

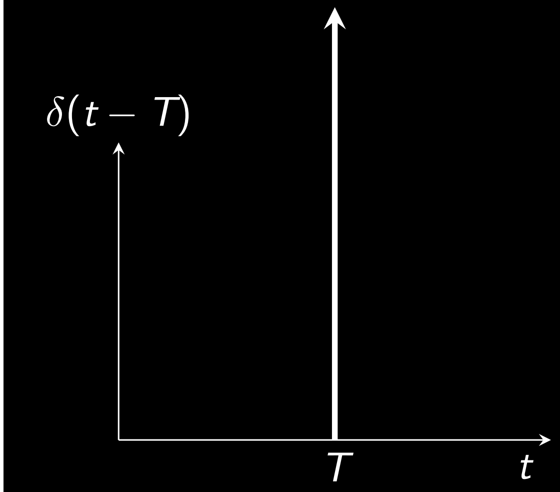
Leonhard Euler (04/15/1707 - 09/18/1783):

- Swiss mathematician, physicist, astronomer, geographer, logician and engineer
- studied under Johann Bernoulli
- teacher of Lagrange
- wrote 380 articles within 25 years at Berlin
- produced on average one paper per week at age 67, when almost blind!

Examples: Cosine

- $f(t) = \cos(\omega t)$
- $F(s) = \frac{s}{s^2+\omega^2}$

Examples: Dirac impulse

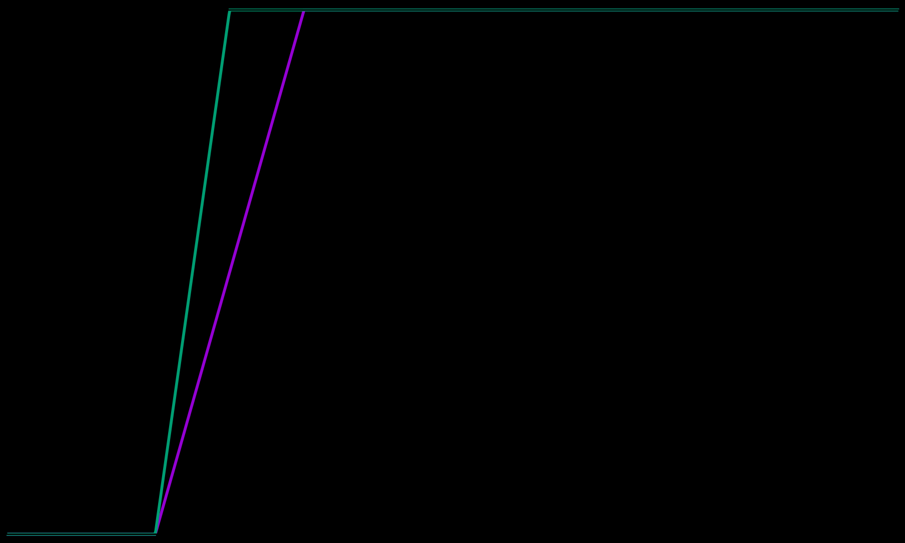


- a generalized function (formally, a distribution)
- e.g., consider $\dot{y} - ay = \dot{u} + bu$
  - if u is a unit step $1(t)$
  - $\dot{u}$ has a jump at 0
  - cannot directly evaluate $\dot{u}$!

Approximating the unit step

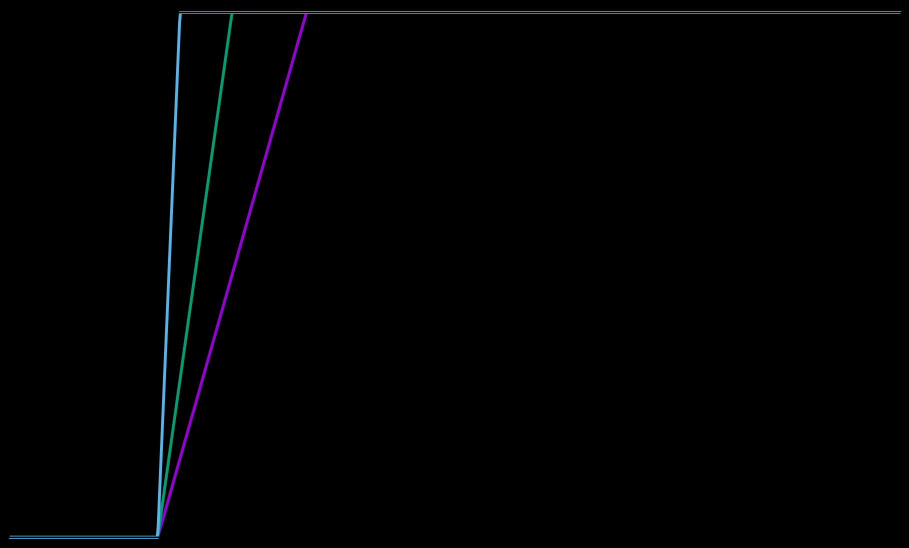
$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

Approximating the unit step



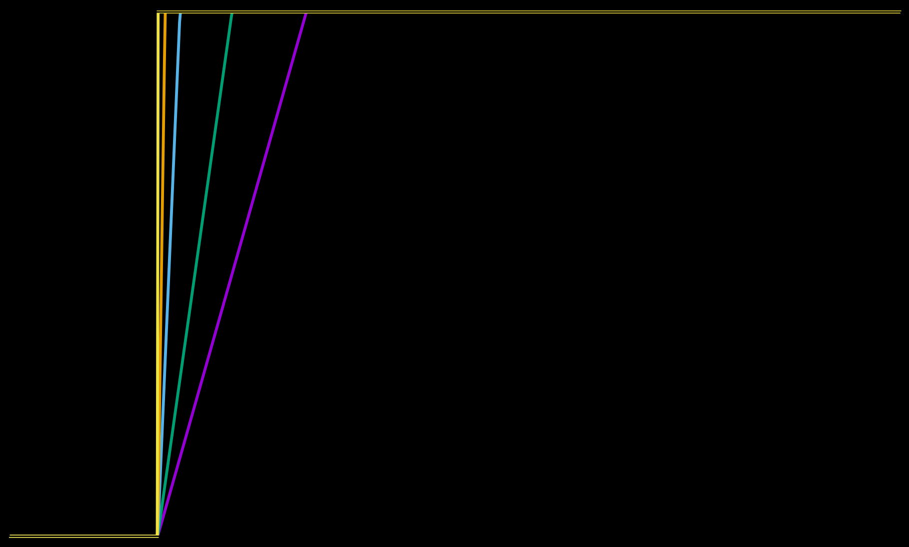
$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

Approximating the unit step



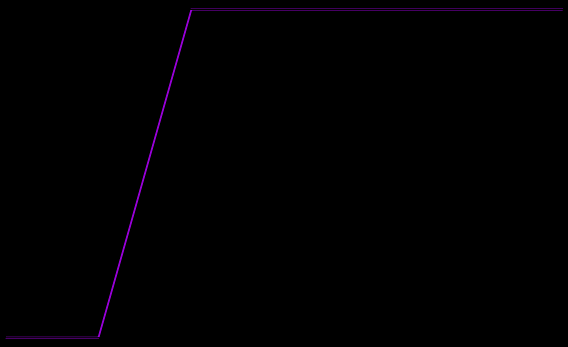
$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

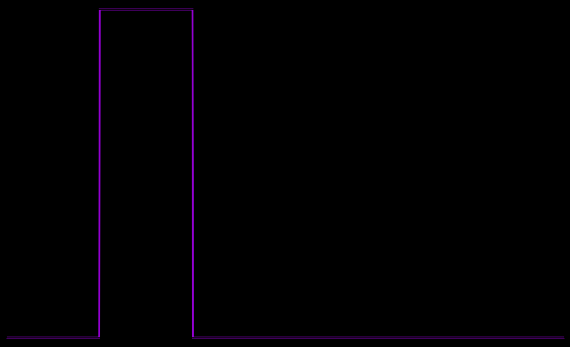
Approximating the unit step



$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

Approximating $\dot{1}(t)$

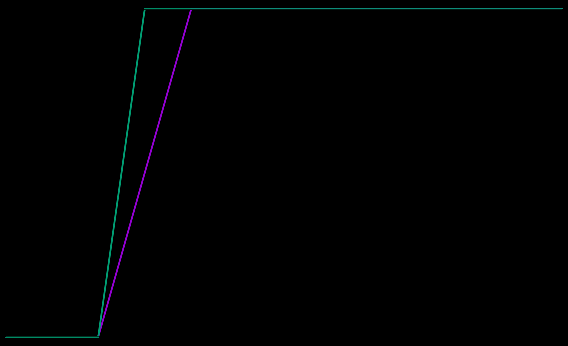


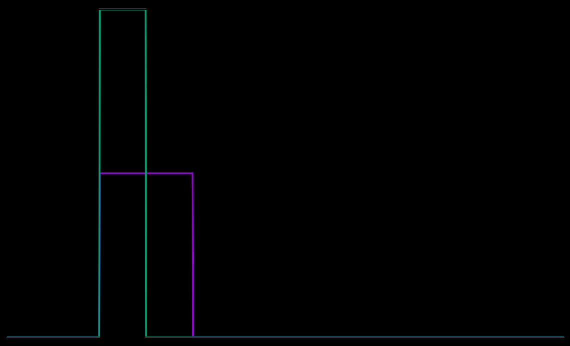


$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon}t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t \end{cases}$$

Approximating $\dot{1}(t)$

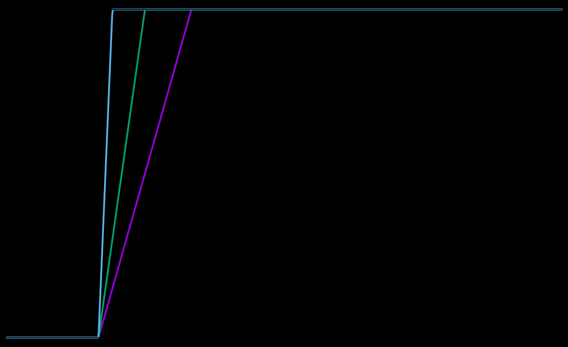


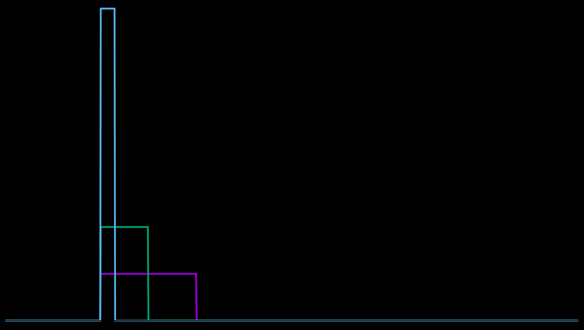


$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon}t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t \end{cases}$$

Approximating $\dot{1}(t)$

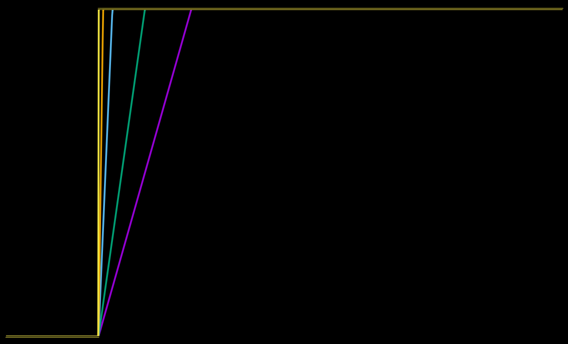


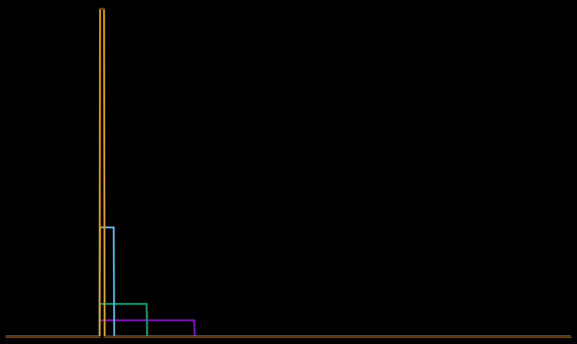


$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon}t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t \end{cases}$$

Approximating $\dot{1}(t)$



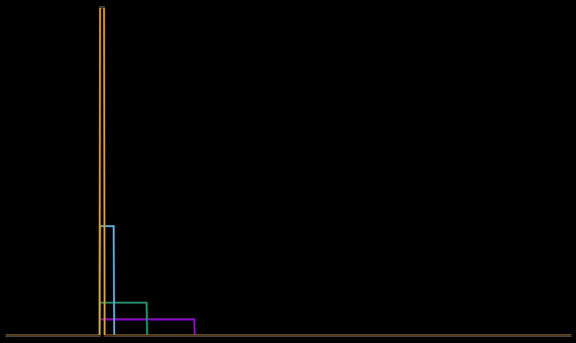


$$\mu_\epsilon(t) := \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon}t & \text{for } 0 \le t < \epsilon \\ 1 & \text{for } \epsilon \le t \end{cases}$$

$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t \end{cases}$$

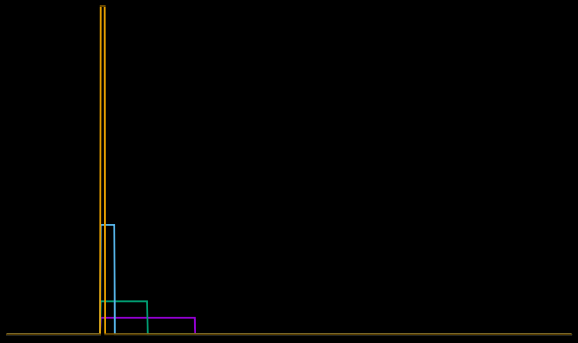
Properties of the first-order approximation

$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t \end{cases}$$

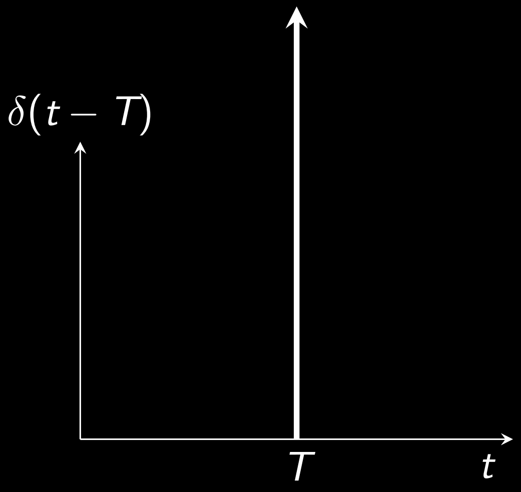


- $\int_{-\infty}^{\infty} \dot{\mu}_\epsilon(t)dt = 1$
- $\lim_{\epsilon\to 0} \int_{-\infty}^{\infty} f(t)\dot{\mu}_\epsilon(t)dt = \lim_{\epsilon\to 0} \int_{0}^{\epsilon} f(t)\frac{1}{\epsilon}dt = f(0)$

General Dirac impulse properties

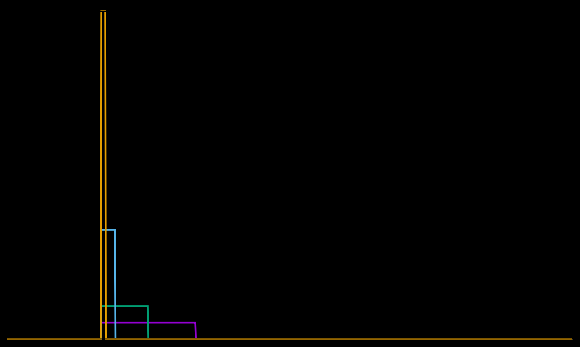


- $\int_{-\infty}^{\infty} \dot{\mu}_\epsilon(t)dt = 1$
- $\lim_{\epsilon\to 0} \int_{-\infty}^{\infty} f(t)\dot{\mu}_\epsilon(t)dt = \lim_{\epsilon\to 0} \int_{0}^{\epsilon} f(t)\frac{1}{\epsilon}dt = f(0)$



- $\int_{0}^{\infty} \delta(t-T)dt = 1$
- $\int_{0}^{\infty} \delta(t-T)f(t)dt = f(T)$

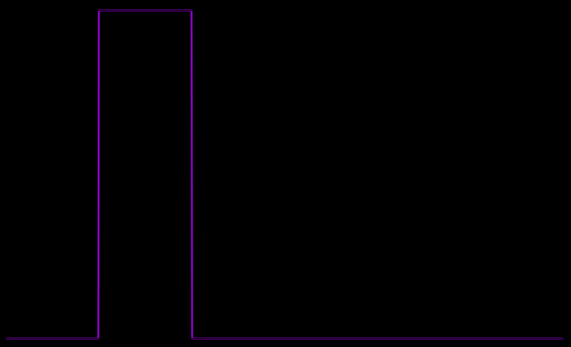
Challenges with the first-order approximation

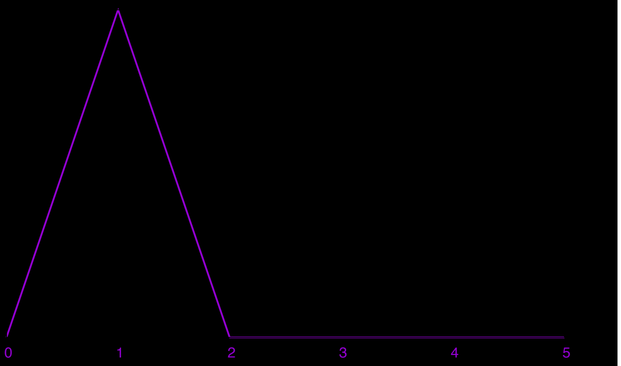


- $\int_{-\infty}^{\infty} \dot{\mu}_\epsilon(t)dt = 1$
- $\lim_{\epsilon\to 0} \int_{-\infty}^{\infty} f(t)\dot{\mu}_\epsilon(t)dt = \lim_{\epsilon\to 0} \int_{0}^{\epsilon} f(t)\frac{1}{\epsilon}dt = f(0)$

- $\dot{\mu}_\epsilon(t)$ is piecewise-continuous and not fully differentiable
- $\mu_\epsilon(t) \approx 1(t)$ is only first-order differentiable
- cannot handle, e.g., $\ddot{y} + 2\dot{y} - ay = \ddot{u} + 3\dot{u} + bu$

Second-order approximation of $\dot{1}(t)$

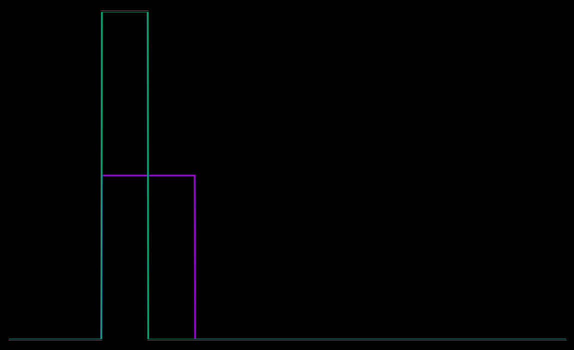


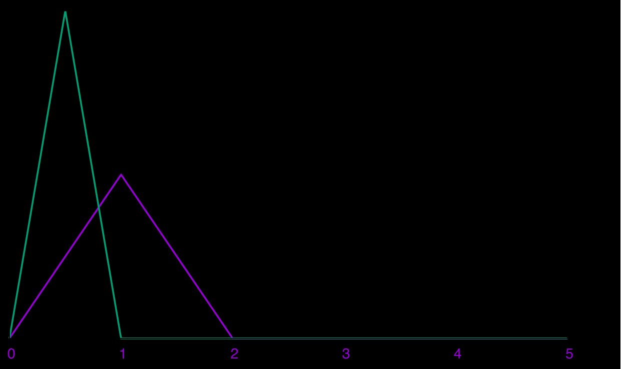


$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t < 2\epsilon \\ 0 & \text{for } \epsilon < t \end{cases}$$

$$\delta_\epsilon(t) := \begin{cases} 0 & \text{for} \quad t < 0 \\ \frac{t}{\epsilon^2} & \text{for} \quad 0 < t < \epsilon \\ \frac{2\epsilon - t}{\epsilon^2} & \text{for} \quad \epsilon < t < 2\epsilon \\ 0 & \text{for} \quad 2\epsilon < t \end{cases}$$

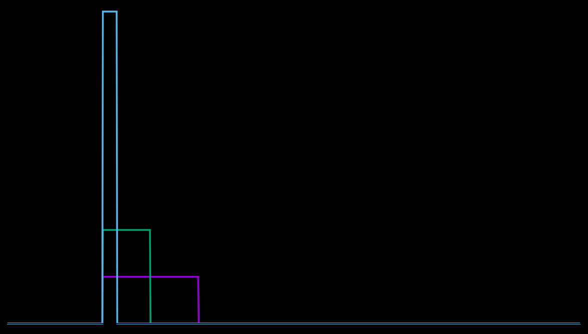
Second-order approximation of $\dot{1}(t)$

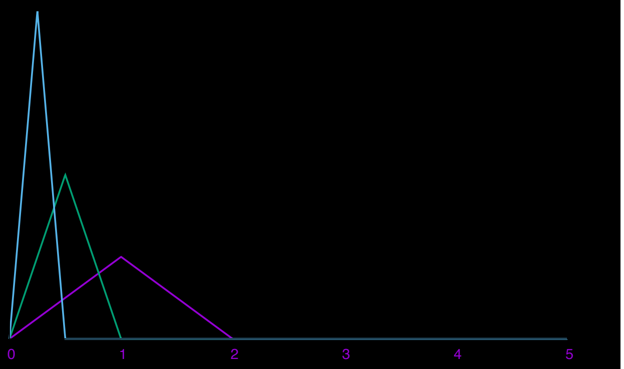




$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t < 2\epsilon \\ 0 & \text{for } \epsilon < t \end{cases} \qquad \delta_\epsilon(t) := \begin{cases} 0 & \text{for} \quad t < 0 \\ \frac{t}{\epsilon^2} & \text{for} \quad 0 < t < \epsilon \\ \frac{2\epsilon - t}{\epsilon^2} & \text{for} \quad \epsilon < t < 2\epsilon \\ 0 & \text{for} \quad 2\epsilon < t \end{cases}$$

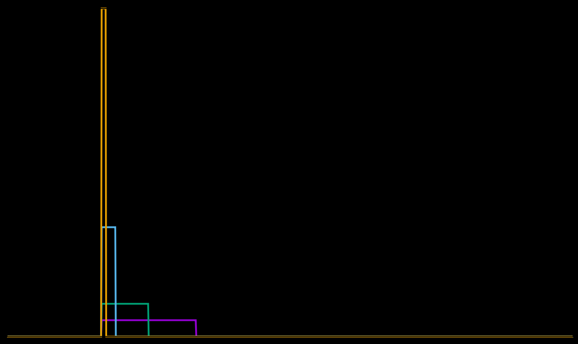
Second-order approximation of $\dot{1}(t)$

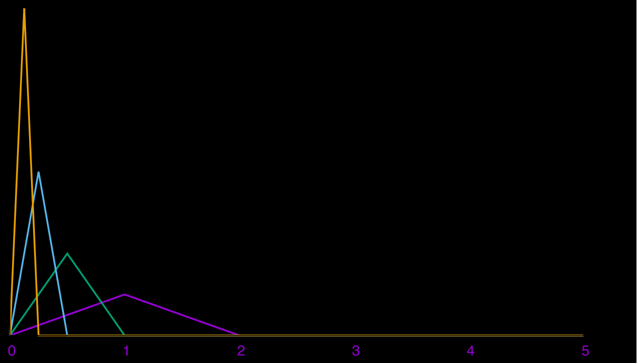




$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t < 2\epsilon \\ 0 & \text{for } \epsilon < t \end{cases} \qquad \delta_\epsilon(t) := \begin{cases} 0 & \text{for} \quad t < 0 \\ \frac{t}{\epsilon^2} & \text{for} \quad 0 < t < \epsilon \\ \frac{2\epsilon - t}{\epsilon^2} & \text{for} \quad \epsilon < t < 2\epsilon \\ 0 & \text{for} \quad 2\epsilon < t \end{cases}$$

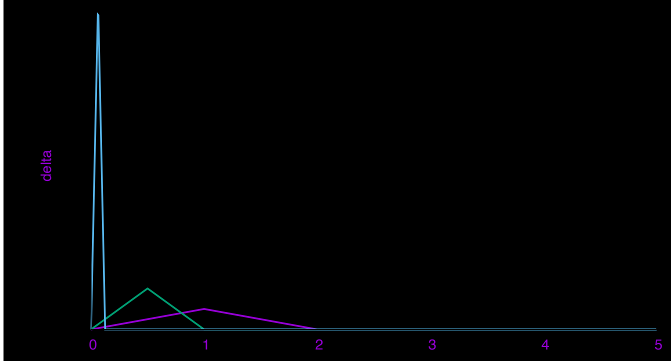
Second-order approximation of $\dot{1}(t)$

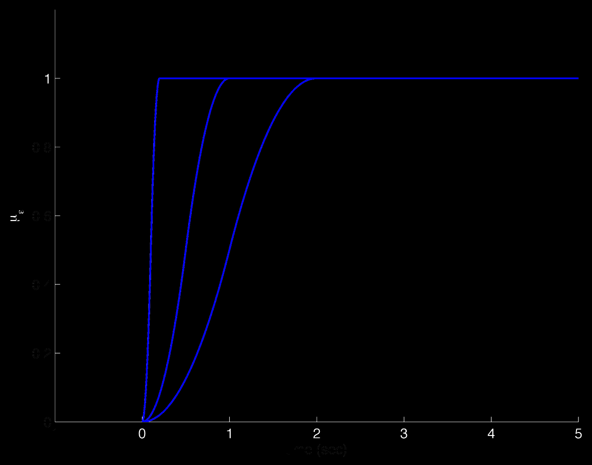




$$\dot{\mu}_\epsilon(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{\epsilon} & \text{for } 0 < t < \epsilon \\ 0 & \text{for } \epsilon < t < 2\epsilon \\ 0 & \text{for } \epsilon < t \end{cases} \qquad \delta_\epsilon(t) := \begin{cases} 0 & \text{for} \quad t < 0 \\ \frac{t}{\epsilon^2} & \text{for} \quad 0 < t < \epsilon \\ \frac{2\epsilon - t}{\epsilon^2} & \text{for} \quad \epsilon < t < 2\epsilon \\ 0 & \text{for} \quad 2\epsilon < t \end{cases}$$

Second-order approximation of $1(t)$

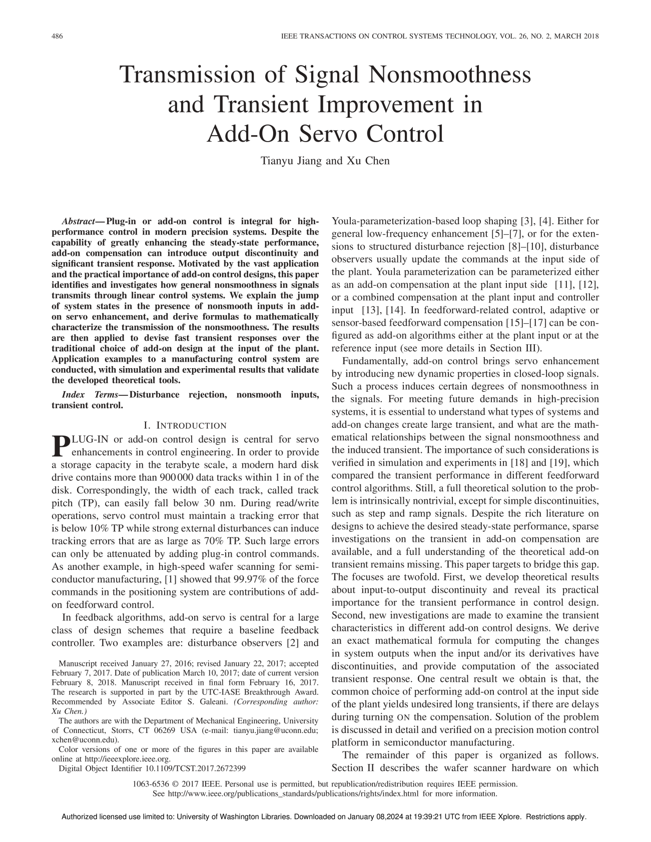




$$\delta_\epsilon(t) := \begin{cases} 0 & \text{for} \quad t < 0 \\ \frac{t}{\epsilon^2} & \text{for} \quad 0 < t < \epsilon \\ \frac{2\epsilon - t}{\epsilon^2} & \text{for} \quad \epsilon < t < 2\epsilon \\ 0 & \text{for} \quad 2\epsilon < t \end{cases}$$

- $\mu_\epsilon(t) = \int_0^t \delta_\epsilon(\tau) d\tau$: a smoother approximation of the unit step!
- is twice differentiable
- can keep on doing this to make δ_ϵ infinitely differentiable

Application of the concept



Transmission of Signal Nonsmoothness and Transient Improvement in Add-On Servo Control

Tianyu Jiang and Xu Chen

***Abstract*—Plug-in or add-on control is integral for high-performance control in modern precision systems. Despite the capability of greatly enhancing the steady-state performance, add-on compensation can introduce output discontinuity and significant transient response. Motivated by the vast application and the practical importance of add-on control designs, this paper identifies and investigates how general nonsmoothness in signals transmits through linear control systems. We explain the jump of system states in the presence of nonsmooth inputs in add-on servo enhancement, and derive formulas to mathematically characterize the transmission of the nonsmoothness. The results are then applied to devise fast transient responses over the traditional choice of add-on design at the input of the plant. Application examples to a manufacturing control system are conducted, with simulation and experimental results that validate the developed theoretical tools.**

***Index Terms*—Disturbance rejection, nonsmooth inputs, transient control.**

I. Introduction

PLUG-IN or add-on control design is central for servo enhancements in control engineering. In order to provide a storage capacity in the terabyte scale, a modern hard disk drive contains more than 900 000 data tracks within 1 in of the disk. Correspondingly, the width of each track, called track pitch (TP), can easily fall below 30 nm. During read/write operations, servo control must maintain a tracking error that is below 10% TP while strong external disturbances can induce tracking errors that are as large as 70% TP. Such large errors can only be attenuated by adding plug-in control commands. As another example, in high-speed wafer scanning for semiconductor manufacturing, [1] showed that 99.97% of the force commands in the positioning system are contributions of add-on feedforward control.

In feedback algorithms, add-on servo is central for a large class of design schemes that require a baseline feedback controller. Two examples are: disturbance observers [2] and Youla-parameterization-based loop shaping [3], [4]. Either for general low-frequency enhancement [5]–[7], or for the extensions to structured disturbance rejection [8]–[10], disturbance observers usually update the commands at the input side of the plant. Youla parameterization can be parameterized either as an add-on compensation at the plant input side [11], [12], or a combined compensation at the plant input and controller input [13], [14]. In feedforward-related control, adaptive or sensor-based feedforward compensation [15]–[17] can be configured as add-on algorithms either at the plant input or at the reference input (see more details in Section III).

Fundamentally, add-on control brings servo enhancement by introducing new dynamic properties in closed-loop signals. Such a process induces certain degrees of nonsmoothness in the signals. For meeting future demands in high-precision systems, it is essential to understand what types of systems and add-on changes create large transient, and what are the mathematical relationships between the signal nonsmoothness and the induced transient. The importance of such considerations is verified in simulation and experiments in [18] and [19], which compared the transient performance in different feedforward control algorithms. Still, a full theoretical solution to the problem is intrinsically nontrivial, except for simple discontinuities, such as step and ramp signals. Despite the rich literature on designs to achieve the desired steady-state performance, sparse investigations on the transient in add-on compensation are available, and a full understanding of the theoretical add-on transient remains missing. This paper targets to bridge this gap. The focuses are twofold. First, we develop theoretical results about input-to-output discontinuity and reveal its practical importance for the transient performance in control design. Second, new investigations are made to examine the transient characteristics in different add-on control designs. We derive an exact mathematical formula for computing the changes in system outputs when the input and/or its derivatives have discontinuities, and provide computation of the associated transient response. One central result we obtain is that, the common choice of performing add-on control at the input side of the plant yields undesired long transients, if there are delays during turning ON the compensation. Solution of the problem is discussed in detail and verified on a precision motion control platform in semiconductor manufacturing.

The remainder of this paper is organized as follows. Section II describes the wafer scanner hardware on which

Manuscript received January 27, 2016; revised January 22, 2017; accepted February 7, 2017. Date of publication March 10, 2017; date of current version February 8, 2018. Manuscript received in final form February 16, 2017. The research is supported in part by the UTC-IASE Breakthrough Award. Recommended by Associate Editor S. Galeani. *(Corresponding author: Xu Chen.)*

The authors are with the Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: tianyu.jiang@uconn.edu; xchen@uconn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCST.2017.2672399

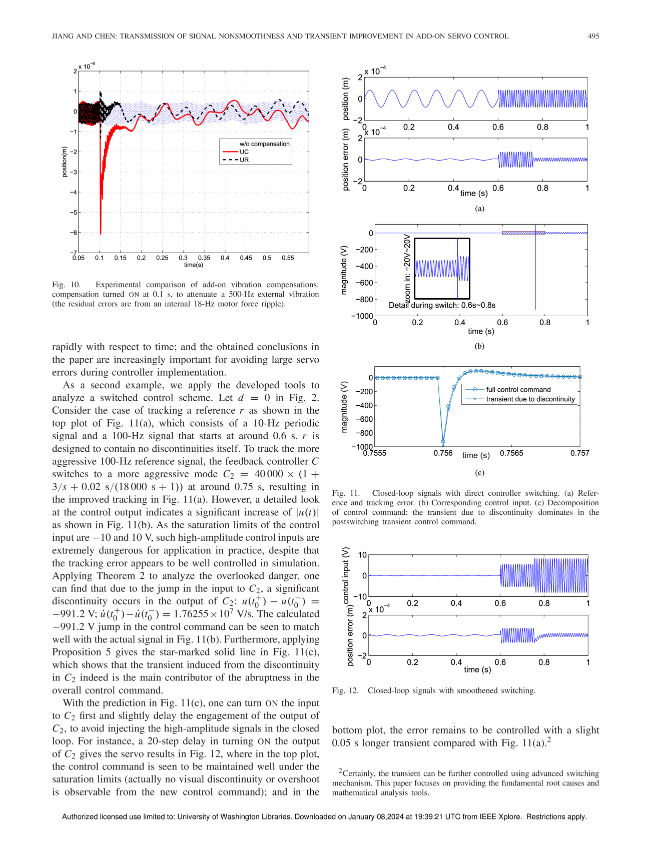


Fig. 10. Experimental comparison of add-on vibration compensations: compensation turned ON at 0.1 s, to attenuate a 500-Hz external vibration (the residual errors are from an internal 18-Hz motor force ripple).

Fig. 11. Closed-loop signals with direct controller switching. (a) Reference and tracking error. (b) Corresponding control input. (c) Decomposition of control command: the transient due to discontinuity dominates in the postswitching transient control command.

Fig. 12. Closed-loop signals with smoothened switching.

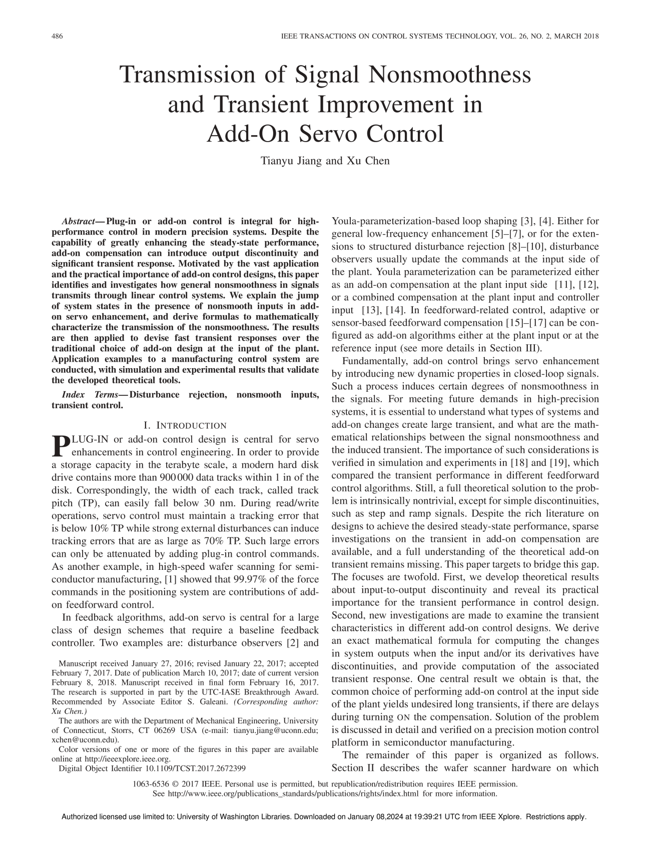
rapidly with respect to time; and the obtained conclusions in the paper are increasingly important for avoiding large servo errors during controller implementation.

As a second example, we apply the developed tools to analyze a switched control scheme. Let $d = 0$ in Fig. 2. Consider the case of tracking a reference r as shown in the top plot of Fig. 11(a), which consists of a 10-Hz periodic signal and a 100-Hz signal that starts at around 0.6 s. r is designed to contain no discontinuities itself. To track the more aggressive 100-Hz reference signal, the feedback controller C switches to a more aggressive mode $C_2 = 40\,000 \times (1 + 3/s + 0.02\ s/(18\,000\ s + 1))$ at around 0.75 s, resulting in the improved tracking in Fig. 11(a). However, a detailed look at the control output indicates a significant increase of $|u(t)|$ as shown in Fig. 11(b). As the saturation limits of the control input are −10 and 10 V, such high-amplitude control inputs are extremely dangerous for application in practice, despite that the tracking error appears to be well controlled in simulation. Applying Theorem 2 to analyze the overlooked danger, one can find that due to the jump in the input to C_2, a significant discontinuity occurs in the output of C_2: $u(t_0^+) - u(t_0^-) = -991.2$ V; $\dot{u}(t_0^+) - \dot{u}(t_0^-) = 1.76255 \times 10^7$ V/s. The calculated −991.2 V jump in the control command can be seen to match well with the actual signal in Fig. 11(b). Furthermore, applying Proposition 5 gives the star-marked solid line in Fig. 11(c), which shows that the transient induced from the discontinuity in C_2 indeed is the main contributor of the abruptness in the overall control command.

With the prediction in Fig. 11(c), one can turn ON the input to C_2 first and slightly delay the engagement of the output of C_2, to avoid injecting the high-amplitude signals in the closed loop. For instance, a 20-step delay in turning ON the output of C_2 gives the servo results in Fig. 12, where in the top plot, the control command is seen to be maintained well under the saturation limits (actually no visual discontinuity or overshoot is observable from the new control command); and in the bottom plot, the error remains to be controlled with a slight 0.05 s longer transient compared with Fig. 11(a).[2]

[2] Certainly, the transient can be further controlled using advanced switching mechanism. This paper focuses on providing the fundamental root causes and mathematical analysis tools.

Application of the concept



486 IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, VOL. 26, NO. 2, MARCH 2018

Transmission of Signal Nonsmoothness and Transient Improvement in Add-On Servo Control

Tianyu Jiang and Xu Chen

***Abstract*—Plug-in or add-on control is integral for high-performance control in modern precision systems. Despite the capability of greatly enhancing the steady-state performance, add-on compensation can introduce output discontinuity and significant transient response. Motivated by the vast application and the practical importance of add-on control designs, this paper identifies and investigates how general nonsmoothness in signals transmits through linear control systems. We explain the jump of system states in the presence of nonsmooth inputs in add-on servo enhancement, and derive formulas to mathematically characterize the transmission of the nonsmoothness. The results are then applied to devise fast transient responses over the traditional choice of add-on design at the input of the plant. Application examples to a manufacturing control system are conducted, with simulation and experimental results that validate the developed theoretical tools.**

***Index Terms*—Disturbance rejection, nonsmooth inputs, transient control.**

I. Introduction

PLUG-IN or add-on control design is central for servo enhancements in control engineering. In order to provide a storage capacity in the terabyte scale, a modern hard disk drive contains more than 900 000 data tracks within 1 in of the disk. Correspondingly, the width of each track, called track pitch (TP), can easily fall below 30 nm. During read/write operations, servo control must maintain a tracking error that is below 10% TP while strong external disturbances can induce tracking errors that are as large as 70% TP. Such large errors can only be attenuated by adding plug-in control commands. As another example, in high-speed wafer scanning for semiconductor manufacturing, [1] showed that 99.97% of the force commands in the positioning system are contributions of add-on feedforward control.

In feedback algorithms, add-on servo is central for a large class of design schemes that require a baseline feedback controller. Two examples are: disturbance observers [2] and Youla-parameterization-based loop shaping [3], [4]. Either for general low-frequency enhancement [5]–[7], or for the extensions to structured disturbance rejection [8]–[10], disturbance observers usually update the commands at the input side of the plant. Youla parameterization can be parameterized either as an add-on compensation at the plant input side [11], [12], or a combined compensation at the plant input and controller input [13], [14]. In feedforward-related control, adaptive or sensor-based feedforward compensation [15]–[17] can be configured as add-on algorithms either at the plant input or at the reference input (see more details in Section III).

Fundamentally, add-on control brings servo enhancement by introducing new dynamic properties in closed-loop signals. Such a process induces certain degrees of nonsmoothness in the signals. For meeting future demands in high-precision systems, it is essential to understand what types of systems and add-on changes create large transient, and what are the mathematical relationships between the signal nonsmoothness and the induced transient. The importance of such considerations is verified in simulation and experiments in [18] and [19], which compared the transient performance in different feedforward control algorithms. Still, a full theoretical solution to the problem is intrinsically nontrivial, except for simple discontinuities, such as step and ramp signals. Despite the rich literature on designs to achieve the desired steady-state performance, sparse investigations on the transient in add-on compensation are available, and a full understanding of the theoretical add-on transient remains missing. This paper targets to bridge this gap. The focuses are twofold. First, we develop theoretical results about input-to-output discontinuity and reveal its practical importance for the transient performance in control design. Second, new investigations are made to examine the transient characteristics in different add-on control designs. We derive an exact mathematical formula for computing the changes in system outputs when the input and/or its derivatives have discontinuities, and provide computation of the associated transient response. One central result we obtain is that, the common choice of performing add-on control at the input side of the plant yields undesired long transients, if there are delays during turning ON the compensation. Solution of the problem is discussed in detail and verified on a precision motion control platform in semiconductor manufacturing.

The remainder of this paper is organized as follows. Section II describes the wafer scanner hardware on which

Manuscript received January 27, 2016; revised January 22, 2017; accepted February 7, 2017. Date of publication March 10, 2017; date of current version February 8, 2018. Manuscript received in final form February 16, 2017. The research is supported in part by the UTC-IASE Breakthrough Award. Recommended by Associate Editor S. Galeani. *(Corresponding author: Xu Chen.)*

The authors are with the Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: tianyu.jiang@uconn.edu; xchen@uconn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCST.2017.2672399

1063-6536 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Washington Libraries. Downloaded on January 08,2024 at 19:39:21 UTC from IEEE Xplore. Restrictions apply.

Laplace transform of the Dirac impulse

- $\mathcal{L}\{\delta(t)\} = \int_0^\infty e^{-st}\delta(t)dt = e^{-s0} = 1$
- because $\int_0^\infty \delta(t)f(t)dt = f(0)$

Properties of Laplace transform

Linearity

For any $\alpha, \beta \in \mathbb{C}$ and functions $f(t), g(t)$, let

$$F(s) = \mathcal{L}\{f(t)\}, \;\; G(s) = \mathcal{L}\{g(t)\}$$

then

$$\boxed{\mathcal{L}\{\alpha f(t) + \beta g(t)\} = \alpha F(s) + \beta G(s)}$$

Differentiation

Defining

$$\dot{f}(t) = \frac{df(t)}{dt}$$
$$F(s) = \mathcal{L}\{f(t)\}$$

- then

$$\boxed{\mathcal{L}\{\dot{f}(t)\} = sF(s) - f(0)}$$

- via integration by parts:

$$\begin{aligned}\mathcal{L}\{\dot{f}(t)\} &= \int_0^\infty e^{-st}\dot{f}(t)dt \\ &= -\int_0^\infty \frac{de^{-st}}{dt}f(t)dt + \left\{e^{-st}f(t)\right\}_{t=0}^{t\to\infty} \\ &= s\int_0^\infty e^{-st}f(t)dt - f(0) = sF(s) - f(0)\end{aligned}$$

Integration

Defining

$$F(s) = \mathcal{L}\{f(t)\}$$

then

$$\boxed{\mathcal{L}\left\{\int_0^t f(\tau)d\tau\right\} = \frac{1}{s}F(s)}$$

Multiplication by e^{-at}

Defining

$$F(s) = \mathcal{L}\{f(t)\}$$

- then

$$\boxed{\mathcal{L}\left\{e^{-at}f(t)\right\} = F(s+a)}$$

- Example:

$$\mathcal{L}\{1(t)\} = \frac{1}{s} \quad \mathcal{L}\{e^{-at}\} = \frac{1}{s+a}$$

$$\mathcal{L}\{\sin(\omega t)\} = \frac{\omega}{s^2+\omega^2} \quad \mathcal{L}\{e^{-at}\sin(\omega t)\} = \frac{\omega}{(s+a)^2+\omega^2}$$

Multiplication by t

Defining

$$F(s) = \mathcal{L}\{f(t)\}$$

- then

$$\boxed{\mathcal{L}\left\{tf(t)\right\} = -\frac{dF(s)}{ds}}$$

- Example:

$$\mathcal{L}\{1(t)\} = \frac{1}{s} \quad \mathcal{L}\{t\} = \frac{1}{s^2}$$

Time delay τ

Defining

$$F(s) = \mathcal{L}\{f(t)\}$$

then

$$\boxed{\mathcal{L}\left\{f(t-\tau)\right\} = e^{-s\tau}F(s)}$$

Convolution

Given $f(t)$, $g(t)$, and

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

- then

$$\boxed{\mathcal{L}\left\{(f \star g)(t)\right\} = F(s)G(s)}$$

- hence we have

$$\delta(t) \longrightarrow \boxed{G(s)} \longrightarrow g(t) = \mathcal{L}^{-1}\left\{G(s)\right\}$$

because

$$1 \longrightarrow \boxed{G(s)} \longrightarrow Y(s) = G(s) \times 1$$

Initial Value Theorem

If $f(0_+) = \lim_{t \to 0_+} f(t)$ exists, then

$$f(0_+) = \lim_{s \to \infty} sF(s)$$

Final Value Theorem

If $\lim_{t\to\infty} f(t)$ exists,

- then
$$\lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s)$$
- Example: find the final value of the system corresponding to:
$$Y_1(s) = \frac{3(s+2)}{s(s^2+2s+10)}, \quad Y_2(s) = \frac{3}{s-2}$$

Common Laplace transform pairs

$f(t)$	$F(s)$	$f(t)$	$F(s)$
$\sin\omega t$	$\frac{\omega}{s^2+\omega^2}$	e^{-at}	$\frac{1}{s+a}$
$\cos\omega t$	$\frac{s}{s^2+\omega^2}$	t	$\frac{1}{s^2}$
$tx(t)$	$-\frac{dX(s)}{ds}$	t^2	$\frac{2}{s^3}$
$\frac{x(t)}{t}$	$\int_s^\infty X(s)\,ds$	te^{-at}	$\frac{1}{(s+a)^2}$
$\delta(t)$	1	$e^{-at}\sin(\omega t)$	$\frac{\omega}{(s+a)^2+\omega^2}$
$1(t)$	$\frac{1}{s}$	$e^{-at}\cos(\omega t)$	$\frac{s+a}{(s+a)^2+\omega^2}$

ME547: Linear Systems

Modeling: Inverse Laplace Transform

Xu Chen

University of Washington

Common Laplace transform pairs

$f(t)$	$F(s)$	$f(t)$	$F(s)$
$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$	e^{-at}	$\frac{1}{s+a}$
$\cos \omega t$	$\frac{s}{s^2+\omega^2}$	t	$\frac{1}{s^2}$
$tx(t)$	$-\frac{dX(s)}{ds}$	t^2	$\frac{2}{s^3}$
$\frac{x(t)}{t}$	$\int_s^\infty X(s)\, ds$	te^{-at}	$\frac{1}{(s+a)^2}$
$\delta(t)$	1	$e^{-at}\sin(\omega t)$	$\frac{\omega}{(s+a)^2+\omega^2}$
$1(t)$	$\frac{1}{s}$	$e^{-at}\cos(\omega t)$	$\frac{s+a}{(s+a)^2+\omega^2}$

Overview of inverse Laplace transform: modularity and decomposition

- goal: to break a large Laplace transform into small blocks, so that we can use elemental examples of Laplace transfer functions:

$$G(s)=\frac{B(s)}{A(s)}=\frac{B_1(s)}{A_1(s)}+\frac{B_2(s)}{A_2(s)}+\dots$$

- we will use examples to demonstrate strategies for common fractional expansions

Real and distinct roots in $A(s)$

example 1

$$G(s)=\frac{B(s)}{A(s)}=\frac{32}{s(s+4)(s+8)}=\frac{K_1}{s}+\frac{K_2}{s+4}+\frac{K_3}{s+8}$$

residues:

- $K_1=\lim_{s\to 0} sG(s)=1$
- $K_2=\lim_{s\to -4}(s+4)G(s)=-2$
- $K_3=\lim_{s\to -8}(s+8)G(s)=1$

Coding partial fraction expansions

```
% MATLAB
syms s
G = 32/s/(s+4)/(s+8)
partfrac(G)
```

```
# Python
import sympy
s = sympy.symbols('s')
G = 32/s/(s+4)/(s+8)
print(sympy.apart(G))
```

```
1/(s + 8) - 2/(s + 4) + 1/s
```

Real and repeated roots in $A(s)$

example 2

$$G(s)=\frac{2}{(s+1)(s+2)^2}=\frac{K_1}{s+1}+\frac{K_2}{s+2}+\frac{K_3}{(s+2)^2}$$

- $K_3=\lim_{s\to-2}(s+2)^2G(s)=-2$
- $K_1=\lim_{s\to-1}(s+1)G(s)=2$
- for K_2, we hit both sides with $(s+2)^2$ then differentiate once w.r.t. s, to get

$$K_2=\lim_{s\to-2}\frac{d}{ds}(s+2)^2G(s)=-2$$

Coding partial fraction expansions

```python
# Python
import sympy
s = sympy.symbols('s')
G = 2/(s+1)/(s+2)**2
print(sympy.apart(G))
```

```
-2/(s + 2) - 2/(s + 2)**2 + 2/(s + 1)
```

Solution of a first-order ODE

example 1: Let $a > 0, b > 0, y(0) = y_0 \in \mathbb{R}$, obtain the solution to the ODE:

$$\dot{y}(t) = -ay(t) + b1(t)$$

where $1(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}$

- Laplace transform: $\mathcal{L}\{\dot{y}(t)\} = sY(s) - y(0)$
- $\Rightarrow$ solution in Laplace domain:

$$Y(s) = \frac{1}{s+a}y(0) + \frac{b}{s(s+a)} = \frac{1}{s+a}y(0) + \frac{b}{a}\left(\frac{1}{s} - \frac{1}{s+a}\right)$$

- apply inverse Laplace transform: $y(t) = \mathcal{L}^{-1}\{Y(s)\} = \dots$
- solution:

$$y(t) = e^{-at}y(0) + \frac{b}{a}(1(t) - e^{-at})$$

Solution of a first-order ODE

example 1: $a > 0, b > 0, y(0) = y_0 \in \mathbb{R}$:

$$\dot{y}(t) = -ay(t) + b1(t) \Rightarrow Y(s) = \frac{1}{s+a}y(0) + \frac{b}{s(s+a)}$$

$$y(t) = e^{-at}y(0) + \frac{b}{a}(1(t) - e^{-at})$$

observations:

- from the ODE, $y(\infty) = \frac{b}{a}$
- using final value theorem,

$$\lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s) = \frac{b}{a}$$

Solution of a first-order ODE

example 2: Let $a > 0, b > 0, y(0) = y_0 \in \mathbb{R}$, obtain the solution to the ODE:

$$\dot{y}(t) = -ay(t) + b\delta(t)$$

- Laplace transform: $\mathcal{L}\{\dot{y}(t)\} = sY(s) - y_0$
- $\Rightarrow$ solution in Laplace domain: $Y(s) = \frac{1}{s+a}(y_0 + b)$
- apply inverse Laplace transform: $y(t) = \mathcal{L}^{-1}\{Y(s)\} = e^{-at}(y_0 + b)$
- Q: what's the initial value from initial value theorem? what does the impulse do to the initial condition?

Connecting two domains

- n-th order differential equation:

$$\frac{d^n y}{dt^n}+a_{n-1}\frac{d^{n-1}y}{dt^{n-1}}+\cdots+a_1\dot{y}+a_0y=b_m\frac{d^m u}{dt^m}+b_{m-1}\frac{d^{m-1}u}{dt^{m-1}}+\cdots+b_1\dot{u}+b_ou$$

where $y(0)=0, \frac{dy}{dt}|_{t=0}=0, \dots, \frac{d^{n-1}y}{dt^{n-1}}|_{t=0}=0$

- applying Laplace transform yields

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

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

Transfer functions

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

- $A(s)=0$: characteristic equation (C.E.)
- roots of C.E.: poles of $G(s)$
- roots of $B(s)=0$: zeros of $G(s)$
- $m\leq n$: realizability condition
- $G(s)$ is called
  - *proper* if $n\geq m$
  - *strictly proper* if $n>m$
- examples: $G_1(s)=K, \; G_2(s)=\frac{k}{s+a}$

Coding transfer functions in Python

```python
import control as co
import matplotlib.pyplot as plt
import numpy as np
num = [1,2] # Numerator co-efficients
den = [1,2,3] # Denominator co-efficients
sys_tf = co.tf(num,den)
print(sys_tf)
poles = co.pole(sys_tf)
zeros = co.zero(sys_tf)
print('\nSystem Poles = ', poles, '\nSystem Zeros = ', zeros)

T,yout = co.step_response(sys_tf)
plt.figure(1,figsize = (6,4))
plt.plot(T,yout)
plt.grid(True)
plt.ylabel("y")
plt.xlabel("Time (sec)")
plt.show()
```

Coding transfer functions in Python

```python
import control as co
import matplotlib.pyplot as plt
import numpy as np
num = [1,2] # Numerator co-efficients
den = [1,2,3] # Denominator co-efficients
sys_tf = co.tf(num,den)

T,yout = co.step_response(sys_tf)
u1 = np.full((1,len(T)),2) # Create an array of 2's
u2 = np.sin(T)
T,yout_u1 = co.forced_response(sys_tf,T,u1) # Response to input 1
T,yout_u2 = co.forced_response(sys_tf,T,u2) # Response to input 2
```

The DC gain

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_0}$$

- DC gain: the ratio of a stable system's output to its input after all transients have decayed
- can use the Final Value Theorem to find the DC gain:

$$\underline{\text{DC gain of } G(s)} = \lim_{s\to 0} sY(s) = \lim_{s\to 0} sG(s)\frac{1}{s} = \underline{\lim_{s\to 0} G(s)}$$

- example: find the DC gain of $G_1(s) = K$ and $G_2(s) = \frac{k}{s+a}$. Try (i) solve the ODE and (ii) the Final Value Theorem

The DC gain in Matlab and Python

```
% MATLAB
s = tf('s');
G = (2*s+3)/(4*s^2+3*s+1);
dcgain(G)
```

```
# Python
import control as co
s = co.tf('s')
G = (2*s+3)/(4*s**2+3*s+1);
print(co.dcgain(G))
```

The DC gain in Matlab and Python

find the DC gain of the system corresponding to $Y_2(s) = \frac{3}{s-2}$

```
# Python
import control as co
H = co.tf([0, 3],[1, -2])
print(co.dcgain(H))
T, yout = co.step_response(H)
print(yout)
```

- exercise: verify the result in Matlab
- is the result correct?

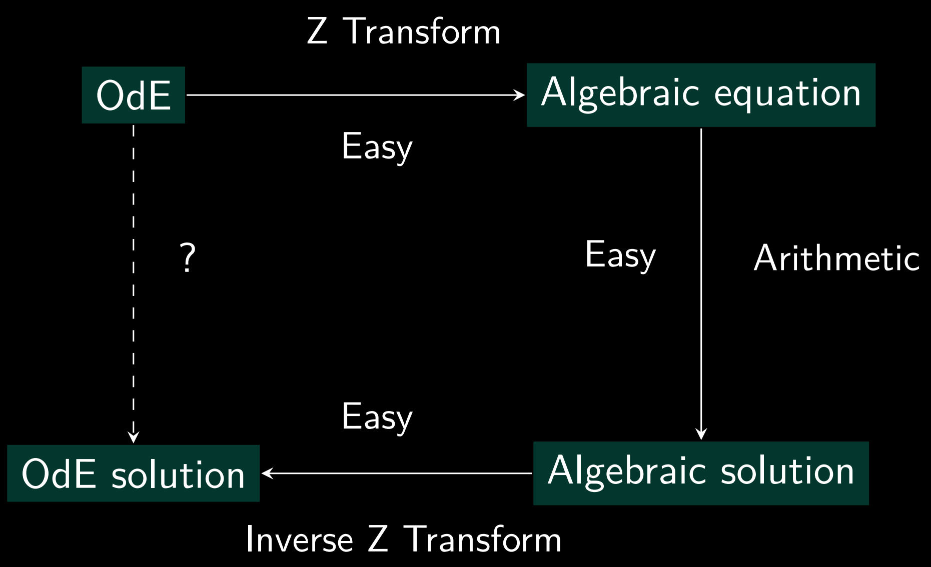
ME547: Linear Systems

Modeling: Z Transform

Xu Chen

University of Washington

The Z transform approach to Ordinary difference Equations (OdEs)



OdE

Z Transform

Easy

Algebraic equation

?

Easy

Arithmetic

Easy

OdE solution

Algebraic solution

Inverse Z Transform

- analogous to Laplace transform for continuous-time signals

Definition

- let $x(k)$ be a real discrete-time sequence that is zero if $k < 0$
- the (one-sided) Z transform of $x(k)$ is

$$\begin{aligned} X(z) &\triangleq \mathcal{Z}\{x(k)\} = \sum_{k=0}^{\infty} x(k) z^{-k} \\ &= x(0) + x(1)z^{-1} + x(2)z^{-2} + \dots \end{aligned}$$

  where $z \in \mathbb{C}$
- a linear operator: $\mathcal{Z}\{\alpha f(k) + \beta g(k)\} = \alpha \mathcal{Z}\{f(k)\} + \beta \mathcal{Z}\{g(k)\}$
- the series $1 + \gamma + \gamma^2 + \dots$ converges to $\frac{1}{1-\gamma}$ for $|\gamma| < 1$ [region of convergence (ROC)]
- (also, recall that $\sum_{k=0}^{N} \gamma^k = \frac{1-\gamma^{N+1}}{1-\gamma}$ if $\gamma \neq 1$)

Example: geometric sequence $\{a^k\}_{k=0}^{\infty}$

$$\boxed{\sum_{k=0}^{\infty} \gamma^k = \frac{1}{1-\gamma}}$$

- $x(k) = a^k$
- $\mathcal{Z}\{a^k\} = \sum_{k=0}^{\infty} a^k z^{-k} = \boxed{\dfrac{1}{1 - az^{-1}}} = \frac{z}{z-a}$

Example: step sequence (discrete-time unit step function)

$$\boxed{\mathcal{Z}\{a^k\} = \frac{1}{1-az^{-1}}}$$

- $1(k) = \begin{cases} 1, & \forall k = 1, 2, \dots \\ 0, & \forall k = \dots, -1, 0 \end{cases}$
- $\mathcal{Z}\{1(k)\} = \mathcal{Z}\{a^k\}\big|_{a=1} = \boxed{\dfrac{1}{1-z^{-1}}} = \frac{z}{z-1}$

Example: discrete-time impulse

- $\delta(k) = \begin{cases} 1, & k = 0 \\ 0, & \text{otherwise} \end{cases}$
- $\mathcal{Z}\{\delta(k)\} = 1$

Exercise: $\cos(\omega_0 k)$

$f(k)$	$F(z)$	ROC
$\delta(k)$	1	All z
$a^k 1(k)$	$\dfrac{1}{1-az^{-1}}$	$\lvert z\rvert > \lvert a\rvert$
$-a^k 1(-k-1)$	$\dfrac{1}{1-az^{-1}}$	$\lvert z\rvert < \lvert a\rvert$
$ka^k 1(k)$	$\dfrac{az^{-1}}{(1-az^{-1})^2}$	$\lvert z\rvert > \lvert a\rvert$
$-ka^k 1(-k-1)$	$\dfrac{az^{-1}}{(1-az^{-1})^2}$	$\lvert z\rvert < \lvert a\rvert$
$\cos(\omega_0 k)$	$\dfrac{1-z^{-1}\cos(\omega_0)}{1-2z^{-1}\cos(\omega_0)+z^{-2}}$	$\lvert z\rvert > 1$
$\sin(\omega_0 k)$	$\dfrac{z^{-1}\sin(\omega_0)}{1-2z^{-1}\cos(\omega_0)+z^{-2}}$	$\lvert z\rvert > 1$
$a^k\cos(\omega_0 k)$	$\dfrac{1-az^{-1}\cos(\omega_0)}{1-2az^{-1}\cos(\omega_0)+a^2z^{-2}}$	$\lvert z\rvert > \lvert a\rvert$
$a^k\sin(\omega_0 k)$	$\dfrac{az^{-1}\sin(\omega_0)}{1-2az^{-1}\cos(\omega_0)+a^2z^{-2}}$	$\lvert z\rvert > \lvert a\rvert$

Properties of Z transform: time shift

- let $\mathcal{Z}\{x(k)\} = X(z)$ and $x(k) = 0\ \forall k < 0$
- one-step delay:

$$\begin{aligned}\mathcal{Z}\{x(k-1)\} &= \sum_{k=0}^{\infty} x(k-1)z^{-k} = \sum_{k=1}^{\infty} x(k-1)z^{-k} + x(-1) \\ &= \sum_{k=1}^{\infty} x(k-1)z^{-(k-1)}z^{-1} + x(-1) \\ &= z^{-1}X(z) + \cancel{x(-1)} = \boxed{z^{-1}X(z)}\end{aligned}$$

- analogously, $\underline{\mathcal{Z}\{x(k+1)\}} = \sum_{k=0}^{\infty} x(k+1)z^{-k} \boxed{= zX(z) - zx(0)}$
- thus, if $x(k+1) = Ax(k) + Bu(k)$ and $x(0) = 0$,

$$zX(z) = AX(z) + BU(z) \Rightarrow X(z) = (zI - A)^{-1}BU(z)$$

provided that $(zI - A)$ is invertible

Solving difference equations

Solve the difference equation

$$y(k) + 3y(k-1) + 2y(k-2) = u(k-2)$$

where $y(-2) = y(-1) = 0$ and $u(k) = 1(k)$.

- $\mathcal{Z}\{y(k-1)\} = z^{-1}\mathcal{Z}\{y(k)\} = z^{-1}Y(z)$
- $\mathcal{Z}\{y(k-2)\} = z^{-1}\mathcal{Z}\{y(k-1)\} = z^{-2}Y(z)$
- $\mathcal{Z}\{U(k-2)\} = z^{-2}U(z)$
- $\Rightarrow (1 + 3z^{-1} + 2z^{-2})Y(z) = z^{-2}U(z)$
- $\Rightarrow \boxed{Y(z) = \frac{1}{z^2 + 3z + 2}U(z)}$

Solving difference equations

Solve the difference equation

$$y(k) + 3y(k-1) + 2y(k-2) = u(k-2)$$

where $y(-2) = y(-1) = 0$ and $u(k) = 1(k)$.

- $\boxed{Y(z) = \frac{1}{z^2 + 3z + 2} U(z) = \frac{1}{(z+2)(z+1)} U(z)}$
- $u(k) = 1(k) \Rightarrow U(z) = 1/(1 - z^{-1})$
- $\Rightarrow Y(z) = \frac{z}{(z-1)(z+2)(z+1)} = \frac{1}{6}\frac{z}{z-1} + \frac{1}{3}\frac{z}{z+2} - \frac{1}{2}\frac{z}{z+1}$ (careful with the partial fraction expansion)
- inverse Z transform then gives $y(k) = \frac{1}{6}1(k) + \frac{1}{3}(-2)^k - \frac{1}{2}(-1)^k,\ k \geq 0$

From difference equation to transfer functions

- general discrete-time OdE:

  $$y(k)+a_{n-1}y(k-1)+\cdots+a_0y(k-n) = b_mu(k+m-n)+\cdots+b_0u(k-n)$$

  where $y(k) = 0\ \forall k < 0$
- applying Z transform to the OdE yields

  $$\left(z^n + a_{n-1}z^{n-1} + \cdots + a_0\right) Y(z) = \left(b_m z^m + b_{m-1}z^{m-1} + \cdots + b_0\right) U(z)$$
- hence

  $$Y(z) = \underbrace{\frac{b_m z^m + b_{m-1}z^{m-1} \cdots + b_1 z + b_0}{z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0}}_{G_{yu}(z)\text{: discrete-time transfer function}} U(z)$$

DC gain of discrete-time transfer functions

- general discrete-time OdE and transfer function:

$$y(k)+a_{n-1}y(k-1)+\cdots+a_0y(k-n)=b_mu(k+m-n)+\cdots+b_0u(k-n)$$

$$Y(z)=\underbrace{\frac{b_mz^m+b_{m-1}z^{m-1}\cdots+b_1z+b_0}{z^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0}}_{G_{yu}(z)\text{: discrete-time transfer function}}U(z)$$

- assuming constant input and convergent output, then at steady state,
  - $y(k)=y(k-1)=\cdots=y(k-n)\triangleq y_{ss}$ and
    $u(k+m-n)=u(k+m-n-1)=\cdots=u(k-n)\triangleq u_{ss}$
  - $y_{ss}+a_{n-1}y_{ss}+\cdots+a_0y_{ss}=b_mu_{ss}+\cdots+b_0u_{ss}$
- thus,

$$\underline{\text{DC gain of } G_{yu}(z)}=\frac{b_m+b_{m-1}+\cdots+b_0}{1+a_{n-1}+\cdots+a_0}=\underline{G_{yu}(z)\big|_{z=1}}$$

Transfer functions in two domains

$$y(k)+a_{n-1}y(k-1)+\cdots+a_0y(k-n)=b_mu(k+m-n)+\cdots+b_0u(k-n)$$

$$\Longleftrightarrow G_{yu}(z)=\frac{B(z)}{A(z)}=\frac{b_mz^m+b_{m-1}z^{m-1}\cdots+b_1z+b_0}{z^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0}$$

v.s.

$$\frac{d^ny(t)}{dt^n}+a_{n-1}\frac{d^{n-1}y(t)}{dt^{n-1}}+\cdots+a_0y(t)=b_m\frac{d^mu(t)}{dt^m}+b_{m-1}\frac{d^{m-1}u(t)}{dt^{m-1}}+\cdots+b_0u(t)$$

$$\Longleftrightarrow G_{yu}(s)=\frac{B(s)}{A(s)}=\frac{b_ms^m+\cdots+b_1s+b_0}{s^n+a_{n-1}s^{n-1}+\cdots+a_1s+a_0}$$

Properties	$G_{yu}(s)$	$G_{yu}(z)$
poles and zeros	roots of $A(s)$ and $B(s)$	roots of $A(z)$ and $B(z)$
causality condition	$n\geq m$	$n\geq m$
DC gain / steady-state response to unit step	$G_{yu}(0)$	$G_{yu}(1)$

Additional useful properties of Z transform

- time shifting (assuming $x(k) = 0$ if $k < 0$):

$$\mathcal{Z}\left\{x\left(k - n_d\right)\right\} = z^{-n_d} X\left(z\right)$$

- Z-domain scaling: $\mathcal{Z}\left\{a^k x\left(k\right)\right\} = X\left(a^{-1} z\right)$
- differentiation: $\mathcal{Z}\left\{k x\left(k\right)\right\} = -z \frac{dX(z)}{dz}$
- time reversal: $\mathcal{Z}\left\{x\left(-k\right)\right\} = X\left(z^{-1}\right)$
- convolution: let $f(k) * g(k) \triangleq \sum_{j=0}^{k} f\left(k - j\right) g\left(j\right)$, then

$$\mathcal{Z}\left\{f(k) * g(k)\right\} = F\left(z\right) G\left(z\right)$$

- initial value theorem: $f(0) = \lim_{z \to \infty} F\left(z\right)$
- final value theorem: $\lim_{k \to \infty} f\left(k\right) = \lim_{z \to 1} \left(z - 1\right) F\left(z\right)$, if $\lim_{k \to \infty} f(k)$ exists and is finite

Mortgage payment

- image you borrow \$100,000 (e.g., for a mortgage)
- annual percent rate: $APR = 4.0\%$
- plan to pay off in 30 years with fixed monthly payments
- interest computed monthly
- what is your monthly payment?

Mortgage payment

- borrow \$100,000 ⇒ initial debt $y(0) = 100,000$
- $APR = 4.0\% \Rightarrow MPR = \frac{4.0\%}{12} = 0.0033$
- pay off in 30 years ($N = 30 \times 12 = 360$ months) ⇒ $y(N) = 0$
- debt at month $k+1$:
  $$y(k+1) = \underbrace{(1+MPR)}_{a} y(k) - \underbrace{b}_{\text{monthly payment}} 1(k)$$
- $$\Rightarrow Y(z) = \frac{z}{z-a} y(0) + \frac{1}{z-a}\frac{b}{1-z^{-1}}$$
  $$\Rightarrow Y(z) = \frac{1}{1-az^{-1}} y(0) + \frac{b}{1-a}\left(\frac{1}{1-az^{-1}} - \frac{1}{1-z^{-1}}\right)$$
- $$\Rightarrow y(k) = a^k y(0) + \frac{b}{1-a}\left(a^k - 1\right)$$
- need $y(N) = 0 \Rightarrow a^N y(0) = -\frac{b}{1-a}\left(a^N - 1\right)$
- $$\Rightarrow b = \frac{a^N y(0)(a-1)}{a^N - 1} = \$477.42$$

ME547: Linear Systems

Modeling: State-Space Models

Xu Chen

University of Washington

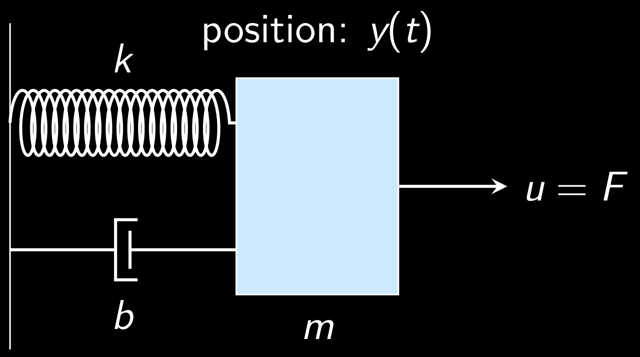
Why state space?

- static/memoryless system: *present* output depends only on its present input: $y(k) = f(u(k))$
- dynamic system: *present* output depends on past and its present input,
  - e.g., $y(k) = f(u(k), u(k-1), \dots, u(k-n), \dots)$
  - described by differential or difference equations, or have time delays
- how much information from the past is needed?

The concept of states of a dynamic system

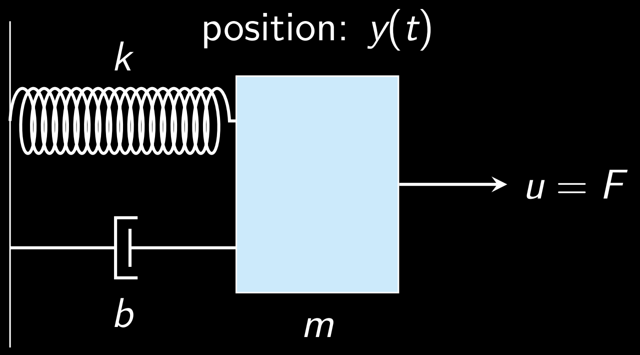
- the *state* $x(t)$ is the information you need at time t that together with future values of the input, will let you compute future values of the output y
- loosely speaking:
  - the "aggregated effect of past inputs"
  - the necessary "memory" that the dynamic system keeps at each time instance

Example



- to predict the future motion, we need to know
  - *current* position and velocity
  - *future* force
- $\Rightarrow$ states: position and velocity

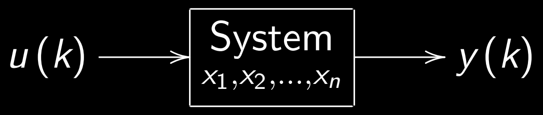
The order of a dynamic system



- the number, n of state variables that is *necessary and sufficient* to uniquely describe the system
- for a given dynamic system,
  - the choice of state variables is *not unique*
  - however, its order n is fixed
  - i.e. you need not more than n but not less than n state variables

States of a discrete-time system

consider a discrete-time dynamic system:



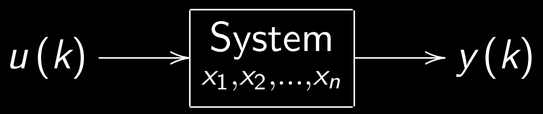
$u(k)$ → System $x_1, x_2, \ldots, x_n$ → $y(k)$

- the state at any instance k_o is the minimum set of variables,

$$x_1(k_o), x_2(k_o), \cdots, x_n(k_o)$$

  that fully describe the system and its response for $k \geq k_o$ to any given set of inputs
- loosely speaking, $x_1(k_o), x_2(k_o), \cdots, x_n(k_o)$ defines the system's memory

Discrete-time state-space description



general case

$$x(k+1) = f(x(k), u(k), k)$$
$$y(k) = h(x(k), u(k), k)$$

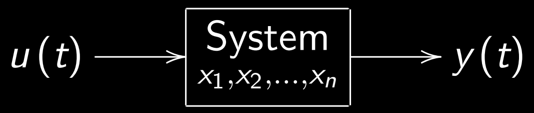
- $u(k)$: input
- $y(k)$: output
- $x(k)$: state
- $x(k+1) = f(\cdot)$: state Eq
- $y(k) = h(\cdot)$: output Eq

linear time-invariant (LTI) case

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k) = Cx(k) + Du(k)$$

- $\Sigma(A, B, C, D)$ denotes a state-space realization
- also written as $\Sigma = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$

Continuous-time state-space description



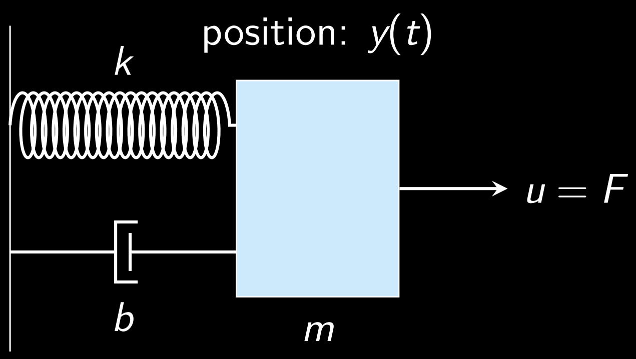
general case

$$\frac{dx(t)}{dt} = f(x(t), u(t), t)$$
$$y(t) = h(x(t), u(t), t)$$

LTI case

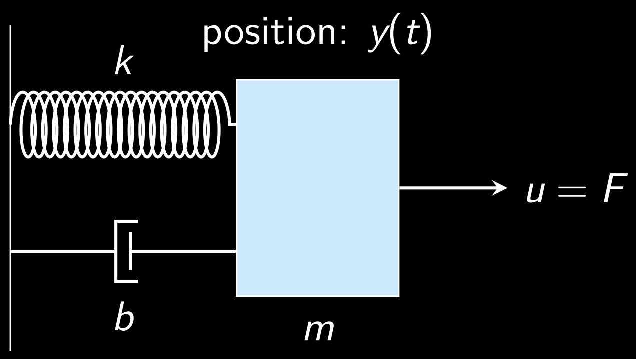
$$\frac{dx(t)}{dt} = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$

Example: mass-spring-damper



$$x(t) = \begin{bmatrix} \overbrace{p(t)}^{\text{mass position}} \\ \underbrace{v(t)}_{\text{mass velocity}} \end{bmatrix} \in \mathbb{R}^2$$

Example: mass-spring-damper



$$\frac{d}{dt}\underbrace{\begin{bmatrix} p(t) \\ v(t) \end{bmatrix}}_{x(t)} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} p(t) \\ v(t) \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}}_{B} u(t)$$

$$y(t) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C}\underbrace{\begin{bmatrix} p(t) \\ v(t) \end{bmatrix}}_{x(t)}$$

Coding a state-space system in MATLAB

```
A = [0,1;-3,-2];
B = [0;1];
C = [2,1];
D = 0;
sys_ss = ss(A,B,C,D)

[yout, T] = step(sys_ss);
figure, plot(T, yout)
```

Coding a state-space system in Python

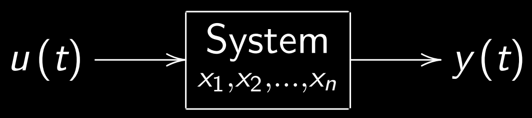
```
import control as co
import matplotlib.pyplot as plt
import numpy as np
A = np.array([[0,1],[-3,-2]])
B = np.array([[0],[1]])
C = np.array([2,1])
D = np.array([0])

sys_ss = co.ss(A,B,C,D)
print(sys_ss)

T,yout = co.step_response(sys_ss)

plt.figure(1,figsize = (6,4))
plt.plot(T,yout)
plt.grid(True)
plt.ylabel("y")
plt.xlabel("Time (sec)")
plt.show()
```

Modeling: Relationship Between State-Space Models and Transfer Functions

Xu Chen

University of Washington

Continuous-time LTI state-space description


$u(t)$

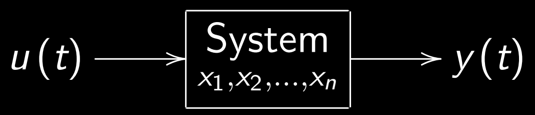
System

$x_1, x_2, \ldots, x_n$

$y(t)$

$$
\begin{aligned}
\frac{d}{dt}x(t) &= Ax(t) + Bu(t)\\
y(t) &= Cx(t) + Du(t)
\end{aligned}
$$

Recap: LTI input/output description



let $u(t) \in \mathbb{R}$ and $y(t) \in \mathbb{R}$, then

$$\boxed{\begin{aligned} y(t) &= (g \star u)(t) \\ &= \int_0^t g(t-\tau)u(\tau)d\tau \end{aligned}}$$

where $g(t)$ is the system's impulse response

Laplace domain:

$$\boxed{Y(s) = G(s)U(s)}$$

$Y(s) = \mathcal{L}\{y(t)\}, U(s) = \mathcal{L}\{u(t)\}, G(s) = \mathcal{L}\{g(t)\}$

From state space to transfer function

given $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$, $C \in \mathbb{R}^{1\times n}$, $D \in \mathbb{R}$,

$$\begin{aligned} \frac{d}{dt}x(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

$\overset{\mathcal{L}}{\Rightarrow}$

$$\begin{aligned} sX(s) - x(0) &= AX(s) + BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned}$$

when $x(0) = 0$, we have

$$\boxed{\frac{Y(s)}{U(s)} = C(sI-A)^{-1}B + D \triangleq: G(s)}$$

–the transfer function between u and y

Analogously for discrete-time systems

for $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$, $C \in \mathbb{R}^{1\times n}$, $D \in \mathbb{R}$,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned}$$

$\overset{\mathcal{Z}}{\Longrightarrow}$

$$\begin{aligned} zX(z) - zx(0) &= AX(z) + BU(z) \\ Y(z) &= CX(z) + DU(z) \end{aligned}$$

when $x(0) = 0$, we have

$$\boxed{\frac{Y(z)}{U(z)} = C(zI - A)^{-1}B + D \triangleq: G(z)}$$

–the transfer function between u and y

From state space to transfer function: Observations

$$\begin{aligned} \frac{d}{dt}x(t) &= A_{n\times n}x(t) + B_{n\times 1}u(t) \\ y(t) &= C_{1\times n}x(t) + Du(t) \end{aligned}$$

- dimensions:

$$G(s) = \underbrace{C}_{1\times n}\underbrace{(sI - A)^{-1}}_{n\times n}\underbrace{B}_{n\times 1} + D$$

$$\Sigma = \left[\begin{array}{c|c} A_{n\times n} & B_{n\times 1} \\ \hline C_{1\times n} & D_{1\times 1} \end{array}\right]$$

- uniqueness: $G(s)$ is unique given the state-space model

Matrix inverse

$$M^{-1} = \frac{1}{\det(M)}\text{Adj}(M)$$

where $\text{Adj}(M) = \{\text{Cofactor matrix of } M\}^T$

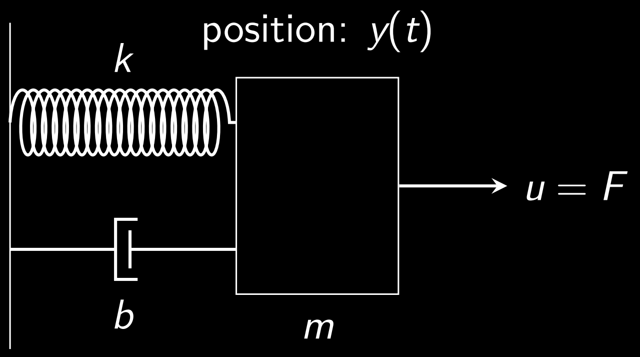
e.g.: $M = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 1 & 0 & 6 \end{bmatrix}$, $\{\text{Cofactor matrix of } M\} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$

where $c_{11} = \begin{vmatrix} 4 & 5 \\ 0 & 6 \end{vmatrix} = 24,\ c_{12} = -\begin{vmatrix} 0 & 5 \\ 1 & 6 \end{vmatrix} = 5,\ c_{13} = \begin{vmatrix} 0 & 4 \\ 1 & 0 \end{vmatrix} = -4,$

$c_{21} = -\begin{vmatrix} 2 & 3 \\ 0 & 6 \end{vmatrix} = -12,\ c_{22} = \begin{vmatrix} 1 & 3 \\ 1 & 6 \end{vmatrix} = 3,\ c_{23} = -\begin{vmatrix} 1 & 2 \\ 1 & 0 \end{vmatrix} = 2,$

$c_{31} = \begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = -2,\ c_{32} = -\begin{vmatrix} 1 & 3 \\ 0 & 5 \end{vmatrix} = -5,\ c_{33} = \begin{vmatrix} 1 & 2 \\ 0 & 4 \end{vmatrix} = 4$

Mass-spring-damper



$$\frac{d}{dt}\underbrace{\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}}_{x(t)} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}}_{B} u(t)$$

$$y(t) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C}\underbrace{\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}}_{x(t)}$$

Mass-spring-damper

$$
\begin{aligned}
\frac{d}{dt}\begin{bmatrix} y(t) \\ v(t) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}\begin{bmatrix} y(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t) \\
y(t) &= \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}
\end{aligned}
$$

$$\boxed{G(s) = C(sI - A)^{-1}B + D}$$

$\Rightarrow$

$$G(s) = \begin{bmatrix} 1 & 0 \end{bmatrix}\left[\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}\right]^{-1}\begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}$$

Mass-spring-damper

$$
\begin{aligned}
\left[\begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}\right]^{-1} &= \begin{bmatrix} s & -1 \\ \frac{k}{m} & s + \frac{b}{m} \end{bmatrix}^{-1} \\
&= \frac{1}{s^2 + \frac{b}{m}s + \frac{k}{m}}\begin{bmatrix} s + \frac{b}{m} & 1 \\ -\frac{k}{m} & s \end{bmatrix}
\end{aligned}
$$

Mass-spring-damper

Putting the inverse in yields

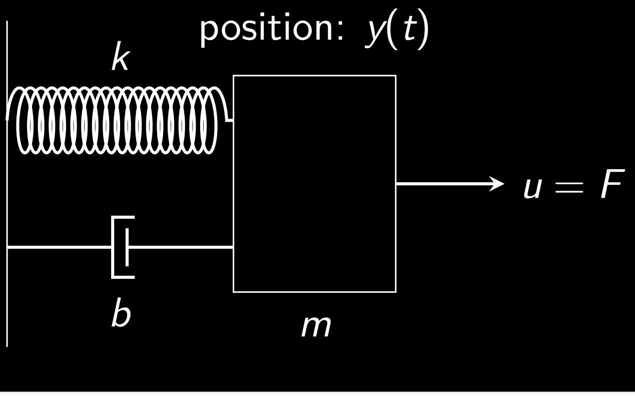
$$G(s) = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s & -1 \\ \frac{k}{m} & s + \frac{b}{m} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}$$

$$= \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s + \frac{b}{m} & 1 \\ -\frac{k}{m} & s \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}}{s^2 + \frac{b}{m}s + \frac{k}{m}}$$

namely

$$G(s) = \frac{\frac{1}{m}}{s^2 + \frac{b}{m}s + \frac{k}{m}}$$

Numerical example in MATLAB



```
m = 1; k = 2; b = 1;
A = [0 1; -k/m -b/m];
B = [0; 1/m];
C = [1 0];
D = 0;
sys = ss(A,B,C,D)
[num,den] = ss2tf(A,B,C,D);
sys_tf = tf(num,den)
figure, step(sys)
figure, step(sys_tf)
```

Numerical example in Python

```python
import control as co
import numpy as np
m = 1
k = 2
b = 1
A = np.array([[0,1],[-k/m,-b/m]])
B = np.array([[0], [1/m]])
C = np.array([1,0])
D = np.array([0])
sys = co.ss(A,B,C,D)
print(sys)
sys_tf = co.ss2tf(sys)
print(sys_tf)

print(co.poles(sys))
print(co.poles(sys_tf))
```

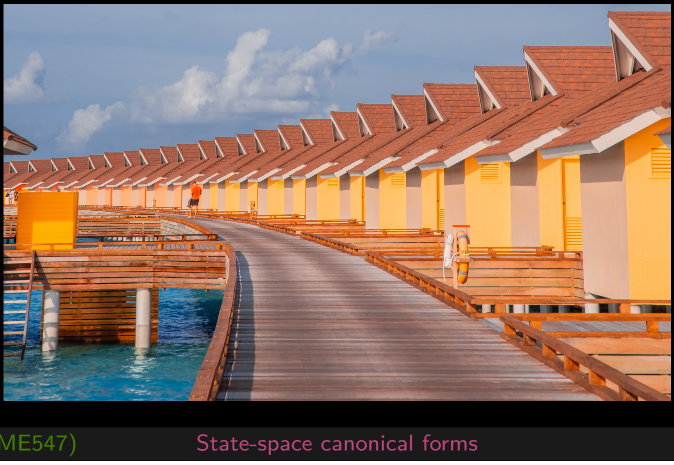
Exercise

Given the following state-space system parameters: $A = \begin{bmatrix} 0 & -6 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$, $B = \begin{bmatrix} -6 & 0 & -3 \\ -2 & 1 & 0 \\ 0 & 2 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$, obtain the transfer function $G(s)$.

Canonical Forms of State-Space Systems

Xu Chen

University of Washington



Goal

the realization problem:

$$\boxed{G(s)=\frac{B(s)}{A(s)}}\overset{?}{\Longrightarrow}\boxed{\Sigma=\left[\begin{array}{c|c}A & B\\\hline C & D\end{array}\right]}$$

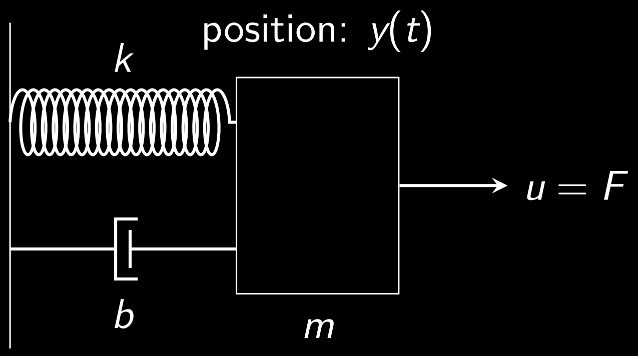
- existence and uniqueness: the same system can have infinite amount of state-space representations: e.g.

$$\begin{cases}\dot{x} &= Ax+Bu\\ y &= Cx\end{cases}\quad\begin{cases}\dot{x} &= Ax+\frac{1}{2}Bu\\ y &= 2Cx\end{cases}$$

- canonical realizations exist
- relationship between different realizations?
- unit problem:

$$G(s)=\frac{b_2s^2+b_1s+b_0}{s^3+a_2s^2+a_1s+a_0}$$

Recall



- $G(s) = \frac{1}{ms^2+bs+k}$
- chose position $y(t)$ and velocity $\dot{y}(t)$ as state variables

From spring mass damper to modules with unity numerator

$$u \longrightarrow \boxed{\frac{1}{s^3+a_2s^2+a_1s+a_0}} \longrightarrow y$$

- choose similarly:

$$x_1 = y,\ x_2 = \dot{x}_1 = \dot{y},\ x_3 = \dot{x}_2 = \ddot{y}$$

- $\Rightarrow$

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Controllable canonical form (ccf)

$$u \longrightarrow \boxed{\frac{b_2s^2+b_1s+b_0}{s^3+a_2s^2+a_1s+a_0}} \longrightarrow y$$

- choose x_1 such that

$$u \longrightarrow \boxed{\frac{1}{s^3+a_2s^2+a_1s+a_0}} \xrightarrow{x_1} \boxed{b_2s^2+b_1s+b_0} \longrightarrow y$$

- the first part

$$u \longrightarrow \boxed{\frac{1}{s^3+a_2s^2+a_1s+a_0}} \longrightarrow \tilde{y}(=x_1)$$

  is now familiar

Controllable canonical form (ccf)

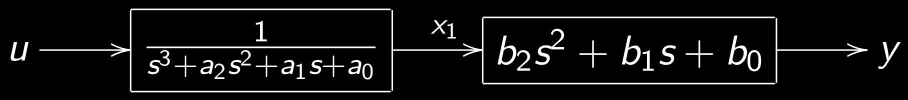
$$u \longrightarrow \boxed{\frac{1}{s^3+a_2s^2+a_1s+a_0}} \xrightarrow{x_1} \boxed{b_2s^2+b_1s+b_0} \longrightarrow y$$

$$X_1(s) = \frac{U(s)}{s^3+a_2s^2+a_1s+a_0} \Rightarrow \dddot{x}_1 + a_2\ddot{x}_1 + a_1\dot{x}_1 + a_0x_1 = u$$

- let $x_2 = \dot{x}_1,\ x_3 = \dot{x}_2 \Rightarrow \dot{x}_3 = -a_2x_3 - a_1x_2 - a_0x_1 + u$
- $\Rightarrow$

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t)$$

Controllable canonical form (ccf)



$$X_1(s)=\frac{U(s)}{s^3+a_2s^2+a_1s+a_0}\Rightarrow \dddot{x}_1+a_2\ddot{x}_1+a_1\dot{x}_1+a_0x_1=u$$

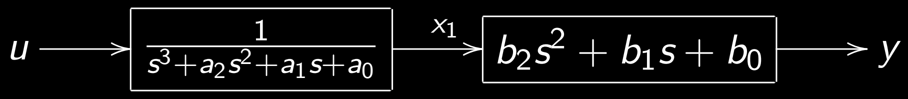
- let $x_2=\dot{x}_1,\ x_3=\dot{x}_2\Rightarrow\dot{x}_3=-a_2x_3-a_1x_2-a_0x_1+u$
- for the output:

$$Y(s)=\left(b_2s^2+b_1s+b_0\right)X_1(s)\Rightarrow y=b_2\underbrace{\ddot{x}_1}_{x_3}+b_1\underbrace{\dot{x}_1}_{x_2}+b_0x_1$$

- $\Rightarrow$

$$y(t)=\left[\begin{array}{ccc}b_0 & b_1 & b_2\end{array}\right]\left[\begin{array}{c}x_1(t)\\x_2(t)\\x_3(t)\end{array}\right]$$

Controllable canonical form (ccf)



- $x_2=\dot{x}_1,\ x_3=\dot{x}_2$
- $y=b_2\underbrace{\ddot{x}_1}_{x_3}+b_1\underbrace{\dot{x}_1}_{x_2}+b_0x_1$
- putting everything in matrix form:

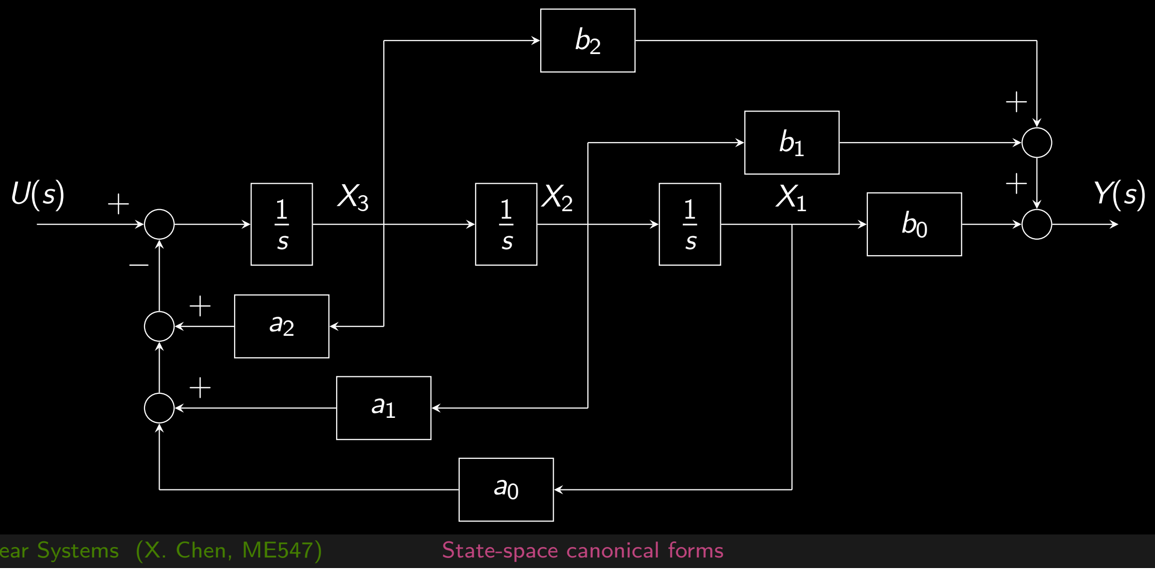
$$\frac{d}{dt}\left[\begin{array}{c}x_1(t)\\x_2(t)\\x_3(t)\end{array}\right]=\left[\begin{array}{ccc}0 & 1 & 0\\0 & 0 & 1\\-a_0 & -a_1 & -a_2\end{array}\right]\left[\begin{array}{c}x_1(t)\\x_2(t)\\x_3(t)\end{array}\right]+\left[\begin{array}{c}0\\0\\1\end{array}\right]u(t)$$

$$y(t)=\left[\begin{array}{ccc}b_0 & b_1 & b_2\end{array}\right]\left[\begin{array}{c}x_1(t)\\x_2(t)\\x_3(t)\end{array}\right]$$

Block diagram realization of ccf

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t)$$

$$y(t) = \begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix}$$



General ccf

general single-input single-output transfer function:

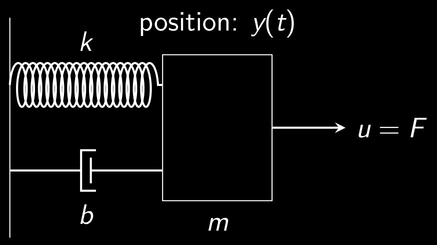
$$G(s) = \frac{b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} + d$$

- the following realizes $G(s)$

$$\Sigma_c = \left[\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right] = \left[\begin{array}{ccccc|c} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \vdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} & 1 \\ \hline b_0 & b_1 & \cdots & b_{n-2} & b_{n-1} & d \end{array}\right]$$

- this realization is called the *controllable canonical form*

ccf example



$$\frac{d}{dt}\underbrace{\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}}_{x(t)} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}}_{A}\underbrace{\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}}_{x(t)} + \underbrace{\begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix}}_{B} u(t)$$

$$y(t) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{C}\underbrace{\begin{bmatrix} y(t) \\ v(t) \end{bmatrix}}_{x(t)}$$

a slightly modified form of the ccf ⇒

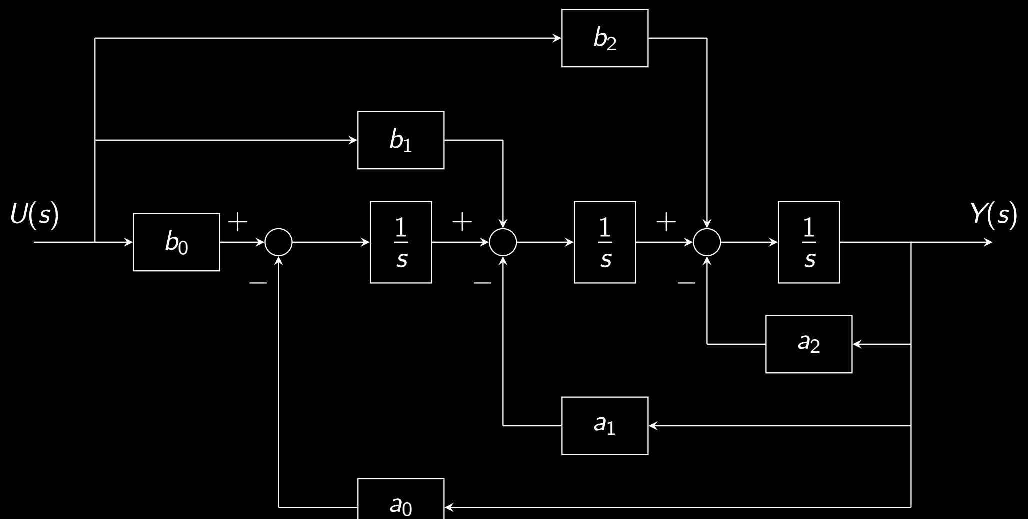
$$G(s) = \frac{1}{m}\frac{1}{s^2 + \frac{b}{m}s + \frac{k}{m}} = \frac{1}{ms^2 + bs + k}$$

Observable canonical form (ocf)

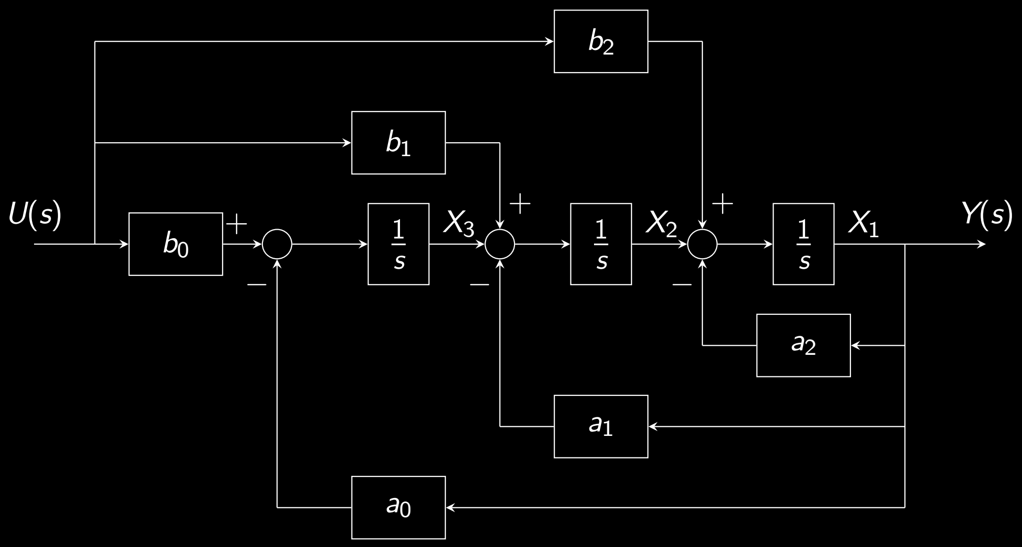
$$Y(s) = \frac{b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} U(s)$$

$$\Rightarrow Y(s) = -\frac{a_2}{s} Y(s) - \frac{a_1}{s^2} Y(s) - \frac{a_0}{s^3} Y(s) + \frac{b_2}{s} U(s) + \frac{b_1}{s^2} U(s) + \frac{b_0}{s^3} U(s)$$

in a block diagram, the above looks like



Observable canonical form



here, the states are connected by

$$\begin{aligned} Y(s) &= X_1(s) & y(t) &= x_1(t) \\ sX_1(s) &= -a_2X_1(s) + X_2(s) + b_2U(s) & \dot{x}_1(t) &= -a_2x_1(t) + x_2(t) + b_2u(t) \\ sX_2(s) &= -a_1X_1(s) + X_3(s) + b_1U(s) \quad \Rightarrow & \dot{x}_2(t) &= -a_1x_1(t) + x_3(t) + b_1u(t) \\ sX_3(s) &= -a_0X_1(s) + b_0U(s) & \dot{x}_3(t) &= -a_0x_1(t) + b_0u(t) \end{aligned}$$

Observable canonical form

$$\begin{cases} \dot{x}_1(t) &= -a_2x_1(t) + x_2(t) + b_2u(t) \\ \dot{x}_2(t) &= -a_1x_1(t) + x_3(t) + b_1u(t) \\ \dot{x}_3(t) &= -a_0x_1(t) + b_0u(t) \\ y(t) &= x_1(t) \end{cases}$$

$$\Rightarrow \dot{x}(t) = \underbrace{\begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix}}_{A_o} x(t) + \underbrace{\begin{bmatrix} b_2 \\ b_1 \\ b_0 \end{bmatrix}}_{B_o} u(t)$$

$$y(t) = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{C_o} x(t)$$

this is called the *observable canonical form* realization of $G(s)$

General ocf

general case for:

$$G(s)=\frac{b_{n-1}s^{n-1}+\cdots+b_1s+b_0}{s^n+a_{n-1}s^{n-1}+\cdots+a_1s+a_0}+d$$

observable canonical form:

$$\Sigma_o=\left[\begin{array}{c|c} A_o & B_o \\ \hline C_o & D_o \end{array}\right]=\left[\begin{array}{ccccc|c} -a_{n-1} & 1 & 0 & \dots & 0 & b_{n-1} \\ -a_{n-2} & 0 & \ddots & \ddots & \vdots & b_{n-2} \\ \vdots & \vdots & \ddots & \ddots & 0 & \vdots \\ -a_1 & 0 & \cdots & 0 & 1 & b_1 \\ -a_0 & 0 & \cdots & 0 & 0 & b_0 \\ \hline 1 & 0 & \cdots & 0 & 0 & d \end{array}\right]$$

ocf in Python

```python
import control as ct
Gs  = ct.tf2ss([1,0,1],[1,2,10])
Gc, T = ct.canonical_form(Gs,'observable')

Gc.A
Gc.B
Gc.C
Gc.D
```

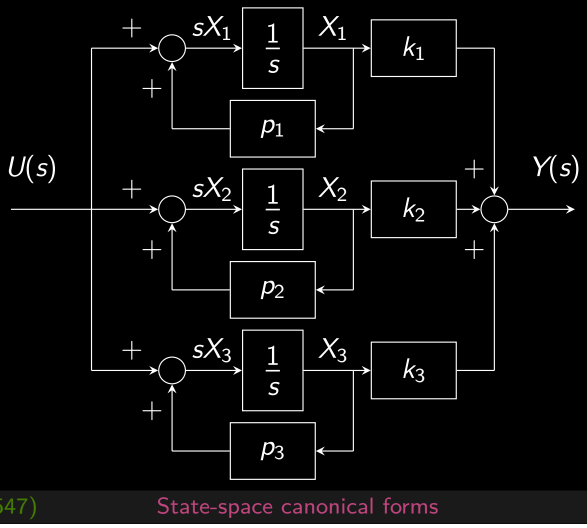
ccf and ocf: no direct Matlab commands

Diagonal form

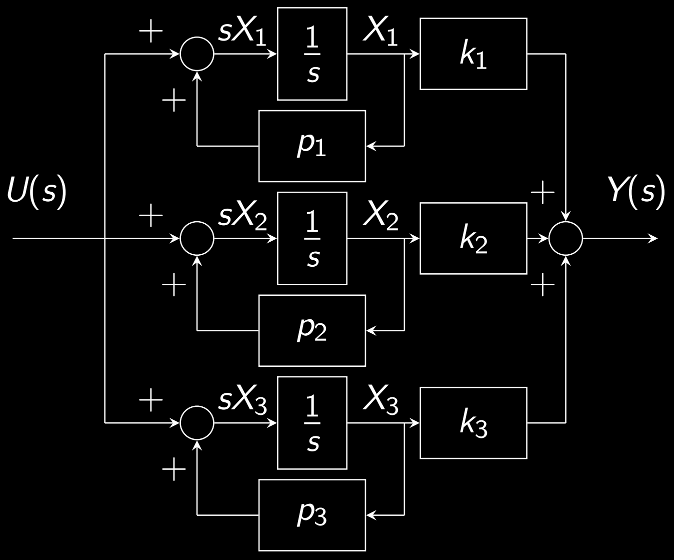
$$G(s) = \frac{B(s)}{A(s)} = \frac{b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0}$$

when the poles $p_1 \neq p_2 \neq p_3$, partial fractional expansion yields

$$G(s) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \frac{k_3}{s - p_3}, \quad k_i = \lim_{p \to p_i} (s - p_i) \frac{B(s)}{A(s)}$$



Diagonal form



state-space realization:

$$\boxed{A = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \; C = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix}, \; D = 0}$$

Jordan form

if poles repeat, say,

$$G(s) = \frac{b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} = \frac{b_2 s^2 + b_1 s + b_0}{(s-p_1)(s-p_m)^2}, \quad p_1 \neq p_m \in \mathbb{R}$$

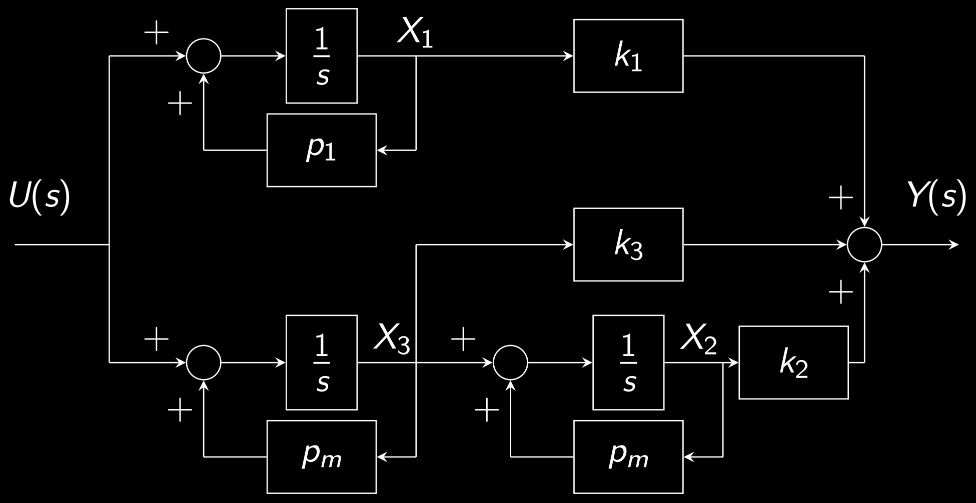
then partial fraction expansion gives

$$G(s) = \frac{k_1}{s-p_1} + \frac{k_2}{(s-p_m)^2} + \frac{k_3}{s-p_m} \quad \text{w/} \begin{cases} k_1 & = \lim_{s\to p_1} G(s)(s-p_1) \\ k_2 & = \lim_{s\to p_m} G(s)(s-p_m)^2 \\ k_3 & = \lim_{s\to p_m} \frac{d}{ds}\left\{G(s)(s-p_m)^2\right\} \end{cases}$$

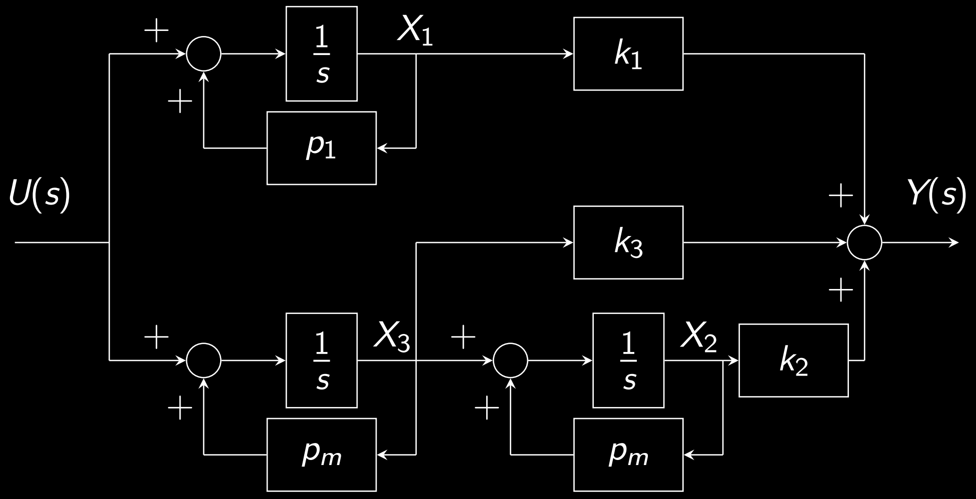
Jordan form

$$G(s) = \frac{k_1}{s-p_1} + \frac{k_2}{(s-p_m)^2} + \frac{k_3}{s-p_m}$$

has the block diagram realization:



Jordan form



state-space realization (called the Jordan canonical form):

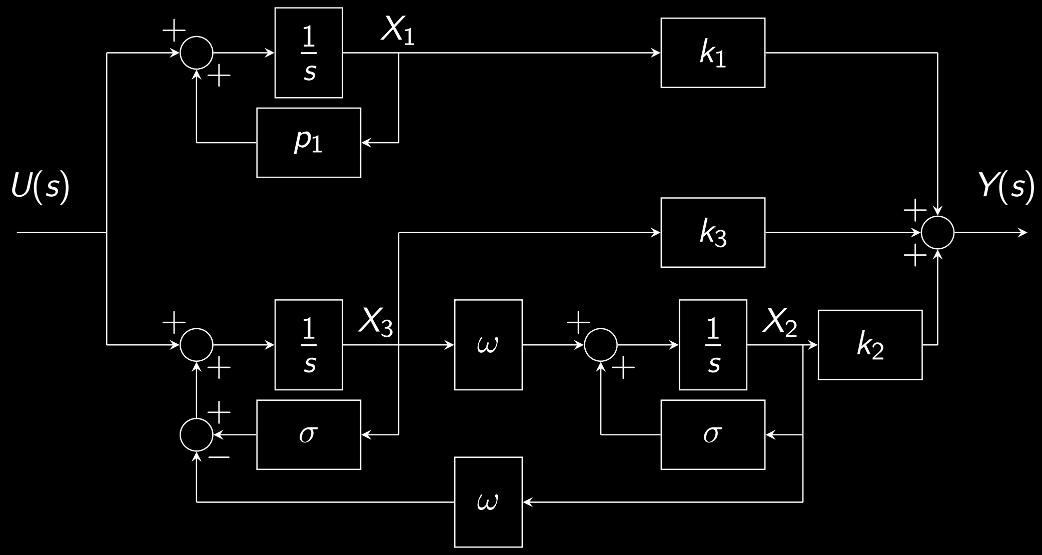
$$\boxed{A = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_m & 1 \\ 0 & 0 & p_m \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ C = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix},\ D = 0}$$

Modified canonical form

if the system has complex poles, say,

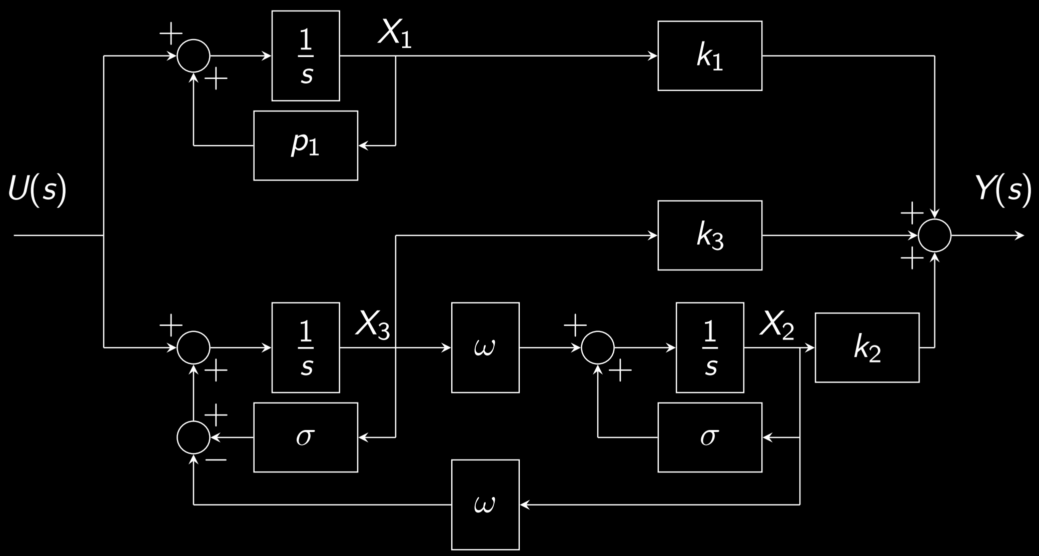
$$G(s) = \frac{b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} = \frac{k_1}{s - p_1} + \frac{\alpha s + \beta}{(s - \sigma)^2 + \omega^2}$$

then



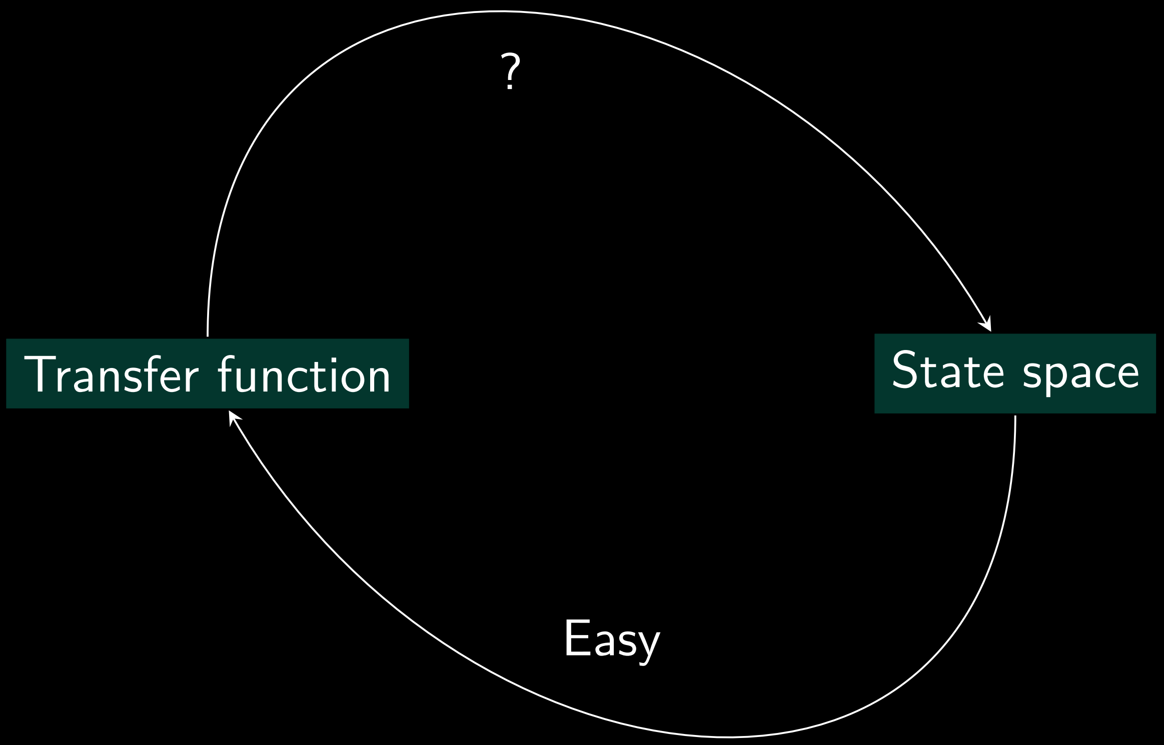
where $k_2 = (\beta + \alpha\sigma)/\omega$ and $k_3 = \alpha$

Modified canonical form



$\Rightarrow$ *modified Jordan form:*

$$\boxed{A=\begin{bmatrix} p_1 & 0 & 0\\ 0 & \sigma & \omega\\ 0 & -\omega & \sigma\end{bmatrix},\ B=\begin{bmatrix}1\\0\\1\end{bmatrix},\ C=\begin{bmatrix} k_1 & k_2 & k_3\end{bmatrix},\ D=0}$$



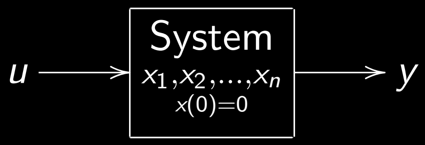
Transfer function

State space

?

Easy

Continuous- and discrete-time state-space descriptions



$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \qquad \begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned}$$

$$\begin{aligned} sX(s) &= AX(s) + BU(s) \\ Y(s) &= CX(s) + DU(s) \end{aligned} \qquad \begin{aligned} zX(z) &= AX(z) + BU(z) \\ Y(z) &= CX(z) + DU(z) \end{aligned}$$

- previous procedure applies to discrete-time systems
- replace t with k, and $\dot{x}(t)$ with $x(k+1)$
- replace s with z, and $\boxed{\frac{1}{s}}$ with $\boxed{z^{-1}}$ in block diagrams

DT controllable canonical form

$$G(z) = \frac{b_2 z^2 + b_1 z + b_0}{z^3 + a_2 z^2 + a_1 z + a_0}$$

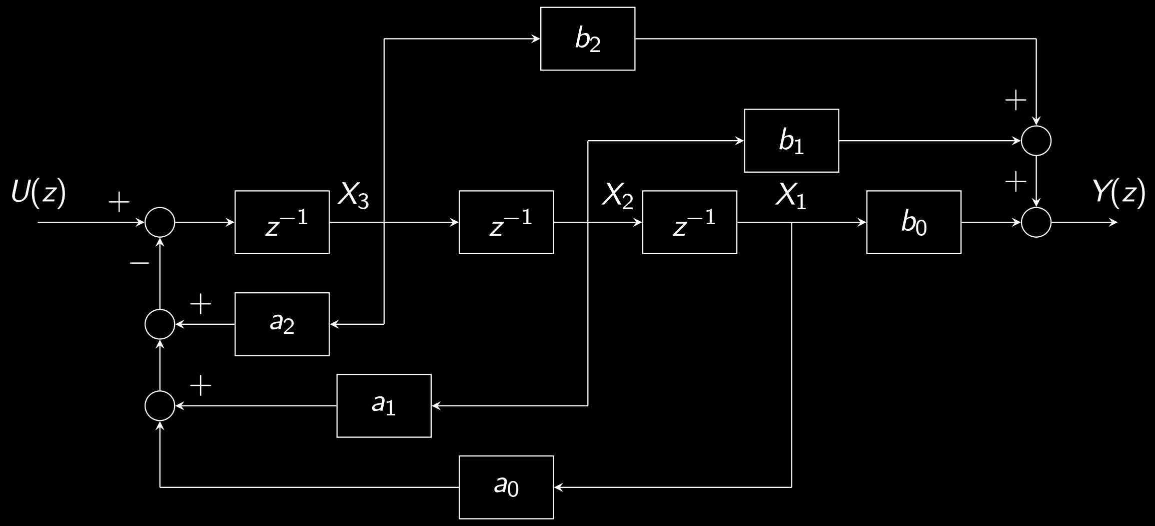
- same transfer-function structure
- $\Rightarrow$ same A, B, C, D matrices as those in CT
- controllable canonical form:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

DT controllable canonical form

$$G(z) = \frac{b_2 z^2 + b_1 z + b_0}{z^3 + a_2 z^2 + a_1 z + a_0}$$



DT observable canonical form

$$G(z)=\frac{b_2z^2+b_1z+b_0}{z^3+a_2z^2+a_1z+a_0}$$

- observable canonical form:

$$\begin{bmatrix} x_1(k+1)\\ x_2(k+1)\\ x_3(k+1)\end{bmatrix}=\begin{bmatrix} -a_2 & 1 & 0\\ -a_1 & 0 & 1\\ -a_0 & 0 & 0\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}+\begin{bmatrix} b_2\\ b_1\\ b_0\end{bmatrix}u(k)$$

$$y(k)=\begin{bmatrix} 1 & 0 & 0\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}$$

DT diagonal form

$$G(z)=\frac{b_2z^2+b_1z+b_0}{z^3+a_2z^2+a_1z+a_0}$$

- diagonal form (distinct poles):

$$G(z)=\frac{k_1}{z-p_1}+\frac{k_2}{z-p_2}+\frac{k_3}{z-p_3}$$

$$\begin{bmatrix} x_1(k+1)\\ x_2(k+1)\\ x_3(k+1)\end{bmatrix}=\begin{bmatrix} p_1 & 0 & 0\\ 0 & p_2 & 0\\ 0 & 0 & p_3\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}+\begin{bmatrix} 1\\ 1\\ 1\end{bmatrix}u(k)$$

$$y(k)=\begin{bmatrix} k_1 & k_2 & k_3\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}$$

DT Jordan form 1

$$G(z)=\frac{b_2z^2+b_1z+b_0}{z^3+a_2z^2+a_1z+a_0}$$

- Jordan form (2 repeated poles):

$$G(z)=\frac{k_1}{z-p_1}+\frac{k_2}{(z-p_m)^2}+\frac{k_3}{z-p_m}$$

$$\begin{bmatrix} x_1(k+1)\\ x_2(k+1)\\ x_3(k+1)\end{bmatrix}=\begin{bmatrix} p_1 & 0 & 0\\ 0 & p_m & 1\\ 0 & 0 & p_m\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}+\begin{bmatrix}1\\0\\1\end{bmatrix}u(k)$$

$$y(k)=\begin{bmatrix} k_1 & k_2 & k_3\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}$$

DT Jordan form 2

$$G(z)=\frac{b_2z^2+b_1z+b_0}{z^3+a_2z^2+a_1z+a_0}$$

- Jordan form (2 complex poles):

$$G(s)=\frac{k_1}{z-p_1}+\frac{\alpha z+\beta}{(z-\sigma)^2+\omega^2}$$

$$\begin{bmatrix} x_1(k+1)\\ x_2(k+1)\\ x_3(k+1)\end{bmatrix}=\begin{bmatrix} p_1 & 0 & 0\\ 0 & \sigma & \omega\\ 0 & -\omega & \sigma\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}+\begin{bmatrix}1\\0\\1\end{bmatrix}u(k)$$

$$y(k)=\begin{bmatrix} k_1 & k_2 & k_3\end{bmatrix}\begin{bmatrix} x_1(k)\\ x_2(k)\\ x_3(k)\end{bmatrix}$$

where $k_2=(\beta+\alpha\sigma)/\omega,\ k_3=\alpha$.

Exercise

obtain the controllable canonical form:

- $G(z) = \frac{z^{-1} - z^{-3}}{1 + 2z^{-1} + z^{-2}}$

Relation between different realizations

- given one realization Σ and a nonsingular $T \in \mathbb{R}^{n\times n}$
- can define *new* states: $Tx^* = x$
- then

$$\dot{x}(t) = Ax(t) + Bu(t) \Rightarrow \frac{d}{dt}\left(Tx^*(t)\right) = ATx^*(t) + Bu(t),$$

$$\Rightarrow \Sigma^* : \begin{cases} \dot{x}^*(t) & = & T^{-1}ATx^*(t) + T^{-1}Bu(t) \\ y(t) & = & CTx^*(t) + Du(t) \end{cases}$$

- namely

$$\Sigma^* = \left[\begin{array}{c|c} T^{-1}AT & T^{-1}B \\ \hline CT & D \end{array}\right]$$

also realizes $G(s)$ and is said to be *similar* to Σ

Relation between different realizations

verify that the following realize the same system

$$\Sigma = \left[\begin{array}{ccc|c} -a_2 & 1 & 0 & b_2 \\ -a_1 & 0 & 1 & b_1 \\ -a_0 & 0 & 0 & b_0 \\ \hline 1 & 0 & 0 & d \end{array}\right], \quad \Sigma^* = \left[\begin{array}{ccc|c} 0 & 0 & -a_0 & b_0 \\ 1 & 0 & -a_1 & b_1 \\ 0 & 1 & -a_2 & b_2 \\ \hline 0 & 0 & 1 & d \end{array}\right]$$

1 From Transfer Function to State Space: State-Space Canonical Forms

It is straightforward to derive the *unique* transfer function corresponding to a state-space model. The inverse problem, i.e., building internal descriptions from transfer functions, is less trivial and is the subject of *realization theory*.

A single transfer function has infinite amount of state-space representations. Consider, for example, the two models

$$\begin{cases} \dot{x} &= Ax + Bu \\ y &= Cx \end{cases}, \qquad \begin{cases} \dot{x} &= Ax + \frac{1}{2}Bu \\ y &= 2Cx \end{cases}$$

which share the same transfer function $C(sI-A)^{-1}B$.

We start with the most common realizations: controller canonical form, observable canonical form, and Jordan form, using the following unit problem:

$$G(s) = \frac{b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0}. \tag{1}$$

1.1 Controllable Canonical Form.

Consider first:

$$Y(s) = \frac{1}{s^3 + a_2 s^2 + a_1 s + a_0} U(s). \tag{2}$$

Similar to choosing position and velocity in the spring-mass-damper example, we can choose

$$x_1 = y,\ x_2 = \dot{x}_1 = \dot{y},\ x_3 = \dot{x}_2 = \ddot{y}, \tag{3}$$

which gives

$$\begin{aligned} \frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned} \tag{4}$$

For the general case in (1), i.e., $\dddot{y} + a_2\ddot{y} + a_1\dot{y} + a_0 y = b_2\ddot{u} + b_1\dot{u} + b_0 u$, there are terms with respect to the derivative of the input. Choosing simply (3) does not generate a proper state equation. However, we can decompose (1) as

$$u \longrightarrow \boxed{\frac{1}{s^3 + a_2 s^2 + a_1 s + a_0}} \longrightarrow \boxed{b_2 s^2 + b_1 s + b_0} \longrightarrow y \tag{5}$$

The first part of the connection

$$u \longrightarrow \boxed{\frac{1}{s^3 + a_2 s^2 + a_1 s + a_0}} \longrightarrow \tilde{y} \tag{6}$$

looks exactly like what we had in (2). Denote the output here as $\tilde{y}$. Then we have

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u,$$

where

$$x_1 = \tilde{y},\ x_2 = \dot{x}_1,\ x_3 = \dot{x}_2. \tag{7}$$

Introducing the states in (7) also addresses the problem of the rising differentiations in u. Notice now, that the second part of (5) is nothing but

$$x_1 \longrightarrow \boxed{b_2 s^2 + b_1 s + b_0} \longrightarrow y$$

So

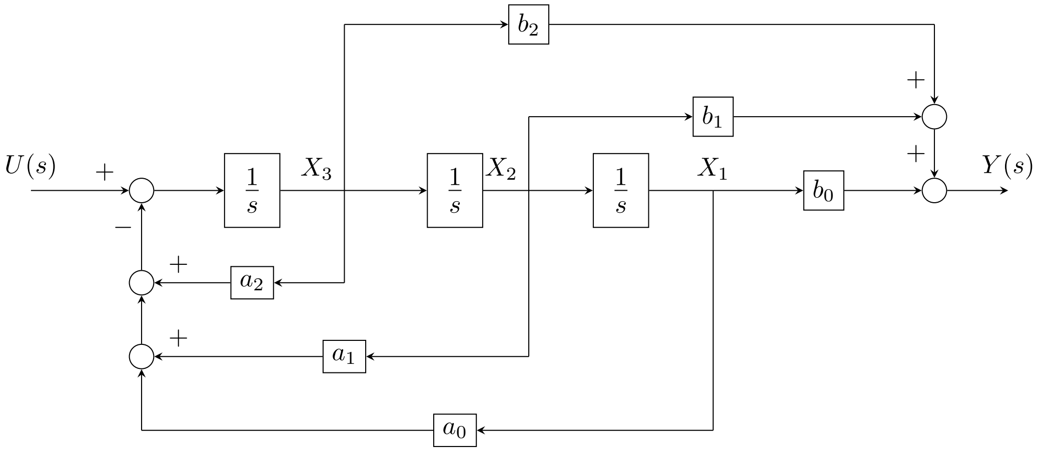
$$y = b_2\ddot{x}_1 + b_1\dot{x}_1 + b_0x_1 = b_2x_3 + b_1x_2 + b_0x_1 = \left[\begin{array}{ccc} b_0 & b_1 & b_2 \end{array}\right]\left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right].$$

The above procedure constructs the *controllable canonical form* of the third-order transfer function (1):

$$\frac{d}{dt}\left[\begin{array}{c} x_1(t) \\ x_2(t) \\ x_3(t) \end{array}\right] = \left[\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{array}\right]\left[\begin{array}{c} x_1(t) \\ x_2(t) \\ x_3(t) \end{array}\right] + \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right] u(t) \tag{8}$$

$$y(t) = \left[\begin{array}{ccc} b_0 & b_1 & b_2 \end{array}\right]\left[\begin{array}{c} x_1(t) \\ x_2(t) \\ x_3(t) \end{array}\right]$$

In a block diagram, the state-space system looks like



Example 1. Obtain the controllable canonical forms of the following systems

- $G(s) = \dfrac{s^2+1}{s^3+2s+10}$
  - Comparing the transfer function with the general form yields $A = \begin{bmatrix} & 1 & \\ & & 1 \\ -10 & -2 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$
- $G(s) = \dfrac{b_0s^2+b_1s+b_2}{s^3+a_0s^2+a_1s+a_2}$
  - Notice the difference in the coefficients. We have $A = \begin{bmatrix} & 1 & \\ & & 1 \\ -a_2 & -a_1 & a_0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $C = \begin{bmatrix} b_2 & b_1 & b_0 \end{bmatrix}$

General Case.

For a single-input single-output transfer function

$$G(s) = \frac{b_{n-1}s^{n-1} + \cdots + b_1s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1s + a_0} + d,$$

we can verify that

$$\Sigma_c = \left[\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right] = \left[\begin{array}{ccccc|c} 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} & 1 \\ \hline b_0 & b_1 & \cdots & b_{n-2} & b_{n-1} & d \end{array}\right] \tag{9}$$

realizes $G(s)$. This realization is called the *controllable canonical form.*

1.2 Observable Canonical Form.

Consider again

$$Y(s) = G(s)U(s) = \frac{b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} U(s).$$

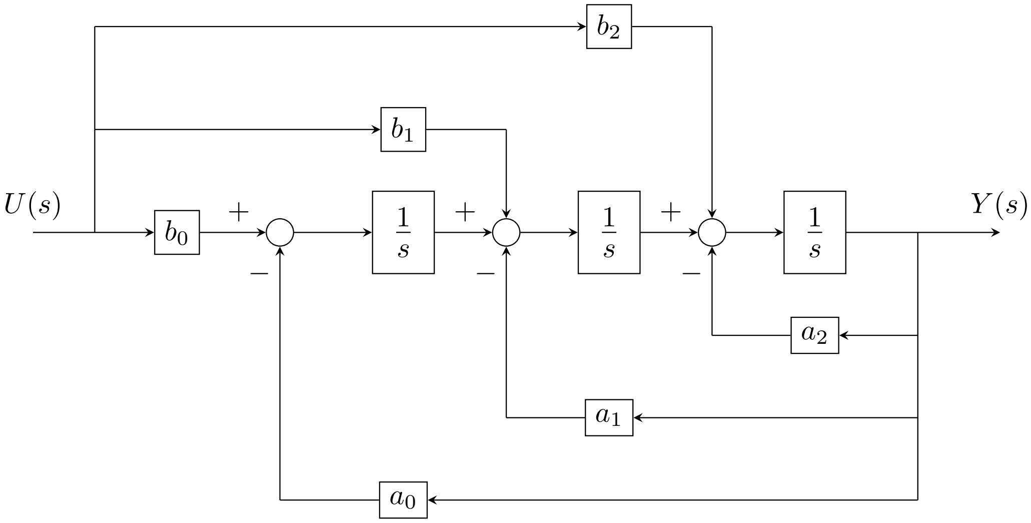
Expanding and dividing by s^3 yield

$$\left(1 + a_2 \frac{1}{s} + a_1 \frac{1}{s^2} + a_0 \frac{1}{s^3}\right) Y(s) = \left(b_2 \frac{1}{s} + b_1 \frac{1}{s^2} + b_0 \frac{1}{s^3}\right) U(s)$$

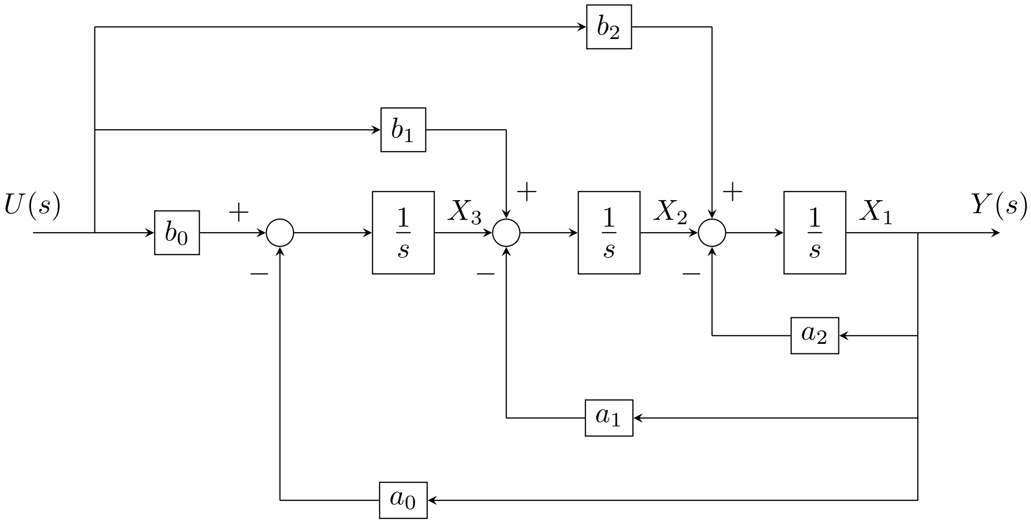
and therefore

$$\begin{aligned} Y(s) = &-a_2 \frac{1}{s} Y(s) - a_1 \frac{1}{s^2} Y(s) - a_0 \frac{1}{s^3} Y(s) \\ &+ b_2 \frac{1}{s} U(s) + b_1 \frac{1}{s^2} U(s) + b_0 \frac{1}{s^3} U(s). \end{aligned}$$

In a block diagram, the above looks like



or more specifically,



Here, the states are connected by

$$\begin{aligned} Y(s) &= X_1(s) \\ sX_1(s) &= -a_2X_1(s) + X_2(s) + b_2U(s) \\ sX_2(s) &= -a_1X_1(s) + X_3(s) + b_1U(s) \\ sX_3(s) &= -a_0X_1(s) + b_0U(s) \end{aligned} \quad \Rightarrow \quad \begin{aligned} y(t) &= x_1(t) \\ \dot{x}_1(t) &= -a_2x_1(t) + x_2(t) + b_2u(t) \\ \dot{x}_2(t) &= -a_1x_1(t) + x_3(t) + b_1u(t) \\ \dot{x}_3(t) &= -a_0x_1(t) + b_0u(t) \end{aligned}$$

or in matrix form:

$$\begin{aligned} \dot{x}(t) &= \underbrace{\begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix}}_{A_o} x(t) + \underbrace{\begin{bmatrix} b_2 \\ b_1 \\ b_0 \end{bmatrix}}_{B_o} u(t) \\ y(t) &= \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{C_o} x(t) \end{aligned} \tag{10}$$

The above is called the *observable canonical form* realization of $G(s)$.

Exercise 1. Verify that $C_o(sI - A_o)^{-1}B_o = G(s)$.

General Case.

In the general case, the *observable canonical form* of the transfer function

$$G(s) = \frac{b_{n-1}s^{n-1} + \cdots + b_1s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1s + a_0} + d$$

is

$$\Sigma_o = \left[\begin{array}{c|c} A_o & B_o \\ \hline C_o & D_o \end{array}\right] = \left[\begin{array}{ccccc|c} -a_{n-1} & 1 & \cdots & 0 & 0 & b_{n-1} \\ -a_{n-2} & 0 & \cdots & 0 & 0 & b_{n-2}0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ -a_1 & 0 & \cdots & 0 & 1 & b_1 \\ -a_0 & & \cdots & & 0 & b_0 \\ \hline 1 & & \cdots & & & d \end{array}\right]. \tag{11}$$

Exercise 2. Obtain the controllable and observable canonical forms of

$$G(s) = \frac{k_1}{s - p_1}.$$

1.3 Diagonal and Jordan canonical forms.

1.3.1 Diagonal form.

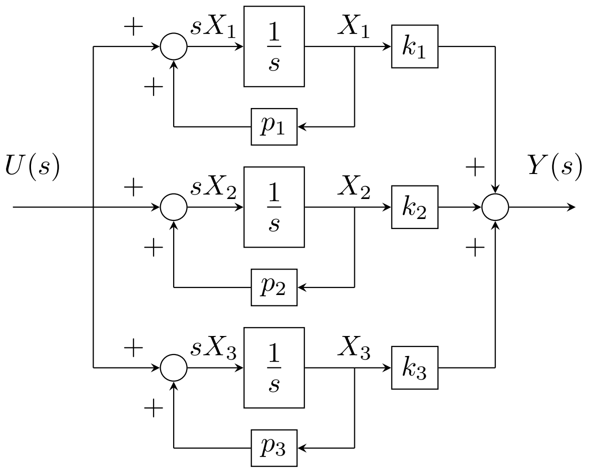
When

$$G(s) = \frac{B(s)}{A(s)} = \frac{b_2s^2 + b_1s + b_0}{s^3 + a_2s^2 + a_1s + a_0}$$

and the poles of the transfer function $p_1 \neq p_2 \neq p_3$, we can write, using partial fractional expansion,

$$G\left(s\right) = \frac{k_1}{s - p_1} + \frac{k_2}{s - p_2} + \frac{k_3}{s - p_3},\ k_i = \lim_{p \to p_i} \left(s - p_i\right) \frac{B(s)}{A(s)},$$

namely



The state-space realization of the above is

$$A=\left[\begin{array}{ccc} p_1 & 0 & 0\\ 0 & p_2 & 0\\ 0 & 0 & p_3\end{array}\right],\ B=\left[\begin{array}{c}1\\1\\1\end{array}\right],\ C=\left[\begin{array}{ccc}k_1 & k_2 & k_3\end{array}\right],\ D=0.$$

1.3.2 Jordan form.

If poles repeat, say,

$$G(s)=\frac{b_2s^2+b_1s+b_0}{s^3+a_2s^2+a_1s+a_0}=\frac{b_2s^2+b_1s+b_0}{(s-p_1)(s-p_m)^2},\ p_1\neq p_m\in\mathbb{R},$$

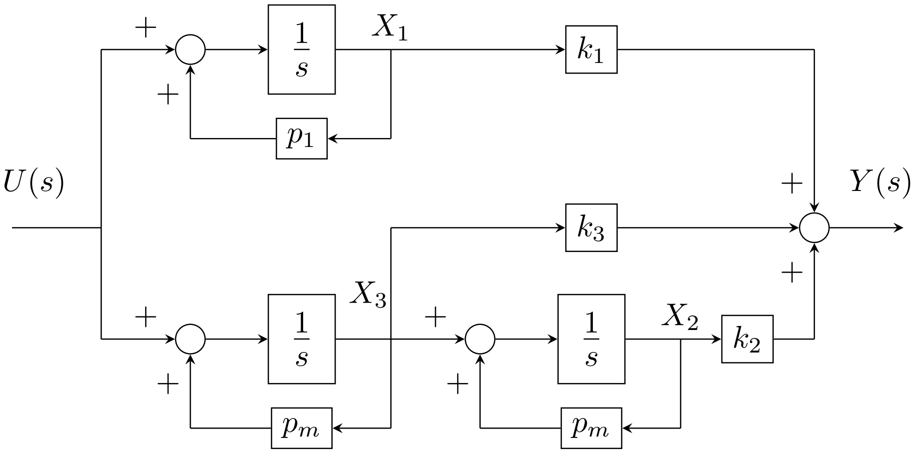
then partial fraction expansion gives

$$G\left(s\right)=\frac{k_1}{s-p_1}+\frac{k_2}{\left(s-p_m\right)^2}+\frac{k_3}{s-p_m},$$

where

$$\begin{aligned} k_1&=\lim_{s\to p_1}G(s)(s-p_1)\\ k_2&=\lim_{s\to p_m}G(s)(s-p_m)^2\\ k_3&=\lim_{s\to p_m}\frac{d}{ds}\left\{G(s)(s-p_m)^2\right\}\end{aligned}$$

In state space, we have



The state-space realization of the above, called the Jordan canonical form,[1] is

$$A=\begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_m & 1 \\ 0 & 0 & p_m \end{bmatrix},\ B=\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ C=\begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix},\ D=0.$$

1.4 Modified canonical form.

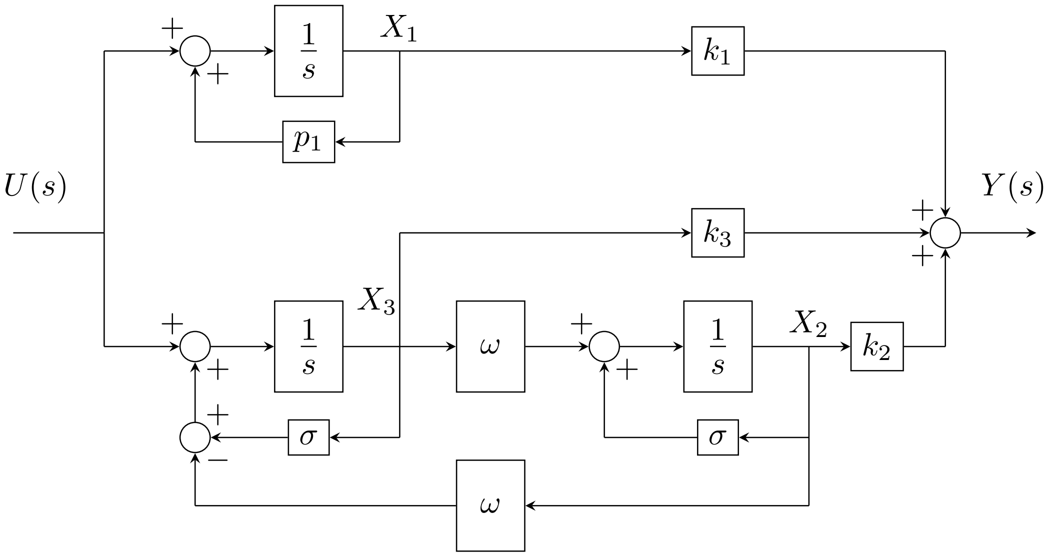
If the system has complex poles, say,

$$G(s)=\frac{b_2s^2+b_1s+b_0}{s^3+a_2s^2+a_1s+a_0}=\frac{b_2s^2+b_1s+b_0}{(s-p_1)\left[(s-\sigma)^2+\omega^2\right]},$$

then partial fraction expansion gives

$$G(s)=\frac{k_1}{s-p_1}+\frac{\alpha s+\beta}{(s-\sigma)^2+\omega^2},$$

which has the graphical representation as below:



Here $k_2=(\beta+\alpha\sigma)/\omega$ and $k_3=\alpha$.

You should be able to check that the block diagram matches with the transfer function realization.

The above can be realized by the modified Jordan form in state space:

$$A=\begin{bmatrix} p_1 & 0 & 0 \\ 0 & \sigma & \omega \\ 0 & -\omega & \sigma \end{bmatrix},\ B=\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ C=\begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix},\ D=0.$$

1.5 Discrete-Time Transfer Functions and Their State-Space Canonical Forms

The procedures for finding state space realizations in discrete time is similar to the continuous time cases. The only difference is that we use

$$\mathcal{Z}\left\{x(k-n)\right\}=z^{-n}X(z),$$

instead of

$$\mathcal{L}\left\{\frac{d^n}{dt^n}x(t)\right\}=s^nX(s),$$

assuming zero state initial conditions.

We have the fundamental relationships:

$$x(k)\longrightarrow\boxed{z^{-1}}\longrightarrow x(k-1)$$

[1] The A matrix is called a Jordan matrix.

$$X(z) \longrightarrow \boxed{z^{-1}} \longrightarrow z^{-1}X(z)$$

$$x(k+n) \longrightarrow \boxed{z^{-1}} \longrightarrow x(k+n-1)$$

The discrete-time state-space description of a general transfer function $G(z)$ is

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned}$$

and satisfies $G(z) = C(zI - A)^{-1}B + D$.

Take again a third-order system as the example:

$$G(z) = \frac{b_2 z^2 + b_1 z + b_0}{z^3 + a_2 z^2 + a_1 z + a_0} = \frac{b_2 z^{-1} + b_1 z^{-2} + b_0 z^{-3}}{1 + a_2 z^{-1} + a_1 z^{-2} + a_0 z^{-3}}.$$

The A, B, C, D matrices of the canonical forms are exactly the same as those in continuous-time cases.

Controllable canonical form:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} b_0 & b_1 & b_2 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

Observable canonical form:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} b_2 \\ b_1 \\ b_0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

Diagonal form (distinct poles):

$$G(z) = \frac{k_1}{z - p_1} + \frac{k_2}{z - p_2} + \frac{k_3}{z - p_3}$$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

Jordan form (2 repeated poles):

$$G(z) = \frac{k_1}{z - p_1} + \frac{k_2}{(z - p_m)^2} + \frac{k_3}{z - p_m}$$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_m & 1 \\ 0 & 0 & p_m \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

Jordan form (2 complex poles):

$$G(s) = \frac{k_1}{z - p_1} + \frac{\alpha z + \beta}{(z - \sigma)^2 + \omega^2}$$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & \sigma & \omega \\ 0 & -\omega & \sigma \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

where $k_2 = (\beta + \alpha\sigma)/\omega$, $k_3 = \alpha$.

Exercise: obtain the controllable canonical form for the following systems

- $G(s) = \frac{z^{-1} - z^{-3}}{1 + 2z^{-1} + z^{-2}}$
- $G(s) = \frac{b_0 z^2 + b_1 z + b_2}{z^3 + a_0 z^2 + a_1 z + a_2}$

1.6 Similar Realizations

Besides the canonical forms, other system realizations exist. Let us begin with the realization Σ of some transfer function $G(s)$. Let $T \in \mathbb{C}^{n \times n}$ be nonsingular. We can define *new* states by:

$$Tx^* = x.$$

We can rewrite the differential equations defining Σ in terms of these new states by plugging in $x = Tx^*$:

$$\frac{d}{dt}(Tx^*(t)) = ATx^*(t) + Bu(t),$$

to obtain

$$\Sigma^* : \begin{cases} \dot{x}^*(t) &= T^{-1}ATx^*(t) + T^{-1}Bu(t) \\ y(t) &= CTx^*(t) + Du(t) \end{cases}$$

This new realization

$$\Sigma^* = \left[\begin{array}{c|c} T^{-1}AT & T^{-1}B \\ \hline CT & D \end{array}\right], \tag{12}$$

also realizes $G(s)$ and is said to be *similar* to Σ.

Similar realizations are fundamentally the same. Indeed, we arrived at Σ_{new} from Σ via nothing more than a change of variables.

Exercise 3 (Another observable canonical form.)**.** Verify that

$$\Sigma = \left[\begin{array}{ccc|c} -a_2 & 1 & 0 & b_2 \\ -a_1 & 0 & 1 & b_1 \\ -a_0 & 0 & 0 & b_0 \\ \hline 1 & 0 & 0 & d \end{array}\right]$$

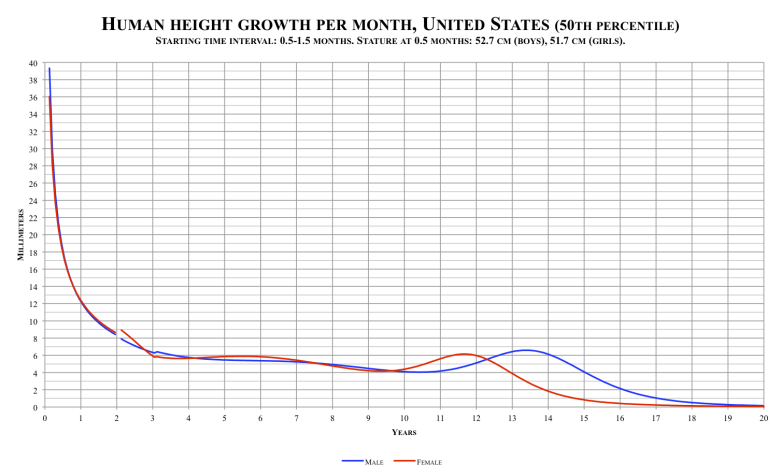
is similar to

$$\Sigma^* = \left[\begin{array}{ccc|c} 0 & 0 & -a_0 & b_0 \\ 1 & 0 & -a_1 & b_1 \\ 0 & 1 & -a_2 & b_2 \\ \hline 0 & 0 & 1 & d \end{array}\right]$$

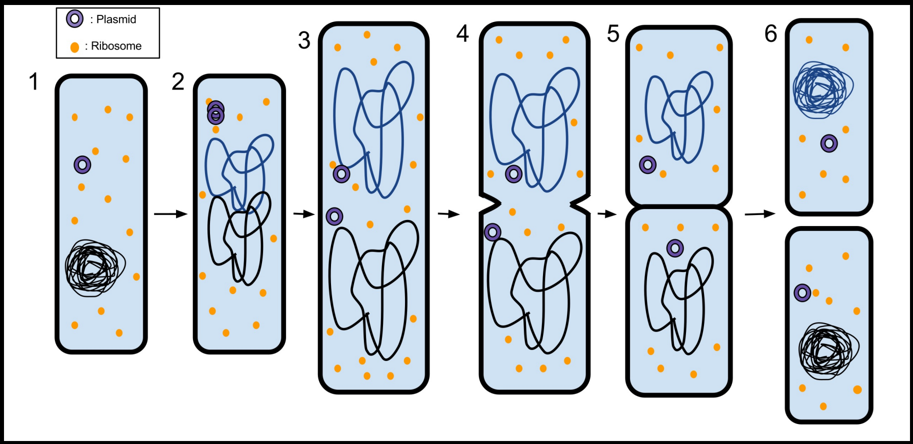
Solution of LTI State-Space Equations

Xu Chen

University of Washington



Population dynamics

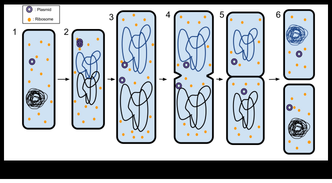


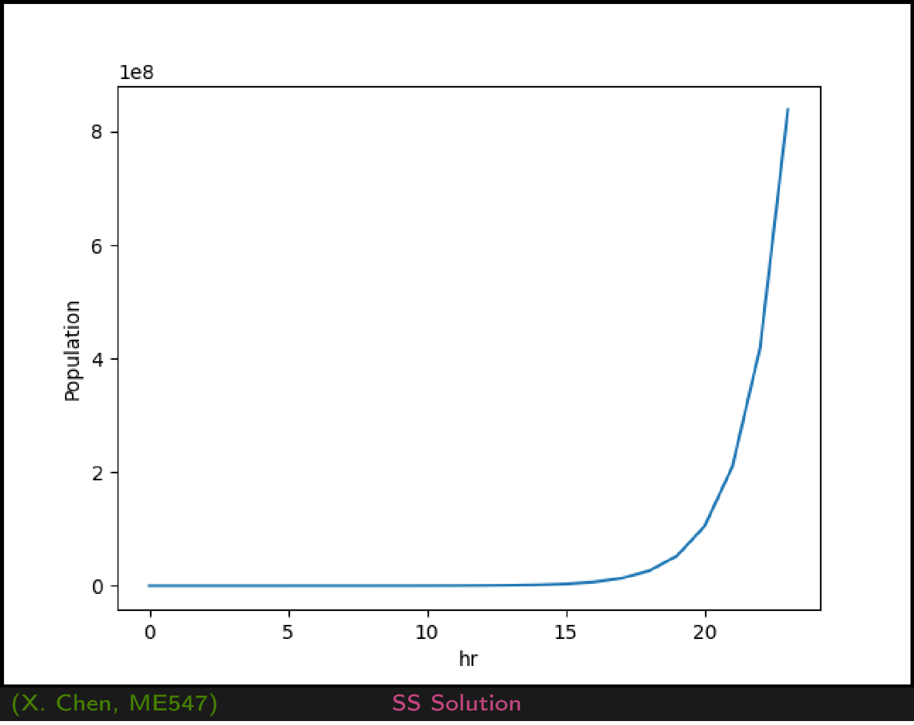
prokaryotic fission

- ~1 hour / division with infinite resource

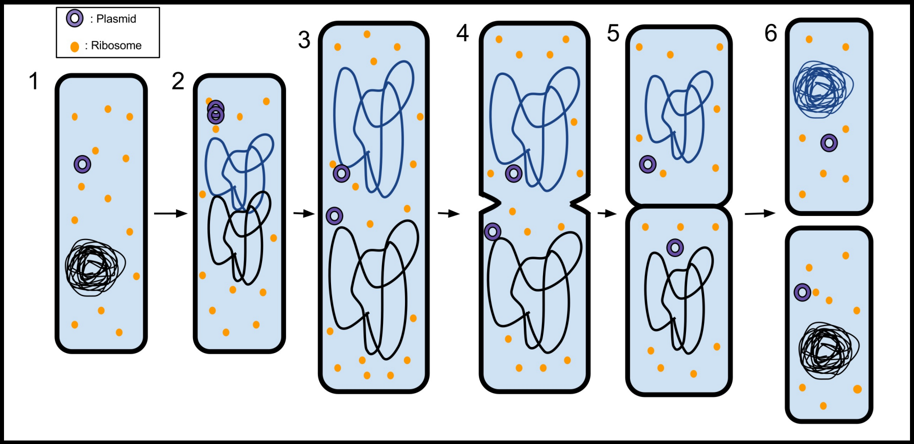
$$100 \xrightarrow{1\text{hr}} 200 \xrightarrow{1\text{hr}} 400 \xrightarrow{1\text{hr}} 800 \xrightarrow{1\text{hr}} \ldots$$

Population dynamics





Population dynamics



prokaryotic fission

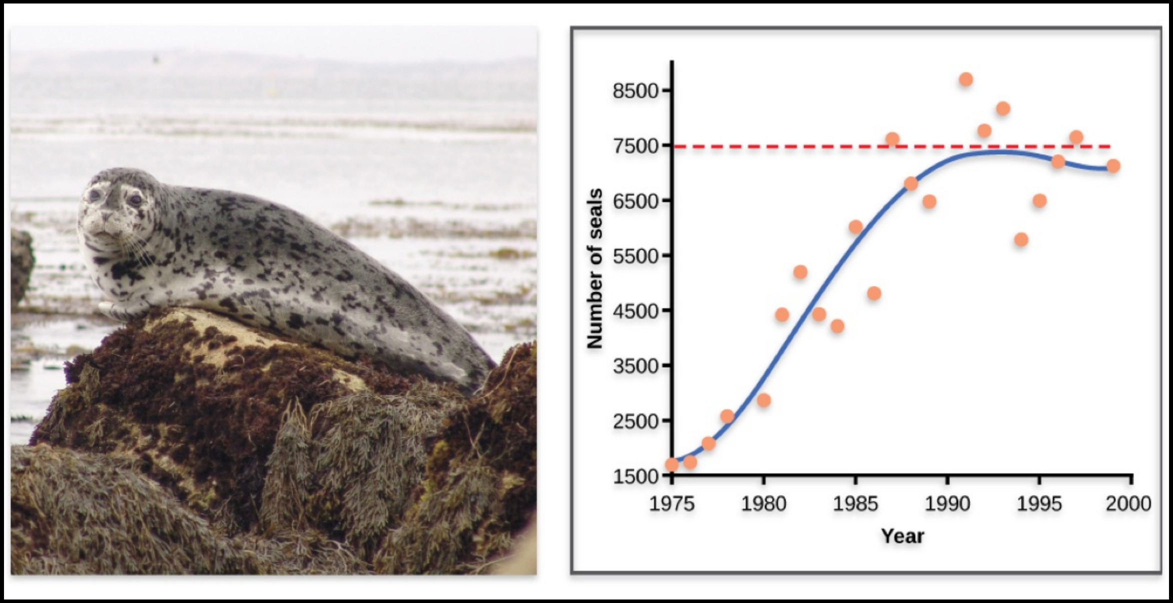
- ~1 hour / division with infinite resource

$$100 \xrightarrow{1\text{hr}} 200 \xrightarrow{1\text{hr}} 400 \xrightarrow{1\text{hr}} 800 \xrightarrow{1\text{hr}} \dots$$

- after 1 day:

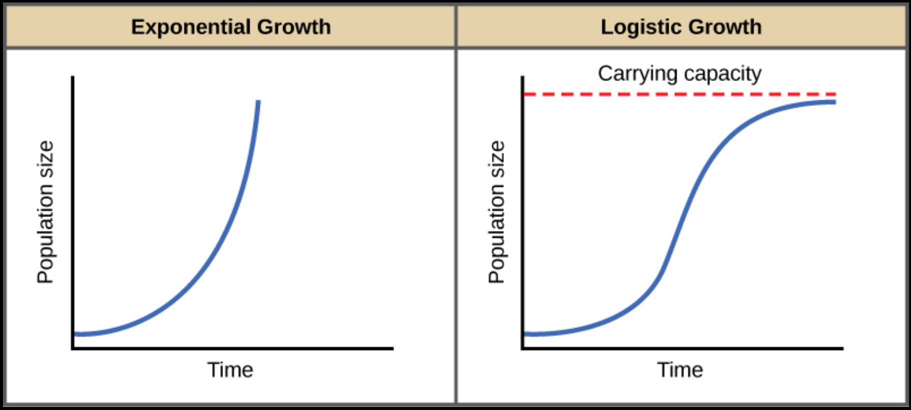
$$100 \xrightarrow[\frac{\Delta N}{N}=1]{1\text{hr}} 200 \xrightarrow{1\text{hr}} 400 \xrightarrow{1\text{hr}} \dots \longrightarrow 100 \times 2^{24} = 1.7\text{B!}$$

Population dynamics



"Environmental limits to population growth: Figure 1," by OpenStax College, Biology, CC BY 4.0.

The exponential function and population dynamics



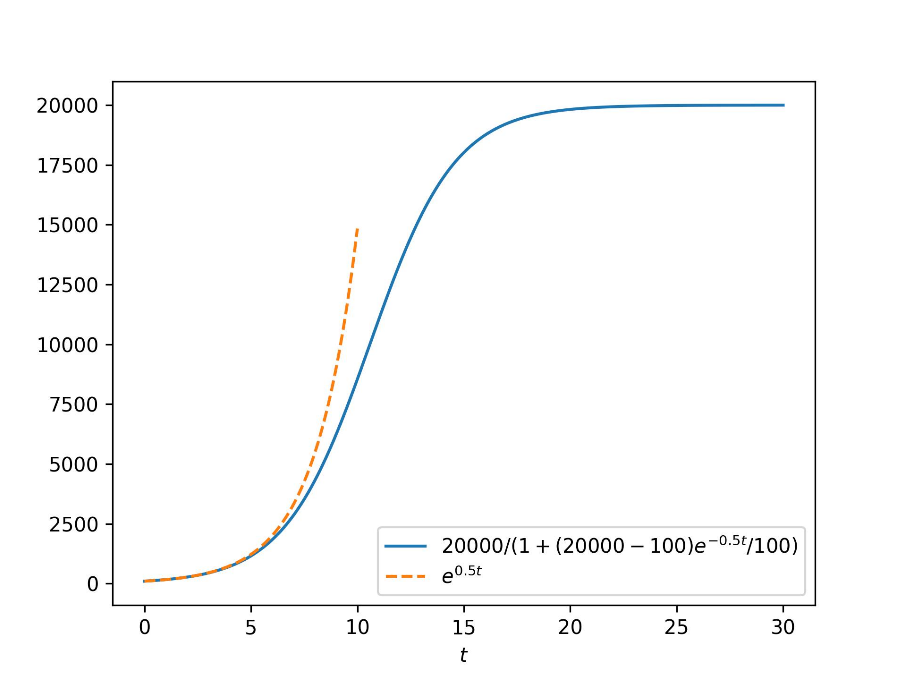
- more general population dynamics (w/ infinite resources)

$$\frac{dN}{dt} = \overbrace{(\text{birth rate} - \text{death rate})}^{r} N \Rightarrow N(t) = e^{rt} N(0)$$

- logistic growth (w/ limited resources in reality)

$$\frac{dN}{dt} = r\frac{K-N}{K}N \Rightarrow N(t) = \frac{KN_0 e^{rt}}{(K-N_0) + N_0 e^{rt}} = \frac{K}{1 + \frac{K-N_0}{N_0} e^{-rt}}$$

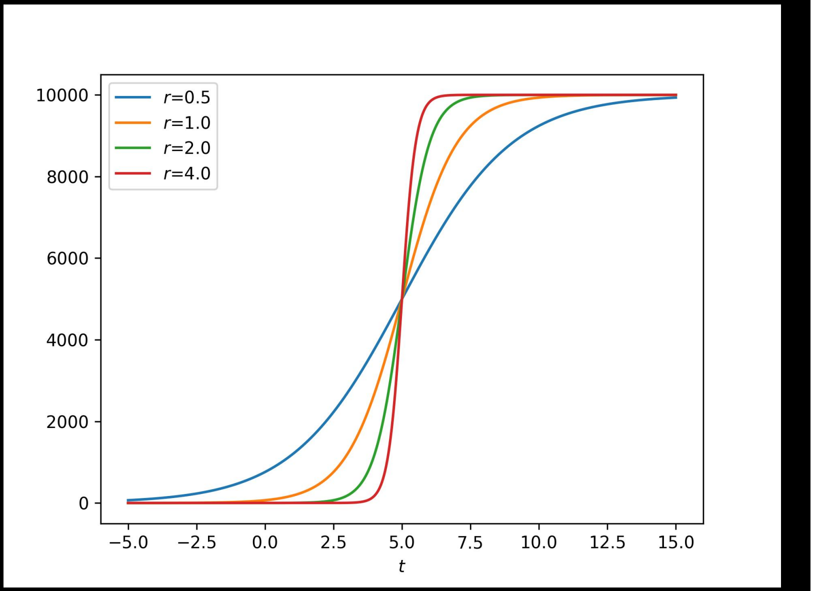
The exponential function and the logistic S curve: example



The logistic S curve

$\frac{K}{1+\frac{K-N_0}{N_0}e^{-rt}}$
can also be written as
$\frac{K}{1+e^{-r(t-t_o)}}$

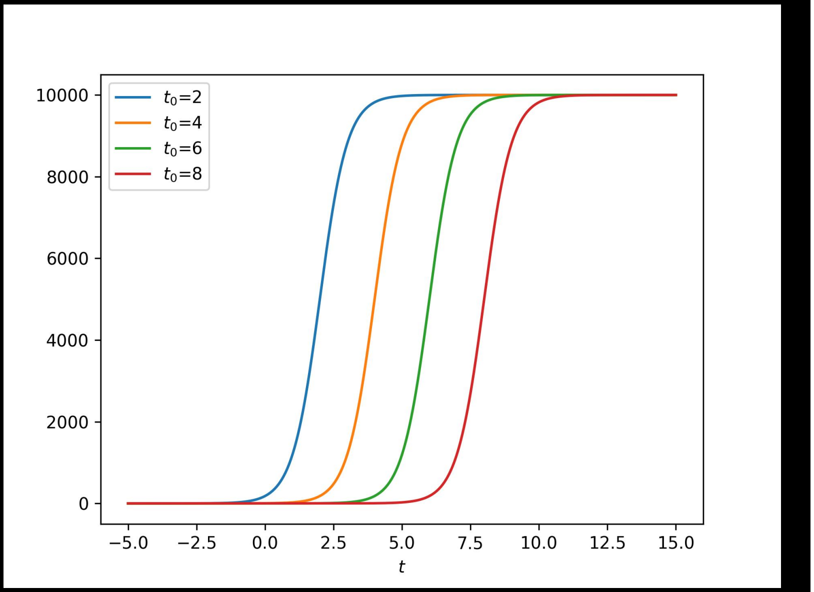
- K: final value
- r: logistic growth rate
- t_o: midpoint



The logistic S curve

$\frac{K}{1+\frac{K-N_0}{N_0}e^{-rt}}$
can also be written as
$\frac{K}{1+e^{-r(t-t_o)}}$

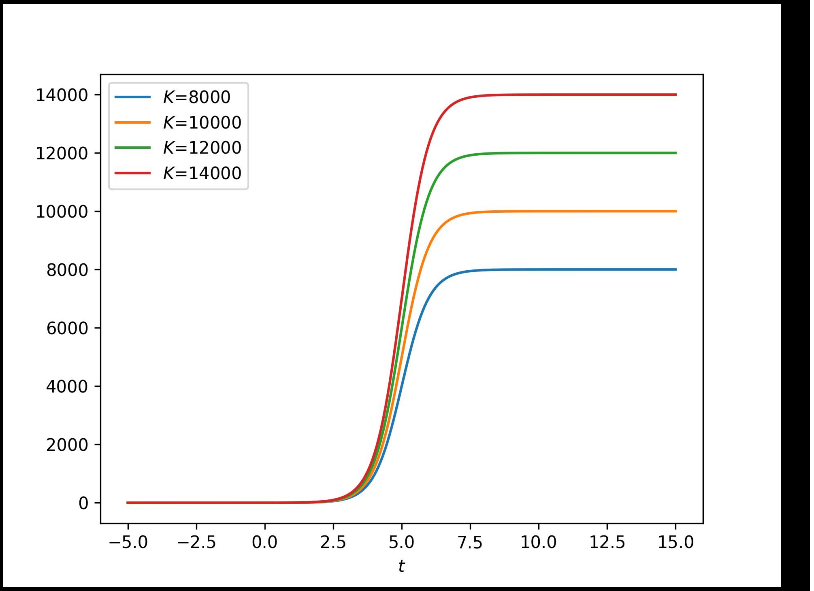
- K: final value
- r: logistic growth rate
- t_o: midpoint



The logistic S curve

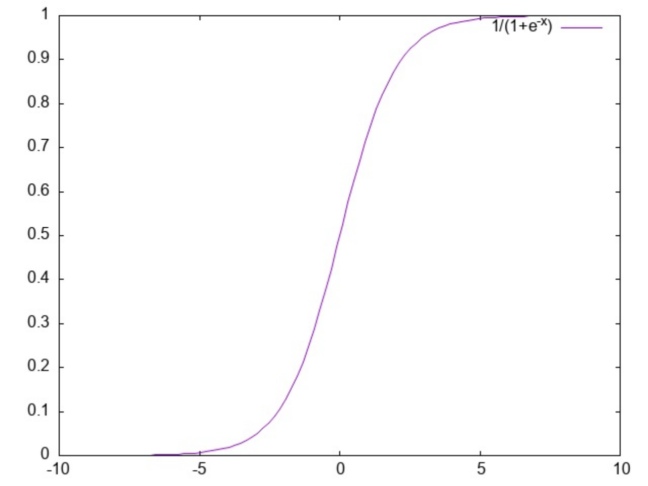
$\frac{K}{1+\frac{K-N_0}{N_0}e^{-rt}}$
can also be written as
$\frac{K}{1+e^{-r(t-t_o)}}$

- K: final value
- r: logistic growth rate
- t_o: midpoint



The logistic function in deep learning

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



- transforms the input variables into a probability value between 0 and 1
- represents the likelihood of the dependent variable being 1 or 0

General LTI continuout-time state equation

$$\frac{dx}{dt} = Ax + Bu$$

$$\Sigma = \left[\begin{array}{c|c} A_{n\times n} & B_{n\times m} \\ \hline C_{n_y\times n} & D_{n_y\times m} \end{array}\right]$$

- to solve the vector equation $\dot{x} = Ax + Bu$, we start with the scalar case when $x, a, b, u \in \mathbb{R}$.

Introduction

The Solution to $\dot{x} = ax + bu$

- fundamental property of exponential functions

$$\frac{d}{dt}e^{at} = ae^{at}, \quad \frac{d}{dt}e^{-at} = -ae^{-at}$$

- $\dot{x}(t) = ax(t) + bu(t),\ a \neq 0 \overset{\because e^{-at}\neq 0}{\Longrightarrow} e^{-at}\dot{x}\left(t\right) - e^{-at}ax\left(t\right) = e^{-at}bu\left(t\right)$
- namely,

$$\frac{d}{dt}\left\{e^{-at}x\left(t\right)\right\} = e^{-at}bu\left(t\right) \Leftrightarrow d\left\{e^{-at}x\left(t\right)\right\} = e^{-at}bu\left(t\right)dt$$

$$\Longrightarrow \boxed{e^{-at}x\left(t\right) = e^{-at_0}x\left(t_0\right) + \int_{t_0}^{t} e^{-a\tau}bu\left(\tau\right)d\tau}$$

The solution to $\dot{x} = ax + bu$

$$e^{-at}x\left(t\right) = e^{-at_0}x\left(t_0\right) + \int_{t_0}^{t} e^{-a\tau}bu\left(\tau\right)d\tau$$

when $t_0 = 0$, we have

$$\boxed{x\left(t\right) = \underbrace{e^{at}x\left(0\right)}_{\text{free response}} + \underbrace{\int_0^t e^{a\left(t-\tau\right)}bu\left(\tau\right)d\tau}_{\text{forced response}}}$$

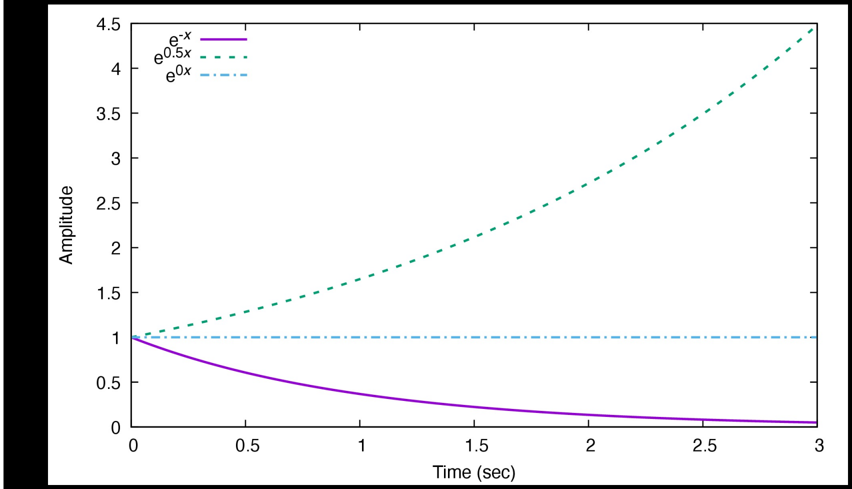
About e

- $e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.71828\ldots$
- also $e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n$
- Python demonstration:

```
import math
math.e
for ii in range(10):
   print(sum(1/math.factorial(k) for k in range(ii)))
for ii in range(1,30):
   print((1+1/ii)**ii)
```

The solution to $\dot{x} = ax + bu$

Solution concepts of $e^{at}x(0)$

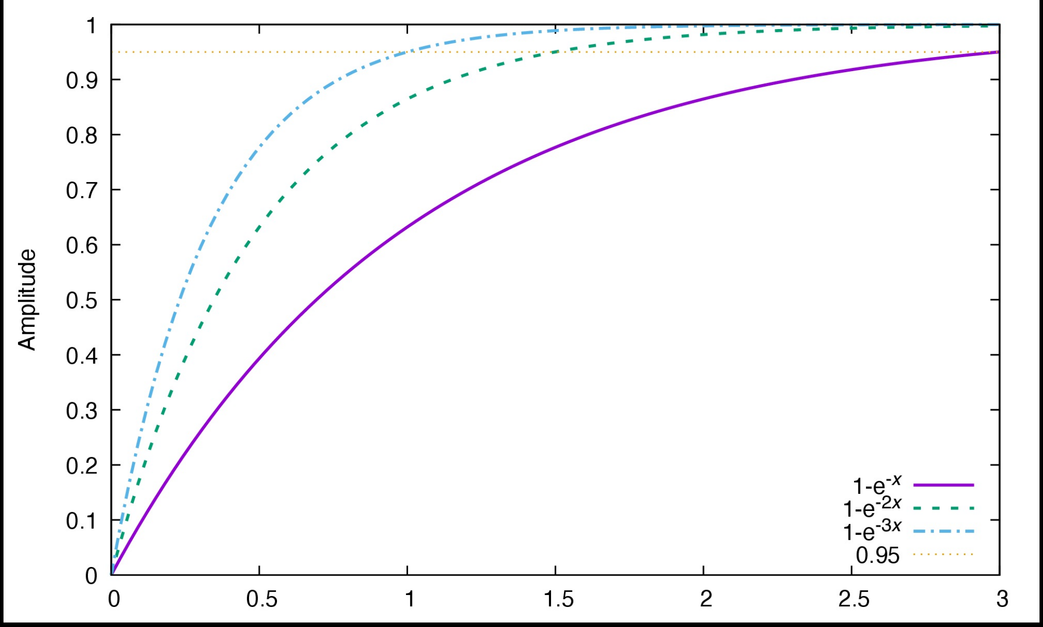


$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.71828\ldots$
$e^{-1} \approx 37\%$,
$e^{-2} \approx 14\%$,
$e^{-3} \approx 5\%$,
$e^{-4} \approx 2\%$
time constant $\tau \triangleq \frac{1}{|a|}$ when $a < 0$: after 3τ, $e^{at}x(0)$, the transient has approximately converged

The solution to $\dot{x} = ax + bu$

Unit step response

When $a < 0$ and $u(t) = 1(t)$ (the step function), the solution is $x(t) = \frac{b}{|a|}(1 - e^{at})$.



* Fundamental Theorem of Differential Equations

addresses the question of whether a dynamical system has a unique solution or not.

Theorem

Consider $\dot{x} = f(x,t)$, $x(t_0) = x_0$, with:

- *$f(x,t)$ piecewise continuous in t (continuous except at finite points of discontinuity)*
- *$f(x,t)$ Lipschitz continuous in x (satisfy the cone constraint:$||f(x,t) - f(y,t)|| \leq k(t)||x - y||$ where $k(t)$ is piecewise continuous)*

then there exists a unique function of time $\phi(\cdot) : \mathbb{R}_+ \to \mathbb{R}^n$ which is continuous almost everywhere and satisfies

- *$\phi(t_0) = x_0$*
- *$\dot{\phi}(t) = f(\phi(t), t)$, $\forall t \in \mathbb{R}_+ \backslash D$, where D is the set of discontinuity points for f as a function of t.*

The solution to n^{th}-order LTI systems

- general state-space equation

$$\Sigma : \begin{cases} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{cases} \qquad x(t_0) = x_0 \in \mathbb{R}^n, \;\; A \in \mathbb{R}^{n\times n}$$

- solution

$$\boxed{x(t) = \underbrace{e^{A(t-t_0)}x_0}_{\text{free response}} + \underbrace{\int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau}_{\text{forced response}}}$$

$$y\left(t\right) = Ce^{A(t-t_0)}x_0 + C\int_{t_0}^{t} e^{A(t-\tau)}Bu(\tau)d\tau + Du\left(t\right)$$

- in both the free and the forced responses, computing e^{At} is key
- $e^{A(t-t_0)}$: called the transition matrix

The state transition matrix e^{At}

scalar case with $a \in \mathbb{R}$: Taylor expansion gives

$$e^{at} = 1 + at + \frac{1}{2}(at)^2 + \cdots + \frac{1}{n!}(at)^n + \ldots$$

the transition scalar $\Phi(t, t_0) = e^{a(t-t_0)}$ satisfies

$$\begin{aligned} \Phi(t,t) &= 1 && \text{(transition to itself)} \\ \Phi(t_3,t_2)\Phi(t_2,t_1) &= \Phi(t_3,t_1) && \text{(consecutive transition)} \\ \Phi(t_2,t_1) &= \Phi^{-1}(t_1,t_2) && \text{(reverse transition)} \end{aligned}$$

The state transition matrix e^{At}

matrix case with $A \in \mathbb{R}^{n\times n}$:

$$\boxed{e^{At} = I_n + At + \frac{1}{2}A^2t^2 + \dots + \frac{1}{n!}A^n t^n + \dots}$$

- as I_n and A^i are matrices of dimension $n \times n$, e^{At} must $\in \mathbb{R}^{n\times n}$
- the transition matrix $\Phi(t, t_0) = e^{A(t-t_0)}$ satisfies

$$\begin{aligned} e^{A0} &= I_n & \Phi(t,t) &= I_n \\ e^{At_1}e^{At_2} &= e^{A(t_1+t_2)} & \Phi(t_3,t_2)\Phi(t_2,t_1) &= \Phi(t_3,t_1) \\ e^{-At} &= \left[e^{At}\right]^{-1} & \Phi(t_2,t_1) &= \Phi^{-1}(t_1,t_2) \end{aligned}$$

- note, however, that $e^{At}e^{Bt} = e^{(A+B)t}$ if and only if $AB = BA$ (check by using Taylor expansion)

Computing a structured e^{At} via Taylor expansion

convenient when A is a diagonal or Jordan matrix

the case with a diagonal matrix $A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$:

- $A^2 = \begin{bmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{bmatrix}, \dots, A^n = \begin{bmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & \lambda_3^n \end{bmatrix}$
- all matrices on the right side of

$$e^{At} = I + At + \frac{1}{2}A^2t^2 + \dots + \frac{1}{n!}A^n t^n + \dots$$

are easy to compute

Computing a structured e^{At} via Taylor expansion

convenient when A is a diagonal or Jordan matrix

the case with a diagonal matrix $A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$:

$$
\begin{aligned}
e^{At} &= I + At + \frac{1}{2}A^2t^2 + \dots + \frac{1}{n!}A^nt^n + \dots \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} \lambda_1 t & 0 & 0 \\ 0 & \lambda_2 t & 0 \\ 0 & 0 & \lambda_3 t \end{bmatrix} + \begin{bmatrix} \frac{1}{2}\lambda_1^2t^2 & 0 & 0 \\ 0 & \frac{1}{2}\lambda_2^2t^2 & 0 \\ 0 & 0 & \frac{1}{2}\lambda_3^2t^2 \end{bmatrix} + \dots \\
&= \begin{bmatrix} 1 + \lambda_1 t + \frac{1}{2}\lambda_1^2t^2 + \dots & 0 & 0 \\ 0 & 1 + \lambda_2 t + \frac{1}{2}\lambda_2^2t^2 + \dots & 0 \\ 0 & 0 & 1 + \lambda_3 t + \frac{1}{2}\lambda_3^2t^2 + \dots \end{bmatrix} \\
&\underline{= \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{bmatrix}.}
\end{aligned}
$$

Computing a structured e^{At} via Taylor expansion

the case with a Jordan matrix $A = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$:

- decompose $A = \underbrace{\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}}_{\lambda I_3} + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{N} \Rightarrow$

  $e^{At} = e^{(\lambda I_3 t + Nt)}$
- also, $(\lambda I_3 t)(Nt) = \lambda N t^2 = (Nt)(\lambda I_3 t)$ and hence $e^{(\lambda I_3 t + Nt)} = e^{\lambda I t}e^{Nt}$
- thus

$$\underline{e^{At} = e^{(\lambda I_3 t + Nt)} = e^{\lambda I t}e^{Nt} \overset{\because e^{\lambda I t} = e^{\lambda t} I}{=\!=\!=} e^{\lambda t}e^{Nt}}$$

Computing a structured e^{At} via Taylor expansion

$$A = \underbrace{\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}}_{\lambda I_3} + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{N}$$

$$e^{At} = e^{\lambda t} e^{Nt}$$

- N is *nilpotent*[1]: $N^3 = N^4 = \dots = 0I_3$, yielding

$$e^{Nt} = I_3 + Nt + \frac{1}{2}N^2t^2 + \cancelto{0}{\frac{1}{3!}N^3t^3} + \cancelto{0}{\dots} = \begin{bmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

- thus

$$e^{At} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} & \frac{t^2}{2}e^{\lambda t} \\ 0 & e^{\lambda t} & te^{\lambda t} \\ 0 & 0 & e^{\lambda t} \end{bmatrix}$$

[1] "nil" $\sim$ zero; "potent" $\sim$ taking powers.

Computing a structured e^{At} via Taylor expansion

Example (mass moving on a straight line with zero friction and no external force)

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$x(t) = e^{At}x(0)$ where

$$e^{At} = I + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t + \frac{1}{2!} \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}} t^2 + \dots \underline{= \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}}.$$

Computing low-order e^{At} via column solutions

an intuition of the matrix entries in e^{At}: consider:

$$\dot{x} = Ax = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x, \quad x(0) = x_0$$

$$x(t) = e^{At}x(0) = \left[\begin{array}{c|c} \overbrace{a_1(t)}^{\text{1st column}} & \overbrace{a_2(t)}^{\text{2nd column}} \end{array} \right] \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} \qquad (1)$$

$$= a_1(t)x_1(0) + a_2(t)x_2(0) \qquad (2)$$

observation

$$x(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow x(t) = a_1(t),$$

$$x(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow x(t) = a_2(t).$$

Computing low-order e^{At} via column solutions

$$\dot{x} = Ax = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x, \quad x(0) = x_0$$

hence, we can obtain e^{At} from:

- write out $\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= -x_2(t) \end{aligned} \Rightarrow \begin{aligned} x_1(t) &= e^{0t}x_1(0) + \int_0^t e^{0(t-\tau)}x_2(\tau)d\tau \\ x_2(t) &= e^{-t}x_2(0) \end{aligned}$
- let $x(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then $\begin{aligned} x_1(t) &\equiv 1 \\ x_2(t) &\equiv 0 \end{aligned}$, namely $x(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- let $x(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then $x_2(t) = e^{-t}$ and $x_1(t) = 1 - e^{-t}$, or more compactly, $x(t) = \begin{bmatrix} 1 - e^{-t} \\ e^{-t} \end{bmatrix}$
- using (1), write out directly $e^{At} = \begin{bmatrix} 1 & 1 - e^{-t} \\ 0 & e^{-t} \end{bmatrix}$

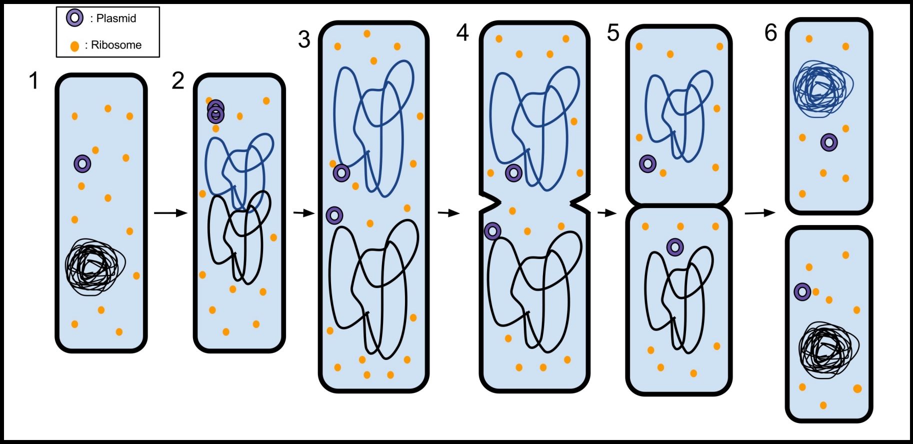
Computing low-order e^{At} via column solutions

Exercise

Compute e^{At} where

$$A=\begin{bmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{bmatrix}$$

Recall: population dynamics



prokaryotic fission

- ~1 hour / division with infinite resource
- after 1 day:

$$100 \xrightarrow[\frac{\Delta N}{N}=1]{1\text{hr}} 200 \xrightarrow{1\text{hr}} 400 \xrightarrow{1\text{hr}} \dots \longrightarrow 100\times 2^{24}=1.7\text{B!}$$

- or: $N(k+1)=2N(k)\Rightarrow N(k)=2^{k}N(0)$

Discrete-time LTI case

discrete-time system:

$$x(k+1) = Ax(k) + Bu(k),\ x(0) = x_0,$$

iteration of the state-space equation gives:

$$x(k) = A^{k-k_0} x(k_o) + \left[A^{k-k_0-1}B, A^{k-k_0-2}B, \cdots, B\right] \begin{bmatrix} u(k_0) \\ u(k_0+1) \\ \vdots \\ u(k-1) \end{bmatrix}$$

$$\Leftrightarrow \boxed{x(k) = \underbrace{A^{k-k_0} x(k_o)}_{\text{free response}} + \underbrace{\sum_{j=k_0}^{k-1} A^{k-1-j} Bu(j)}_{\text{forced response}}}$$

Discrete-time LTI case

$$\boxed{x(k) = \underbrace{A^{k-k_0} x(k_o)}_{\text{free response}} + \underbrace{\sum_{j=k_0}^{k-1} A^{k-1-j} Bu(j)}_{\text{forced response}}}$$

$\Phi(k,j) = A^{k-j}$: the transition matrix:

$$\begin{aligned} \Phi(k,k) &= 1 \\ \Phi(k_3,k_2)\Phi(k_2,k_1) &= \Phi(k_3,k_1) && k_3 \geq k_2 \geq k_1 \\ \Phi(k_2,k_1) &= \Phi^{-1}(k_1,k_2) && \text{if and only if } A \text{ is nonsingular} \end{aligned}$$

The state transition matrix A^k

similar to the continuous-time case, when A is a diagonal or Jordan matrix, A^k is easy

- diagonal matrix $A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$: $A^k = \begin{bmatrix} \lambda_1^k & 0 & 0 \\ 0 & \lambda_2^k & 0 \\ 0 & 0 & \lambda_3^k \end{bmatrix}$

Computing a structured A^k via Taylor expansion

- Jordan canonical form

$$A = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} = \underbrace{\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}}_{\lambda I_3} + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{N}:$$

$$\begin{aligned} A^k &= (\lambda I_3 + N)^k \\ &= (\lambda I_3)^k + k(\lambda I_3)^{k-1} N + \underbrace{\begin{pmatrix} k \\ 2 \end{pmatrix}}_{\text{2 combination}} (\lambda I_3)^{k-2} N^2 + \underbrace{\begin{pmatrix} k \\ 3 \end{pmatrix} (\lambda I_3)^{k-3} N^3 + \dots}_{N^3 = N^4 = \dots = 0 I_3} \\ &= \begin{bmatrix} \lambda^k & 0 & 0 \\ 0 & \lambda^k & 0 \\ 0 & 0 & \lambda^k \end{bmatrix} + k\lambda^{k-1} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} + \frac{k(k-1)}{2} \lambda^{k-2} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \underline{\begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!} k(k-1)\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{bmatrix}} \end{aligned}$$

Computing a structured A^k via Taylor expansion

Exercise

Recall that $\begin{pmatrix} k \\ 3 \end{pmatrix} = \frac{1}{3!}k\left(k-1\right)\left(k-2\right)$. Show

$$A = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}$$

$$\Rightarrow A^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k\left(k-1\right)\lambda^{k-2} & \frac{1}{3!}k\left(k-1\right)\left(k-2\right)\lambda^{k-3} \\ 0 & \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k\left(k-1\right)\lambda^{k-2} \\ 0 & 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & 0 & \lambda^k \end{bmatrix}$$

1. Introduction

2. State-space solution formula
   - Continuous-time case
     - The solution to $\dot{x} = ax + bu$
     - The solution to n^{th}-order LTI systems
     - The state transition matrix e^{At}
     - Computing e^{At} when A is diagonal or in Jordan form
   - Discrete-time LTI case
     - The state transition matrix A^k
     - Computing A^k when A is diagonal or in Jordan form

3. Explicit computation of the state transition matrix e^{At}

4. Explicit Computation of the State Transition Matrix A^k

5. Transition Matrix via Inverse Transformation

Explicit computation of a general e^{At}

- why another method: general matrices may not be diagonal or Jordan
- approach: transform a general matrix to a diagonal or Jordan form, via similarity transformation

Computing e^{At} via similarity transformation

principle concept:

- given

$$\dot{x}(t) = Ax(t) + Bu(t),\ x(t_0) = x_0 \in \mathbb{R}^n,\ A \in \mathbb{R}^{n\times n}$$

- find a nonsingular $T \in \mathbb{R}^{n\times n}$ such that a coordinate transformation defined by $x(t) = Tx^*(t)$ yields

$$\frac{d}{dt}\left(Tx^*(t)\right) = ATx^*(t) + Bu(t)$$

$$\frac{d}{dt}x^*(t) = \underbrace{T^{-1}AT}_{\triangleq\Lambda\text{: diagonal or Jordan}} x^*(t) + \underbrace{T^{-1}B}_{B^*} u(t)$$

$$x^*(0) = T^{-1}x_0$$

Computing e^{At} via similarity transformation

- when $u(t) = 0$

$$\dot{x}(t) = Ax(t) \overset{x=Tx^*}{\Longrightarrow} \frac{d}{dt}x^*(t) = \underbrace{T^{-1}AT}_{\triangleq\Lambda:\ \text{diagonal or Jordan}} x^*(t)$$

- now $x^*(t)$ can be solved easily: e.g., if $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, then

$$x^*(t) = e^{\Lambda t}x^*(0) = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix}\begin{bmatrix} x_1^*(0) \\ x_2^*(0) \end{bmatrix} = \begin{bmatrix} e^{\lambda_1 t}x_1^*(0) \\ e^{\lambda_2 t}x_2^*(0) \end{bmatrix}.$$

- $x(t) = Tx^*(t)$ then yields

$$x(t) = Te^{\Lambda t}x^*(0) = Te^{\Lambda t}T^{-1}x_0$$

- on the other hand, $x(t) = e^{At}x_0 \Rightarrow$

$$\boxed{e^{At} = Te^{\Lambda t}T^{-1}}$$

Similarity transformation

- existence of solutions: T comes from the theory of eigenvalues and eigenvectors in linear algebra
- if A and $B \in \mathbb{C}^{n\times n}$ are similar: $A = TBT^{-1}$, $T \in \mathbb{C}^{n\times n}$, then
  - their A^n and B^n are also similar: e.g.,

$$A^2 = TBT^{-1}TBT^{-1} = TB^2T^{-1}$$

  - their exponential matrices are also similar

$$e^{At} = Te^{Bt}T^{-1}$$

as

$$\begin{aligned} Te^{Bt}T^{-1} &= T(I_n + Bt + \frac{1}{2}B^2t^2 + \ldots)T^{-1} \\ &= TI_nT^{-1} + TBtT^{-1} + \frac{1}{2}TB^2t^2T^{-1} + \ldots \\ &= I + At + \frac{1}{2}A^2t^2 + \cdots = e^{At} \end{aligned}$$

Similarity transformation

- for $A \in \mathbb{R}^{n\times n}$, an eigenvalue $\lambda \in \mathcal{C}$ of A is the solution to the characteristic equation

$$\boxed{\det\left(A-\lambda I\right)=0} \tag{3}$$

- the corresponding eigenvectors are the nonzero solutions to

$$At=\lambda t \Leftrightarrow \left(A-\lambda I\right)t=0 \tag{4}$$

Similarity transformation

The case with distinct eigenvalues (diagonalization)

recall: when $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues such that

$$\begin{aligned} Ax_1 &= \lambda_1 x_1 \\ &\vdots \\ Ax_n &= \lambda_n x_n \end{aligned}$$

or equivalently

$$A\underbrace{\left[x_1,x_2,\dots,x_n\right]}_{\triangleq T}=\left[x_1,x_2,\dots,x_n\right]\underbrace{\begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \dots & 0 & \lambda_n\end{bmatrix}}_{\Lambda}$$

$\left[x_1,x_2,\dots,x_n\right]$ is square and invertible. Hence

$$A=T\Lambda T^{-1},\ \Lambda=T^{-1}AT$$

Example (Mechanical system with strong damping)

$$\frac{d}{dt}\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \underbrace{\left[\begin{array}{cc} 0 & 1 \\ -2 & -3 \end{array}\right]}_{A}\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right]$$

- find eigenvalues: $\det(A-\lambda I) = \det\left[\begin{array}{cc} -\lambda & 1 \\ -2 & -\lambda-3 \end{array}\right] = (\lambda+2)(\lambda+1) \Rightarrow \lambda_1 = -2, \lambda_2 = -1$
- find associate eigenvectors:
  - $\lambda_1 = -2$: $(A-\lambda_1 I)\, t_1 = 0 \Rightarrow t_1 = \left[\begin{array}{c} 1 \\ -2 \end{array}\right]$
  - $\lambda_1 = -1$: $(A-\lambda_2 I)\, t_2 = 0 \Rightarrow t_2 = \left[\begin{array}{c} 1 \\ -1 \end{array}\right]$
- define T and Λ: $T = \left[\begin{array}{cc} t_1 & t_2 \end{array}\right] = \left[\begin{array}{cc} 1 & 1 \\ -2 & -1 \end{array}\right]$, $\Lambda = \left[\begin{array}{cc} \lambda_1 & 0 \\ 0 & \lambda_2 \end{array}\right] = \left[\begin{array}{cc} -2 & 0 \\ 0 & -1 \end{array}\right]$

Example (Mechanical system with strong damping)

$$\frac{d}{dt}\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right] = \underbrace{\left[\begin{array}{cc} 0 & 1 \\ -2 & -3 \end{array}\right]}_{A}\left[\begin{array}{c} x_1 \\ x_2 \end{array}\right]$$

- $T = \left[\begin{array}{cc} 1 & 1 \\ -2 & -1 \end{array}\right]$, $\Lambda = \left[\begin{array}{cc} -2 & 0 \\ 0 & -1 \end{array}\right]$
- compute $T^{-1} = \left[\begin{array}{cc} 1 & 1 \\ -2 & -1 \end{array}\right]^{-1} = \left[\begin{array}{cc} -1 & -1 \\ 2 & 1 \end{array}\right]$
- compute $e^{At} = Te^{\Lambda t}T^{-1} = T\left[\begin{array}{cc} e^{-2t} & 0 \\ 0 & e^{-1t} \end{array}\right]T^{-1} = \left[\begin{array}{cc} -e^{-2t}+2e^{-t} & -e^{-2t}+e^{-t} \\ 2e^{-2t}-2e^{-t} & 2e^{-2t}-e^{-t} \end{array}\right]$

Similarity transform: diagonalization

Physical interpretations

- diagonalized system:
$$x^*(t) = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} x_1^*(0) \\ x_2^*(0) \end{bmatrix} = \begin{bmatrix} e^{\lambda_1 t} x_1^*(0) \\ e^{\lambda_2 t} x_2^*(0) \end{bmatrix}$$
- $x(t) = Tx^*(t) = e^{\lambda_1 t} x_1^*(0) t_1 + e^{\lambda_2 t} x_2^*(0) t_2$ then decomposes the state trajectory into two modes parallel to the two eigenvectors.

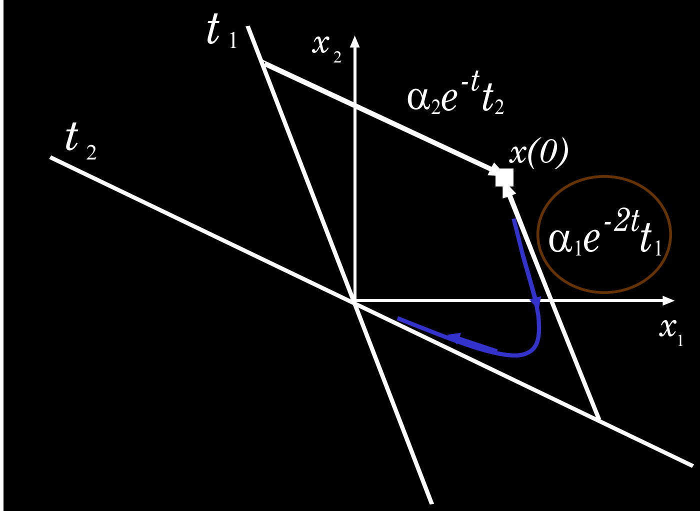
Similarity transform: diagonalization

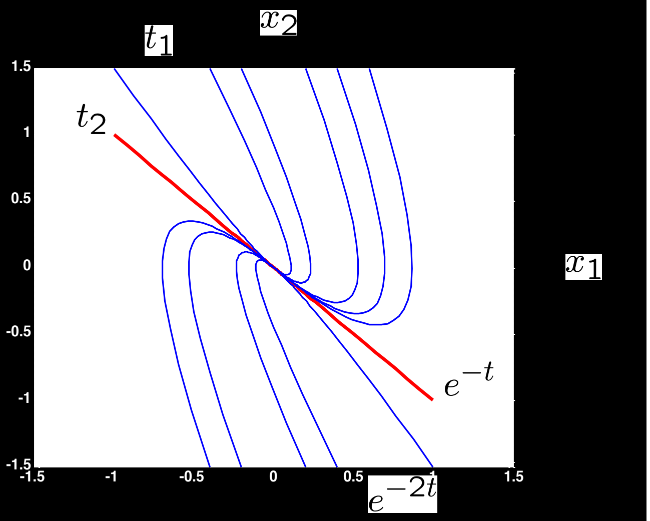
Physical interpretations

- if $x(0)$ is aligned with one eigenvector, say, t_1, then $x_2^*(0) = 0$ and $x(t) = e^{\lambda_1 t} x_1^*(0) t_1 + e^{\lambda_2 t} x_2^*(0) t_2$ dictates that $x(t)$ will stay in the direction of t_1
- i.e., if the state initiates along the direction of one eigenvector, then the free response will stay in that direction without "making turns"
- if $\lambda_1 < 0$, then $x(t)$ will move towards the origin of the state space; if $\lambda_1 = 0$, $x(t)$ will stay at the initial point; and if positive, $x(t)$ will move away from the origin along t_1
- furthermore, the magnitude of λ_1 determines the speed of response

Similarity transform: diagonalization

Physical interpretations: example





Similarity transformation

The case with complex eigenvalues

consider the undamped spring-mass system

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \ \det(A - \lambda I) = \lambda^2 + 1 = 0 \Rightarrow \lambda_{1,2,} = \pm j.$$

the eigenvectors are

$$\lambda_1 = j: \ (A - jI)t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ j \end{bmatrix}$$

$$\lambda_2 = -j: \ (A + jI)t_2 = 0 \Rightarrow t_2 = \begin{bmatrix} 1 \\ -j \end{bmatrix} \text{ (complex conjugate of } t_1\text{).}$$

hence

$$T = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}, \ T^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}$$

Similarity transformation

The case with complex eigenvalues

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- $\lambda_{1,2,} = \pm j$
- $T = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}$, $T^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}$
- we have

$$e^{At} = Te^{\Lambda t}T^{-1} = T\begin{bmatrix} e^{jt} & 0 \\ 0 & e^{-jt} \end{bmatrix}T^{-1} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}.$$

Similarity transformation

The case with complex eigenvalues

for a general $A \in \mathbb{R}^{2\times 2}$ with complex eigenvalues $\sigma \pm j\omega$, by using $T = [t_R, t_I]$, where t_R and t_I are the real and the imaginary parts of t_1, an eigenvector associated with $\lambda_1 = \sigma + j\omega$, $x = Tx^*$ transforms $\dot{x} = Ax$ to

$$\dot{x}^*(t) = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}x^*(t)$$

and

$$e^{\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}t} = \begin{bmatrix} e^{\sigma t}\cos\omega t & e^{\sigma t}\sin\omega t \\ -e^{\sigma t}\sin\omega t & e^{\sigma t}\cos\omega t \end{bmatrix}.$$

Similarity transformation

The case with repeated eigenvalues via generalized eigenvectors

consider $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$: two repeated eigenvalues $\lambda(A) = 1$, and

$$(A - \lambda I)\, t_1 = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

- No other linearly independent eigenvectors exist. What next?
- A is already very similar to the Jordan form. Try instead

$$A \left[\ t_1 \quad t_2\ \right] = \left[\ t_1 \quad t_2\ \right] \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix},$$

which requires $At_2 = t_1 + \lambda t_2$, i.e.,

$$(A - \lambda I)\, t_2 = t_1 \Leftrightarrow \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} t_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow t_2 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$$

t_2 is linearly independent from $t_1 \Rightarrow t_1$ and t_2 span $\mathbb{R}^2$. (t_2 is called a generalized eigenvector.)

Similarity transformation

The case with repeated eigenvalues via generalized eigenvectors

for general 3×3 matrices with $\det(\lambda I - A) = (\lambda - \lambda_m)^3$, i.e., $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_m$, we look for T such that

$$A = TJT^{-1}$$

where J has three canonical forms:

$$i), \begin{bmatrix} \lambda_m & 0 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}, \quad iii), \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

$$ii), \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix} \text{ or } \begin{bmatrix} \lambda_m & 0 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

Similarity transformation

The case with repeated eigenvalues via generalized eigenvectors

$$i),\ A = TJT^{-1},\ J = \begin{bmatrix} \lambda_m & 0 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

this happens

- when A has three linearly independent eigenvectors, i.e., $(A - \lambda_m I)t = 0$ yields t_1, t_2, and t_3 that span $\mathbb{R}^3$.
- mathematically: when nullity $(A - \lambda_m I) = 3$, namely, $\text{rank}(A - \lambda_m I) = 3 - \text{nullity}\,(A - \lambda_m I) = 0$.

Similarity transformation

The case with repeated eigenvalues via generalized eigenvectors

$$ii),\ A = TJT^{-1},\ J = \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix} \text{ or } \begin{bmatrix} \lambda_m & 0 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

- this happens when $(A - \lambda_m I)t = 0$ yields two linearly independent solutions, i.e., when nullity $(A - \lambda_m I) = 2$
- we then have, e.g.,

$$A[t_1, t_2, t_3] = [t_1, t_2, t_3] \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

$$\Leftrightarrow [\lambda_m t_1, t_1 + \lambda_m t_2, \lambda_m t_3] = [At_1, At_2, At_3] \tag{5}$$

- t_1 and t_3 are the directly computed eigenvectors.
- For t_2, the second column of (5) gives

$$(A - \lambda_m I)\, t_2 = t_1$$

Similarity transformation

The case with repeated eigenvalues via generalized eigenvectors

$$iii),\ A = TJT^{-1},\ J = \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

- this is for the case when $(A - \lambda_m I)t = 0$ yields only one linearly independent solution, i.e., when nullity$(A - \lambda_m I) = 1$
- We then have

$$A[t_1, t_2, t_3] = [t_1, t_2, t_3] \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

$$\Leftrightarrow [\lambda_m t_1, t_1 + \lambda_m t_2, t_2 + \lambda_m t_3] = [At_1, At_2, At_3]$$

yielding

$$\begin{aligned} &(A - \lambda_m I)\, t_1 = 0 \\ &(A - \lambda_m I)\, t_2 = t_1,\ \ (t_2 : \text{ generalized eigenvector}) \\ &(A - \lambda_m I)\, t_3 = t_2,\ \ (t_3 : \text{ generalized eigenvector}) \end{aligned}$$

Example

$$A = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix},\ \det(A - \lambda I) = \lambda^2 \Rightarrow \lambda_1 = \lambda_2 = 0,\ J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

- two repeated eigenvalues with rank$(A - 0I) = 1 \Rightarrow$ only one linearly independent eigenvector:$(A - 0I)\, t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
- generalized eigenvector:$(A - 0I)\, t_2 = t_1 \Rightarrow t_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
- coordinate transform matrix:

$$T = [t_1, t_2] = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},\ T^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

$$e^{At} = Te^{Jt}T^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{0t} & te^{0t} \\ 0 & e^{0t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 - t & t \\ -t & 1 + t \end{bmatrix}$$

Example

$A = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$, $\det(A - \lambda I) = \lambda^2 \Rightarrow \lambda_1 = \lambda_2 = 0.$

observation:

- $\lambda_1 = 0,\ t_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ implies that if $x_1(0) = x_2(0)$ then the response is characterized by $e^{0t} = 1$
- i.e., $x_1(t) = x_1(0) = x_2(0) = x_2(t)$. This makes sense because $\dot{x}_1 = -x_1 + x_2$ from the state equation.

Example (Multiple eigenvectors)

Obtain the eigenvectors of

$$A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix} \left(\lambda_1 = 5,\ \lambda_2 = \lambda_3 = -3\right).$$

Generalized eigenvectors

Physical interpretation.

when $\dot{x} = Ax$, $A = TJT^{-1}$ with $J = \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$, we have

$$x(t) = e^{At}x(0) = T \begin{bmatrix} e^{\lambda_m t} & te^{\lambda_m t} & 0 \\ 0 & e^{\lambda_m t} & 0 \\ 0 & 0 & e^{\lambda_m t} \end{bmatrix} T^{-1}x(0)$$

$$= T \begin{bmatrix} e^{\lambda_m t} & te^{\lambda_m t} & 0 \\ 0 & e^{\lambda_m t} & 0 \\ 0 & 0 & e^{\lambda_m t} \end{bmatrix} \cancelto{I}{T^{-1}T} x^*(0)$$

- if the initial condition is in the direction of t_1, i.e., $x^*(0) = [x_1^*(0), 0, 0]^T$ and $x_1^*(0) \neq 0$, the above equation yields $x(t) = x_1^*(0)t_1e^{\lambda_m t}$

Generalized eigenvectors

Physical interpretation Cont'd.

when $\dot{x} = Ax$, $A = TJT^{-1}$ with $J = \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$, we have

$$x(t) = e^{At}x(0) = T \begin{bmatrix} e^{\lambda_m t} & te^{\lambda_m t} & 0 \\ 0 & e^{\lambda_m t} & 0 \\ 0 & 0 & e^{\lambda_m t} \end{bmatrix} T^{-1}x(0)$$

$$= T \begin{bmatrix} e^{\lambda_m t} & te^{\lambda_m t} & 0 \\ 0 & e^{\lambda_m t} & 0 \\ 0 & 0 & e^{\lambda_m t} \end{bmatrix} \cancelto{I}{T^{-1}T} x^*(0)$$

- if $x(0)$ starts in the direction of t_2, i.e., $x^*(0) = [0, x_2^*(0), 0]^T$, then $x(t) = x_2^*(0)(t_1te^{\lambda_m t} + t_2e^{\lambda_m t})$. In this case, the response does not remain in the direction of t_2 but is confined in the subspace spanned by t_1 and t_2

Example

Obtain eigenvalues of J and e^{Jt} by inspection:

$$J = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 & 0 \\ 0 & 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & 0 & -3 \end{bmatrix}.$$

Explicit computation of A^k

everything in getting the similarity transform applies to the DT case:

$$A^k = T\Lambda^k T^{-1} \text{ or } A^k = TJ^kT^{-1}.$$

J	J^k
$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}$
$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$	$\begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k\left(k-1\right)\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{bmatrix}$
$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda^k & k\lambda^{k-1} & 0 \\ 0 & \lambda^k & 0 \\ 0 & 0 & \lambda_3^k \end{bmatrix}$
$\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}$	$r^k \begin{bmatrix} \cos k\theta & \sin k\theta \\ -\sin k\theta & \cos k\theta \end{bmatrix}$ $r = \sqrt{\sigma^2 + \omega^2}$ $\theta = \tan^{-1}\frac{\omega}{\sigma}$

Example

Write down J^k for $J = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$ and

$$J = \begin{bmatrix} -10 & 1 & 0 & 0 & 0 \\ 0 & -10 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & -100 & 1 \\ 0 & 0 & 0 & -1 & -100 \end{bmatrix}.$$

1. Introduction

2. State-space solution formula

   Continuous-time case

   The solution to $\dot{x} = ax + bu$

   The solution to n^{th}-order LTI systems

   The state transition matrix e^{At}

   Computing e^{At} when A is diagonal or in Jordan form

   Discrete-time LTI case

   The state transition matrix A^k

   Computing A^k when A is diagonal or in Jordan form

3. Explicit computation of the state transition matrix e^{At}

4. Explicit Computation of the State Transition Matrix A^k

5. Transition Matrix via Inverse Transformation

Transition matrix via inverse transformation

	Continuous-time system
state eq.	$\dot{x}(t) = Ax(t) + Bu(t),\ x(0) = x_0$
solution	$x(t) = \underbrace{e^{At}x(0)}_{\text{free response}} + \underbrace{\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau}_{\text{forced response}}$
transition matrix	e^{At}

On the other hand, from Laplace transform:

$$\dot{x}(t) = Ax(t) + Bu(t) \Rightarrow X(s) = \underbrace{(sI - A)^{-1}x(0)}_{\text{free response}} + \underbrace{(sI - A)^{-1}BU(s)}_{\text{forced response}}$$

Comparing $x(t)$ and $X(s)$ gives

$$\boxed{e^{At} = \mathcal{L}^{-1}\left\{(sI - A)^{-1}\right\}} \tag{6}$$

Example

Example

$$A = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}$$

$$\begin{aligned} e^{At} &= \mathcal{L}^{-1}\begin{bmatrix} s-\sigma & -\omega \\ \omega & s-\sigma \end{bmatrix}^{-1} \\ &= \mathcal{L}^{-1}\left\{\frac{1}{(s-\sigma)^2+\omega^2}\begin{bmatrix} s-\sigma & \omega \\ -\omega & s-\sigma \end{bmatrix}\right\} \\ &= e^{\sigma t}\begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \end{aligned}$$

Transition matrix via inverse transformation (DT case)

	Discrete-time system
state eq.	$x(k+1)=Ax(k)+Bu(k),\ x(0)=x_0$
solution	$x(k)=\underbrace{A^k x(0)}_{\text{free response}}+\underbrace{\sum_{j=0}^{(k-1)} A^{(k-1-j)}Bu(j)}_{\text{forced response}}$
transition matrix	transition matrix A^k

On the other hand, from Z transform:

$$X(z)=(zI-A)^{-1}zx(0)+(zI-A)^{-1}BU(s)$$

Hence

$$\boxed{A^k=\mathcal{Z}^{-1}\left\{(zI-A)^{-1}z\right\}} \tag{7}$$

Example

Example

$$A=\begin{bmatrix}\sigma & \omega\\ -\omega & \sigma\end{bmatrix}$$

$$\begin{aligned}
A^k &= \mathcal{Z}^{-1}\left\{z\begin{bmatrix}z-\sigma & -\omega\\ \omega & z-\sigma\end{bmatrix}^{-1}\right\}\\
&= \mathcal{Z}^{-1}\left\{\frac{z}{(z-\sigma)^2+\omega^2}\begin{bmatrix}z-\sigma & \omega\\ -\omega & z-\sigma\end{bmatrix}\right\}\\
&= \mathcal{Z}^{-1}\left\{\frac{z}{z^2-2r\cos\theta z+r^2}\begin{bmatrix}z-r\cos\theta & r\sin\theta\\ -r\sin\theta & z-r\cos\theta\end{bmatrix}\right\}\\
&\qquad ,\ r=\sqrt{\sigma^2+\omega^2},\ \theta=\tan^{-1}\frac{\omega}{\sigma}\\
&= r^k\begin{bmatrix}\cos k\theta & \sin k\theta\\ -\sin k\theta & \cos k\theta\end{bmatrix}
\end{aligned}$$

Example

Consider $A = \begin{bmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{bmatrix}$. We have

$$
\begin{aligned}
& (zI - A)^{-1} z \\
= & \begin{bmatrix} \frac{z(z-0.5)}{(z-0.8)(z-0.4)} & \frac{0.3z}{(z-0.8)(z-0.4)} \\ \frac{0.1z}{(z-0.8)(z-0.4)} & \frac{z(z-0.7)}{(z-0.8)(z-0.4)} \end{bmatrix} \\
= & \begin{bmatrix} \frac{0.75z}{z-0.8} + \frac{0.25z}{z-0.4} & \frac{0.75z}{z-0.8} - \frac{0.75z}{z-0.4} \\ \frac{0.25z}{z-0.8} - \frac{0.25z}{z-0.4} & \frac{0.25z}{z-0.8} + \frac{0.75z}{z-0.4} \end{bmatrix} \\
\Rightarrow A^k = & \begin{bmatrix} 0.75\left(0.8\right)^k + 0.25\left(0.4\right)^k & 0.75\left(0.8\right)^k - 0.75\left(0.4\right)^k \\ 0.25\left(0.8\right)^k - 0.25\left(0.4\right)^k & 0.25\left(0.8\right)^k + 0.75\left(0.4\right)^k \end{bmatrix}
\end{aligned}
$$

1 Solution of Time-Invariant State-Space Equations

1.1 Continuous-Time State-Space Solutions

1.1.1 The Solution to $\dot{x} = ax + bu$

To solve the vector equation $\dot{x} = Ax + Bu$, we start with the scalar case when $x, a, b, u \in \mathbb{R}$. The solution can be easily derived using one fundamental property of exponential functions, that

$$\frac{d}{dt}e^{at} = ae^{at},$$

and

$$\frac{d}{dt}e^{-at} = -ae^{-at}.$$

Consider the ODE

$$\dot{x}(t) = ax(t) + bu(t),\ a \neq 0.$$

Since $e^{-at} \neq 0$, the above is equivalent to

$$e^{-at}\dot{x}\left(t\right) - e^{-at}ax\left(t\right) = e^{-at}bu\left(t\right),$$

namely,

$$\begin{gathered}\frac{d}{dt}\left\{e^{-at}x\left(t\right)\right\} = e^{-at}bu\left(t\right),\\ \Leftrightarrow d\left\{e^{-at}x\left(t\right)\right\} = e^{-at}bu\left(t\right)dt.\end{gathered}$$

Integrating both sides from t_0 to t_1 gives

$$e^{-at_1}x\left(t_1\right) = e^{-at_0}x\left(t_0\right) + \int_{t_0}^{t_1} e^{-at}bu\left(t\right)dt.$$

It does not matter whether we use t or τ in the integration $\int_{t_0}^{t_1} e^{-at}bu\left(t\right)dt$. Hence we can change notations and get

$$\begin{gathered}e^{-at}x\left(t\right) = e^{-at_0}x\left(t_0\right) + \int_{t_0}^{t} e^{-a\tau}bu\left(\tau\right)d\tau,\\ \Leftrightarrow x\left(t\right) = e^{a(t-t_o)}x\left(t_0\right) + \int_{t_0}^{t} e^{a(t-\tau)}bu\left(\tau\right)d\tau.\end{gathered}$$

Taking $t_0 = 0$ gives

$$\boxed{x\left(t\right) = \underbrace{e^{at}x\left(0\right)}_{\text{free response}} + \underbrace{\int_0^t e^{a(t-\tau)}bu\left(\tau\right)d\tau}_{\text{forced response}}} \tag{1}$$

where the free response is the part of the solution due only to initial conditions when no input is applied, and the forced response is the part due to the input alone.

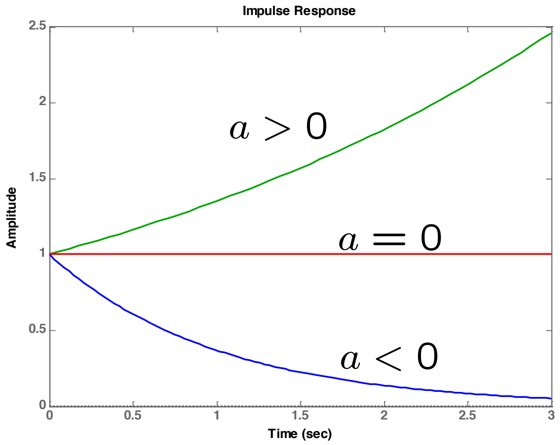
Solution Concepts.

Time Constant. When $a < 0$, e^{at} is a decaying function. For the free response $e^{at}x\left(0\right)$, the exponential function satisfies $e^{-1} \approx 37\%$, $e^{-2} \approx 14\%$, $e^{-3} \approx 5\%$, and $e^{-4} \approx 2\%$. The time constant is defined as

$$T = \frac{1}{|a|}.$$

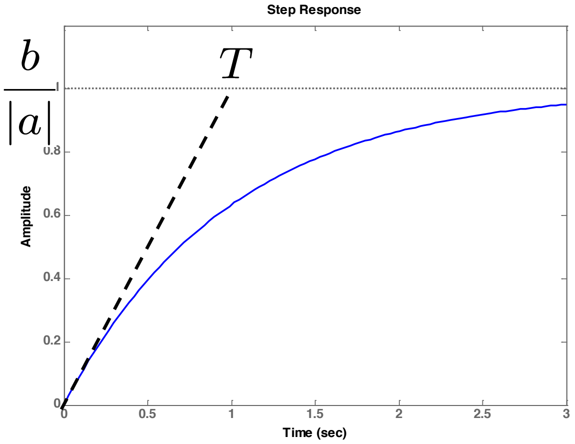
After three time constants, the free response reduces to 5% of its initial value. Roughly, we say the free response has died down.

Graphically, the exponential function looks like:



Unit Step Response. When $a < 0$ and $u(t) = 1(t)$ (the step function), the solution is

$$x(t) = \frac{b}{|a|}(1 - e^{at}).$$



1.1.2 * Fundamental Theorem of Differential Equations

The following theorem addresses the question of whether a dynamical system has a unique solution or not.

Theorem 1. *Consider $\dot{x} = f(x, t)$, $x(t_0) = x_0$, with:*

- $f(x, t)$ piecewise continuous in t
- $f(x, t)$ Lipschitz continuous in x

then there exists a *unique* function of time $\phi(\cdot) : \mathbb{R}_+ \to \mathbb{R}^n$ which is continuous almost everywhere and satisfies

- $\phi(t_0) = x_0$
- $\dot{\phi}(t) = f(\phi(t), t)$, $\forall t \in \mathbb{R}_+ \backslash D$, where D is the set of discontinuity points for f as a function of t.

Remark 1. Piecewise continuous functions are continuous except at finite points of discontinuity.

- example 1: $f(t) = |t|$
- example 2:

$$f(x, t) = \begin{cases} A_1 x, & t \le t_1 \\ A_2 x, & t > t_1 \end{cases}$$

Lipschitz continuous functions are those that satisfy the cone constraint:

$$\|f(x,t)-f(y,t)\| \le k(t)\|x-y\|$$

where $k(t)$ is piecewise continuous.

- example: $f(x)=Ax+B$
- a graphical representation of a Lipschitz function is that it must stay within a cone in the space of $(x,f(x))$
- a function is Lipschitz continuous if it is continuously differentiable with its derivative bounded everywhere. This is a sufficient condition. Functions can be Lipschitz continuous but not differentiable: e.g., the saturation function and $f(x)=|x|$.
- A continuous function is not necessarily Lipschitz continuous at all: e.g., a function whose derivative at $x=0$ is infinity.

1.1.3 The Solution to n^{th}-order LTI System

Consider the general state-space equation

$$\Sigma : \begin{cases} \dot{x}(t) &= Ax(t)+Bu(t) \\ y(t) &= Cx(t)+Du(t) \end{cases} \qquad x(t_0)=x_0\in\mathbb{R}^n,\ A\in\mathbb{R}^{n\times n}$$

Only the first equation here is a differential equation. Once we solve this equation for $x(t)$, we can find $y(t)$ very easily using the second equation. Also, $f(x,t)=Ax+Bu$ satisfies the conditions in Fundamental Theorem for Differential Equations. A unique solution thus exists. The solution of the state-space equations is given in closed form by

$$\boxed{x(t)=\underbrace{e^{A(t-t_0)}x_0}_{\text{free response}}+\underbrace{\int_{t_0}^{t}e^{A(t-\tau)}Bu(\tau)d\tau}_{\text{forced response}}} \tag{2}$$

Derivation of the general state-space solution. Since $e^{-At}\neq 0$, $\dot{x}(t)=Ax(t)+Bu(t)$ is equivalent to

$$e^{-At}\dot{x}(t)-e^{-At}Ax(t)=e^{-At}Bu(t)$$

namely

$$\frac{d}{dt}\left(e^{-At}x(t)\right)=e^{-At}Bu(t)$$
$$\Leftrightarrow d\left(e^{-At}x(t)\right)=e^{-At}Bu(t)\,dt$$

Integrating both sides from t_0 to t_1 gives

$$e^{-At_1}x(t_1)=e^{-At_0}x(t_0)+\int_{t_0}^{t_1}e^{-At}Bu(t)\,dt$$

Changing notations from t to τ in the integral yields

$$e^{-At}x(t)=e^{-At_0}x(t_0)+\int_{t_0}^{t}e^{-A\tau}Bu(\tau)\,d\tau$$
$$\Leftrightarrow x(t)=e^{A(t-t_o)}x(t_0)+\int_{t_0}^{t}e^{A(t-\tau)}Bu(\tau)\,d\tau$$

In both the free and the forced responses, computing the matrix e^{At} is key. $e^{A(t-t_0)}$ is called the transition matrix, and can be computed using a few handy results in linear algebra.

1.1.4 The State Transition Matrix e^{At}

For the scalar case with $a \in \mathbb{R}$, Tylor expansion gives

$$e^{at} = 1 + at + \frac{1}{2}(at)^2 + \cdots + \frac{1}{n!}(at)^n + \ldots \tag{3}$$

The transition scalar $\Phi(t, t_0) = e^{a(t-t_0)}$ satisfies

$$\begin{aligned} \Phi(t,t) &= 1 && \text{(transition to itself)} \\ \Phi(t_3,t_2)\Phi(t_2,t_1) &= \Phi(t_3,t_1) && \text{(consecutive transition)} \\ \Phi(t_2,t_1) &= \Phi^{-1}(t_1,t_2) && \text{(reverse transition)} \end{aligned}$$

For the matrix case with $A \in \mathbb{R}^{n\times n}$

$$\boxed{e^{At} = I + At + \frac{1}{2}A^2t^2 + \cdots + \frac{1}{n!}A^nt^n + \ldots} \tag{4}$$

As I and A^i are matrices of dimension $n \times n$, we confirm that $e^{At} \in \mathbb{R}^{n\times n}$.

The state transition matrix $\Phi(t, t_0) = e^{A(t-t_0)}$ satisfies

$$\begin{aligned} e^{A0} &= I_n \\ e^{At_1}e^{At_2} &= e^{A(t_1+t_2)} \\ e^{-At} &= \left[e^{At}\right]^{-1}. \end{aligned}$$

Similar to the scalar case, it can be shown that

$$\begin{aligned} \Phi(t,t) &= I \\ \Phi(t_3,t_2)\Phi(t_2,t_1) &= \Phi(t_3,t_1) \\ \Phi(t_2,t_1) &= \Phi^{-1}(t_1,t_2). \end{aligned}$$

Note, however, that $e^{At}e^{Bt} = e^{(A+B)t}$ if and only if $AB = BA$. (Check by using Tylor expansion.)

When A is a diagonal or Jordan matrix, the Tylor expansion formula readily generates e^{At}:

Diagonal matrix $A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$**.** In this case $A^n = \begin{bmatrix} \lambda_1^n & 0 & 0 \\ 0 & \lambda_2^n & 0 \\ 0 & 0 & \lambda_3^n \end{bmatrix}$ is also diagonal and hence

$$\begin{aligned} e^{At} &= I + At + \frac{1}{2}A^2t^2 + \cdots + \frac{1}{n!}A^nt^n + \ldots \qquad (5) \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} \lambda_1 t & 0 & 0 \\ 0 & \lambda_2 t & 0 \\ 0 & 0 & \lambda_3 t \end{bmatrix} + \begin{bmatrix} \frac{1}{2}\lambda_1^2t^2 & 0 & 0 \\ 0 & \frac{1}{2}\lambda_2^2t^2 & 0 \\ 0 & 0 & \frac{1}{2}\lambda_3^2t^2 \end{bmatrix} + \ldots \qquad (6) \\ &= \begin{bmatrix} 1+\lambda_1 t + \frac{1}{2}\lambda_1^2t^2 + \ldots & 0 & 0 \\ 0 & 1+\lambda_2 t + \frac{1}{2}\lambda_2^2t^2 + \ldots & 0 \\ 0 & 0 & 1+\lambda_3 t + \frac{1}{2}\lambda_3^2t^2 + \ldots \end{bmatrix} \qquad (7) \\ &= \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{bmatrix}. \qquad (8) \end{aligned}$$

Jordan canonical form $A = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$**.** Decompose

$$A = \underbrace{\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}}_{\lambda I_3} + \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}}_{N}.$$

Then

$$e^{At} = e^{(\lambda I_3 t + Nt)}.$$

As $(\lambda I t)(Nt) = \lambda N t^2 = (Nt)(\lambda I t)$, we have $e^{At} = e^{\lambda I t} e^{Nt} = e^{\lambda t} e^{Nt}$. Also, N has the special property of $N^3 = N^4 = \cdots = 0 I_3$, yielding

$$e^{Nt} = I + Nt + \frac{1}{2} N^2 t^2 = \begin{bmatrix} 1 & t & \frac{t^2}{2} \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus

$$e^{At} = \begin{bmatrix} e^{\lambda t} & t e^{\lambda t} & \frac{t^2}{2} e^{\lambda t} \\ 0 & e^{\lambda t} & t e^{\lambda t} \\ 0 & 0 & e^{\lambda t} \end{bmatrix}. \tag{9}$$

Remark 2 (Nilpotent matrices). The matrix

$$N = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

is a *nilpotent* matrix that equals to zero when raised to a positive integral power. (“nil” $\sim$ zero; “potent” $\sim$ taking powers.) When taking powers of N, the off-diagonal 1 elements march to the top right corner and finally vanish.

Example. Consider a mass moving on a straight line with zero friction and no external force. Let x_1 and x_2 be be the position and the velocity of the mass, respectively. The state-space description of the system is

$$\frac{d}{dt} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Then $x(t) = e^{At} x(0)$ and

$$e^{At} = I + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} t + \frac{1}{2!} \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}} t^2 + \ldots = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Columns of the state-transition matrix. We discuss an intuition of the matrix entries in the e^{At} matrix. Consider the system equation

$$\dot{x} = Ax = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x, \quad x(0) = x_0,$$

with the solution

$$x(t) = e^{At} x(0) = \left[\begin{array}{c|c} | & | \\ a_1(t) & a_2(t) \\ | & | \end{array} \right] \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = a_1(t) x_1(0) + a_2(t) x_2(0), \tag{10}$$

where $a_1(t)$ and $a_2(t)$ are columns of e^{At}. They satisfy

$$x(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow x(t) = a_1(t),$$
$$x(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow x(t) = a_2(t).$$

Hence, we can obtain e^{At} from the following, without using explicitly the Tylor expansion,

1. write out $\begin{aligned} \dot{x}_1(t) =& x_2(t) \\ \dot{x}_2(t) =& -x_2(t) \end{aligned} \Rightarrow \begin{aligned} x_1(t) =& e^{0t} x_1(0) + \int_0^t e^{0(t-\tau)} x_2(\tau) d\tau = e^{0t} x_1(0) + \int_0^t e^{-\tau} x_2(0) d\tau \\ x_2(t) =& e^{-t} x_2(0) \end{aligned}$

2. let $x(0)=\left[\begin{array}{c}1\\0\end{array}\right]$, then $\begin{array}{c}x_1(t)\equiv 1\\x_2(t)\equiv 0\end{array}$, namely $x(t)=\left[\begin{array}{c}1\\0\end{array}\right]$

3. let $x(0)=\left[\begin{array}{c}0\\1\end{array}\right]$, then $x_2(t)=e^{-t}$ and $x_1(t)=1-e^{-t}$, or more compactly, $x(t)=\left[\begin{array}{c}1-e^{-t}\\e^{-t}\end{array}\right]$

4. using the property of (10), write out directly

$$e^{At}=\left[\begin{array}{cc}1 & 1-e^{-t}\\0 & e^{-t}\end{array}\right]$$

Exercise. Use the above method to compute e^{At} where

$$A=\left[\begin{array}{ccc}\lambda & 1 & 0\\0 & \lambda & 1\\0 & 0 & \lambda\end{array}\right].$$

1.2 Discrete-Time LTI State-Space Solutions

For the discrete-time system

$$x(k+1)=Ax(k)+Bu(k),\ x(0)=x_0,$$

iteration of the state-space equation gives

$$x\left(k+1\right)=Ax\left(k\right)+Bu\left(k\right) \tag{11}$$

$$\Rightarrow x\left(k\right)=A^{k-k_0}x\left(k_o\right)+\left[\begin{array}{cccc}A^{k-k_0-1}B & A^{k-k_0-2}B & \cdots & B\end{array}\right]\left[\begin{array}{c}u\left(k_0\right)\\u\left(k_0+1\right)\\\vdots\\u\left(k-1\right)\end{array}\right] \tag{12}$$

$$\Leftrightarrow \boxed{x\left(k\right)=\underbrace{A^{k-k_0}x\left(k_o\right)}_{\text{free response}}+\underbrace{\sum_{j=k_0}^{k-1}A^{k-1-j}Bu\left(j\right)}_{\text{forced response}}} \tag{13}$$

where the transition matrix is defined by $\Phi(k,j)=A^{k-j}$ and satisfies

$$\begin{aligned}\Phi(k,k)&=1\\ \Phi(k_3,k_2)\Phi(k_2,k_1)&=\Phi(k_3,k_1) && k_3\geq k_2\geq k_1\\ \Phi(k_2,k_1)&=\Phi^{-1}(t_1,t_2) && \text{if and only if } A \text{ is nonsingular}\end{aligned}$$

1.2.1 The State Transition Matrix A^k

Similar to the continuous-time case, when A is a diagonal or Jordan matrix, the Tylor expansion formula readily generates A^k. We have

- Diagonal matrix $A=\left[\begin{array}{ccc}\lambda_1 & 0 & 0\\0 & \lambda_2 & 0\\0 & 0 & \lambda_3\end{array}\right]$: $A^k=\left[\begin{array}{ccc}\lambda_1^k & 0 & 0\\0 & \lambda_2^k & 0\\0 & 0 & \lambda_3^k\end{array}\right]$
- Jordan canonical form $A=\left[\begin{array}{ccc}\lambda & 1 & 0\\0 & \lambda & 1\\0 & 0 & \lambda\end{array}\right]=\underbrace{\left[\begin{array}{ccc}\lambda & 0 & 0\\0 & \lambda & 0\\0 & 0 & \lambda\end{array}\right]}_{\lambda I_3}+\underbrace{\left[\begin{array}{ccc}0 & 1 & 0\\0 & 0 & 1\\0 & 0 & 0\end{array}\right]}_{N}$: With the nilpotent N and the

commutative property $(\lambda I_3) N = N (\lambda I_3)$, we have

$$
\begin{aligned}
A^k &= (\lambda I_3 + N)^k = (\lambda I_3)^k + k(\lambda I_3)^{k-1} N + \underbrace{\begin{pmatrix} k \\ 2 \end{pmatrix}}_{\text{2 combination}} (\lambda I_3)^{k-2} N^2 + \underbrace{\begin{pmatrix} k \\ 3 \end{pmatrix} (\lambda I_3)^{k-3} N^3 + \dots}_{N^3 = N^4 = \dots = 0 I_3} \\
&= \begin{bmatrix} \lambda^k & 0 & 0 \\ 0 & \lambda^k & 0 \\ 0 & 0 & \lambda^k \end{bmatrix} + k\lambda^{k-1} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} + \frac{k(k-1)}{2} \lambda^{k-2} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
&= \underline{\begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!} k(k-1)\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{bmatrix}}
\end{aligned}
$$

Exercise. Show that

$$
A = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \Rightarrow A^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!} k(k-1)\lambda^{k-2} & \frac{1}{3!} k(k-1)(k-2)\lambda^{k-3} \\ 0 & \lambda^k & k\lambda^{k-1} & \frac{1}{2!} k(k-1)\lambda^{k-2} \\ 0 & 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & 0 & \lambda^k \end{bmatrix}
$$

1.3 Explicit Computation of the State Transition Matrix e^{At}

General matrices may have structures other than the diagonal and Jordan canonical forms. However, via similar transformation, we can readily transform a general matrix to a diagonal or Jordan form under a different choice of state vectors.

Principle Concept.

1. Given
$$\dot{x}(t) = Ax(t) + Bu(t),\ x(t_0) = x_0 \in \mathbb{R}^n,\ A \in \mathbb{R}^{n \times n}$$
we will find a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that a coordinate transformation defined by $x(t) = Tx^*(t)$ yields
$$
\begin{aligned}
\frac{d}{dt}(Tx^*(t)) &= ATx^*(t) + Bu(t) \\
\frac{d}{dt} x^*(t) &= \underline{\underbrace{T^{-1}AT}_{\triangleq \Lambda} x^*(t) + \underbrace{T^{-1}B}_{B^*} u(t),\ x^*(0) = T^{-1}x_0}
\end{aligned}
$$
where Λ is diagonal or in Jordan form.

2. Now $x^*(t)$ can be solved easily, and the free response is $x^*(t) = e^{\Lambda t} x^*(0)$. For example, when $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$, we would readily obtain $x^*(t) = \begin{bmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{bmatrix} \begin{bmatrix} x_1^*(0) \\ x_2^*(0) \end{bmatrix} = \begin{bmatrix} e^{\lambda_1 t} x_1^*(0) \\ e^{\lambda_2 t} x_2^*(0) \end{bmatrix}$.

3. As $x(t) = Tx^*(t)$, the above implies
$$x(t) = Te^{\Lambda t} T^{-1} x_0$$

4. From the original state-space description, $x(t) = e^{At} x_0$. Hence
$$\boxed{e^{At} = Te^{\Lambda t} T^{-1}}$$

Existence of Solutions. The solution of T comes from the theory of eigenvalues and eigenvectors in linear algebra.

More generally If two matrices A, $B \in \mathbb{C}^{n\times n}$ are similar: $A = TBT^{-1},\ T \in \mathbb{C}^{n\times n}$, then

- their A^n and B^n are also similar: e.g., $AA = TBT^{-1}TBT^{-1} = TB^2T^{-1}$
- their exponential matrices are also similar
$$e^{At} = Te^{Bt}T^{-1}$$
as
$$\begin{aligned} Te^{Bt}T^{-1} &= T(I + Bt + \frac{1}{2}B^2t^2 + \dots)T^{-1} = TIT^{-1} + TBtT^{-1} + \frac{1}{2}TB^2t^2T^{-1} + \dots \\ &= I + At + \frac{1}{2}A^2t^2 + \dots = e^{At} \end{aligned}$$

Eigenvalues and Eigenvectors. The principle concept of computing e^{At} in this section relies on the similarity transform $\Lambda = T^{-1}AT$, where Λ is structurally simple: i.e., in diagonal or Jordan form. We already observed the resulting convenience in computing $x^*(t) = e^{\Lambda t}x^*(0) \overset{\text{e.g.}}{=} \begin{bmatrix} e^{\lambda_1 t}x_1^*(0) \\ e^{\lambda_2 t}x_2^*(0) \end{bmatrix}$. Under the coordinate transformation defined by $x(t) = Tx^*(t)$, we then have

$$x(t) = Te^{\Lambda t}x^*(0) \overset{\text{e.g.}}{=} \underbrace{[t_1, t_2]}_{T} \begin{bmatrix} e^{\lambda_1 t}x_1^*(0) \\ e^{\lambda_2 t}x_2^*(0) \end{bmatrix} = e^{\lambda_1 t}x_1^*(0)t_1 + e^{\lambda_2 t}x_2^*(0)t_2$$

in other words, the state trajectory is conveniently decomposed into two modes along the directions defined by t_1 and t_2, the column vectors of T.

In practice, Λ and T are obtained using the tools of eigenvalues and eigenvectors.

For $A \in \mathbb{R}^{n\times n}$, an eigenvalue $\lambda \in \mathcal{C}$ of A is the solution to the characteristic equation

$$\boxed{\det(A - \lambda I) = 0} \tag{14}$$

The corresponding eigenvectors are the nonzero solutions to

$$At = \lambda t \Leftrightarrow (A - \lambda I)\, t = 0 \tag{15}$$

The case with distinct eigenvalues (diagonalization). When $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues such that

$$\begin{aligned} Ax_1 &= \lambda_1 x_1 \\ Ax_2 &= \lambda_2 x_2 \\ &\vdots \\ Ax_n &= \lambda_n x_n \end{aligned}$$

we can write the above as

$$A\underbrace{[x_1, x_2, \dots, x_n]}_{\triangleq T} = [\lambda_1 x_1, \lambda_2 x_2, \dots, \lambda_n x_n] = [x_1, x_2, \dots, x_n]\underbrace{\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}}_{\Lambda}$$

The matrix $[x_1, x_2, \dots, x_n]$ is square. From linear algebra, the eigenvectors are linearly independent and $[x_1, x_2, \dots, x_n]$ is invertible. Hence

$$A = T\Lambda T^{-1},\ \Lambda = T^{-1}AT$$

Example 1. Mechanical system with strong damping

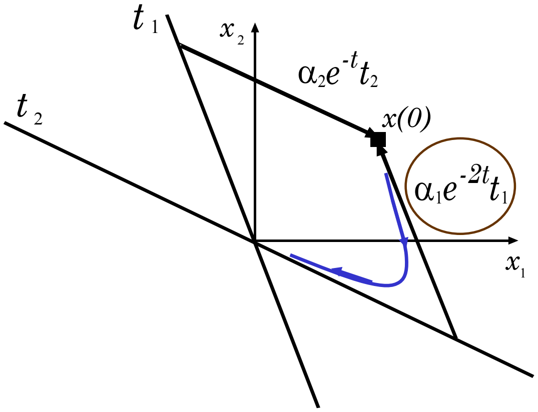
Consider a spring-mass-damper system with $m = 1$, $k = 2$, $b = 3$. Let x_1 and x_2 be the position and velocity of the mass, respectively. We have

$$\begin{cases} \dot{x}_1 & = x_2 \\ \dot{x}_2 + 2x_1 + 3x_2 & = 0 \end{cases} \qquad \Longrightarrow \qquad \frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}}_{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

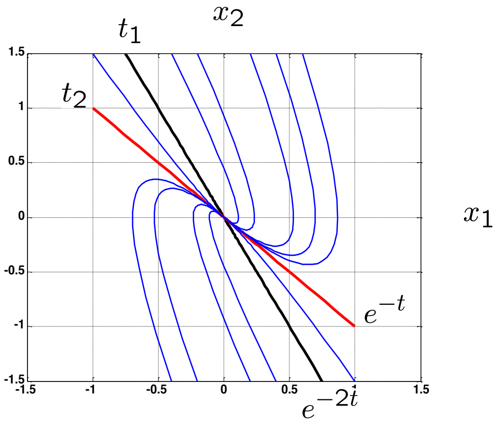
- Find eigenvalues: $\det(A-\lambda I)=\det\begin{bmatrix}-\lambda & 1\\ -2 & -\lambda-3\end{bmatrix}=(\lambda+2)(\lambda+1)\Rightarrow\lambda_1=-2,\lambda_2=-1$
- Find associate eigenvectors:
  - $\lambda_1=-2$: $(A-\lambda_1 I)\,t_1=0\Rightarrow t_1=\begin{bmatrix}1\\ -2\end{bmatrix}$
  - $\lambda_1=-1$: $(A-\lambda_2 I)\,t_2=0\Rightarrow t_2=\begin{bmatrix}1\\ -1\end{bmatrix}$
- Define T and Λ: $T=\begin{bmatrix}t_1 & t_2\end{bmatrix}=\begin{bmatrix}1 & 1\\ -2 & -1\end{bmatrix}$, $\Lambda=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}=\begin{bmatrix}-2 & 0\\ 0 & -1\end{bmatrix}$
- Compute $T^{-1}=\begin{bmatrix}1 & 1\\ -2 & -1\end{bmatrix}^{-1}=\begin{bmatrix}-1 & -1\\ 2 & 1\end{bmatrix}$
- Compute $e^{At}=Te^{\Lambda t}T^{-1}=T\begin{bmatrix}e^{-2t} & 0\\ 0 & e^{-1t}\end{bmatrix}T^{-1}=\begin{bmatrix}-e^{-2t}+2e^{-t} & -e^{-2t}+e^{-t}\\ 2e^{-2t}-2e^{-t} & 2e^{-2t}-e^{-t}\end{bmatrix}$

Physical interpretations Let us revisit the intuition at the beginning of this subsection:

- $x(t)=e^{\lambda_1 t}x_1^*(0)t_1+e^{\lambda_2 t}x_2^*(0)t_2$ decomposes the state trajectory into two modes along the direction of the two eigenvectors t_1 and t_2.
- The two modes are scaled by $x_1^*(0)$ and $x_2^*(0)$ defined from $x(0)=Tx^*(0)$, or more explicitly, $x(0)=[t_1,t_2][x_1^*(0),x_2^*(0)]^T=x_1^*(0)t_1+x_2^*(0)t_2$. This is nothing but decomposing $x(0)$ into the sum of two vectors along the directions of the eigenvectors; and $x_1^*(0)$ and $x_2^*(0)$ are the coefficients of the decomposition!



- If the initial condition $x(0)$ is aligned with one eigenvector, say, t_1, then $x_2^*(0)=0$. The decomposition $x(t)=e^{\lambda_1 t}x_1^*(0)t_1+e^{\lambda_2 t}x_2^*(0)t_2$ then dictates that $x(t)$ will stay in the direction of t_1. In other words, if the state initiates along the direction of one eigenvector, then the free response will stay in that direction without "making turns". If $\lambda_1<0$, then $x(t)$ will move towards the origin of the state space; if $\lambda_1=0$, $x(t)$ will stay at the initial point; and if positive, $x(t)$ will move away from the origin along t_1. Furthermore, the magnitude of λ_1 determines the speed of response.



The case with complex eigenvalues Consider the undamped spring-mass system

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}_{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \ \det(A-\lambda I) = \lambda^2+1=0 \Rightarrow \lambda_{1,2,} = \pm j.$$

The eigenvectors are

$$\lambda_1 = j: \ (A-jI)t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ j \end{bmatrix}$$
$$\lambda_2 = -j: \ (A+jI)t_2 = 0 \Rightarrow t_2 = \begin{bmatrix} 1 \\ -j \end{bmatrix} \text{ (complex conjugate of } t_1\text{).}$$

Hence

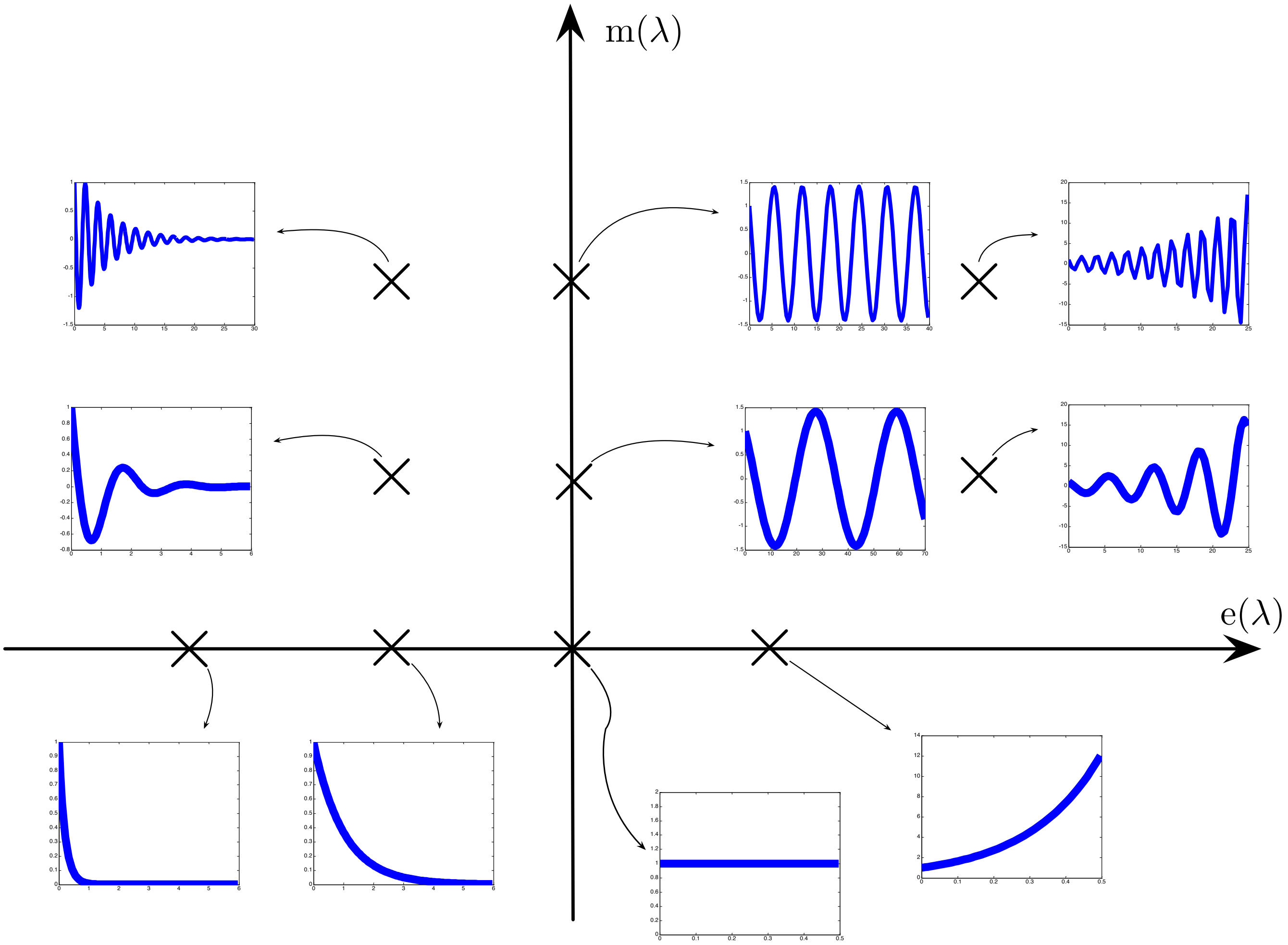
$$T = \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}, \ T^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}, \ e^{At} = Te^{\Lambda t}T^{-1} = T\begin{bmatrix} e^{jt} & 0 \\ 0 & e^{-jt} \end{bmatrix}T^{-1} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}.$$

As an exercise, for a general $A \in \mathbb{R}^{2\times 2}$ with complex eigenvalues $\sigma \pm j\omega$, you can show that by using $T = [t_R, t_I]$ where t_R and t_I are the real and the imaginary parts of t_1, an eigenvector associated with $\lambda_1 = \sigma + j\omega$, $x = Tx^*$ transforms $\dot{x} = Ax$ to

$$\dot{x}^*(t) = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} x^*(t)$$

and

$$e^{\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} t} = \begin{bmatrix} e^{\sigma t}\cos\omega t & e^{\sigma t}\sin\omega t \\ -e^{\sigma t}\sin\omega t & e^{\sigma t}\cos\omega t \end{bmatrix}.$$



The case with repeated eigenvalues, via generalized eigenvectors. Consider $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, which has two repeated eigenvalues $\lambda(A) = 2$ and

$$(A - \lambda I)\, t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

No other linearly independent eigenvectors exist. How do we go further? As A is already very similar to the Jordan form, we try instead

$$A \begin{bmatrix} t_1 & t_2 \end{bmatrix} = \begin{bmatrix} t_1 & t_2 \end{bmatrix} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix},$$

which requires $At_2 = t_1 + \lambda t_2$, i.e.,

$$(A - \lambda I)\, t_2 = t_1 \Leftrightarrow \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} t_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow t_2 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}$$

t_2 is linearly independent from t_1. Together, t_1 and t_2 span the 2-dimensional vector space. As such, t_2 is called a generalized eigenvector.

For general 3×3 matrices with $\det(\lambda I - A) = (\lambda - \lambda_m)^3$, i.e., $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_m$, we look for T such that

$$A = TJT^{-1}$$

where J has three canonical forms:

$$i), \begin{bmatrix} \lambda_m & 0 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}, \; ii), \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix} \text{ or } \begin{bmatrix} \lambda_m & 0 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}, iii), \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix}.$$

- Case i): this happens when A has three linearly independent eigenvectors, i.e., $(A-\lambda_m I)t = 0$ yields t_1, t_2, and t_3 that span the 3-d vector space. This happens when nullity $(A - \lambda_m I) = 3$, namely, $\text{rank}(A-\lambda_m I) = 3 - \text{nullity}\,(A-\lambda_m I) = 0$.

- Case ii): this happens when $(A-\lambda_m I)t = 0$ yields two linearly independent solutions, i.e., when nullity $(A-\lambda_m I) = 2$. We then have, e.g.,

$$A[t_1, t_2, t_3] = [t_1, t_2, t_3] \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix} \Leftrightarrow [\lambda_m t_1, t_1 + \lambda_m t_2, \lambda_m t_3] = [At_1, At_2, At_3]$$

  t_1 and t_3 are the directly computed eigenvectors. For the generalized eigenvector t_2, the second column of the equality gives

$$(A-\lambda_m I)\, t_2 = t_1$$

- Case iii): this is for the case when $(A-\lambda_m I)t = 0$ yields only one linearly independent solution, i.e., when $\text{nullity}(A-\lambda_m I) = 1$. We then have,

$$A[t_1, t_2, t_3] = [t_1, t_2, t_3] \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 1 \\ 0 & 0 & \lambda_m \end{bmatrix} \Leftrightarrow [\lambda_m t_1, t_1 + \lambda_m t_2, t_2 + \lambda_m t_3] = [At_1, At_2, At_3]$$

  yielding

$$\begin{aligned} (A-\lambda_m I)\, t_1 &= 0 \\ (A-\lambda_m I)\, t_2 &= t_1 \\ (A-\lambda_m I)\, t_3 &= t_2 \end{aligned}$$

  where t_2 and t_3 are the generalized eigenvectors.

Example 2. Consider

$$A = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \; \det(A-\lambda I) = (\lambda+1)(\lambda-1) - 1 = \lambda^2 \Rightarrow \lambda_1 = \lambda_2 = 0.$$

Two repeated eigenvalues with $\text{rank}(A - 0I) = 1$ $\Rightarrow$only one linearly independent eigenvector:

$$(A - 0I)\, t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Generalized eigenvector:

$$(A - 0I)\, t_2 = t_1 \Rightarrow t_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Coordinate transform matrix:

$$T = [t_1, t_2] = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \; T^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix},$$

$$J = T^{-1}AT = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \; e^{At} = Te^{Jt}T^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1-t & t \\ -t & 1+t \end{bmatrix}.$$

The first eigenvector implies that if $x_1(0) = x_2(0)$ then the response is characterized by $e^{0t} = 1$, i.e., $x_1(t) = x_1(0) = x_2(0) = x_2(t)$. This makes sense because $\dot{x}_1 = -x_1 + x_2$ from the state equation.

Example 3 (Multiple eigenvectors)**.** Obtain the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}.$$

Analogous procedures give that

$$\lambda_1 = 5,\ \lambda_2 = \lambda_3 = -3.$$

So there are repeated eigenvalues. For $\lambda_1 = 5$, $(A - 5I)t_1 = 0$ gives

$$\begin{bmatrix} -7 & 2 & -3 \\ 2 & -4 & -6 \\ -1 & -2 & -5 \end{bmatrix} t_1 = 0 \Rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

For $\lambda_2 = \lambda_3 = -3$, the characteristic matrix is

$$A + 3I = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{bmatrix}.$$

The second row is the first row multiplied by 2. The third row is the negative of the first row. So the characteristic matrix has only rank 1. The characteristic equation

$$(A - \lambda_2 I)\, t = 0$$

has two linearly independent solutions

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$

Then

$$T = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix},\ J = \begin{bmatrix} 5 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

Physical interpretation. When $\dot{x} = Ax$, $A = TJT^{-1}$ with $J = \begin{bmatrix} \lambda_m & 1 & 0 \\ 0 & \lambda_m & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$, we have

$$x(t) = e^{At}x(0) = T \begin{bmatrix} e^{\lambda_m t} & te^{\lambda_m t} & 0 \\ 0 & e^{\lambda_m t} & 0 \\ 0 & 0 & e^{\lambda_m t} \end{bmatrix} T^{-1}x(0) = T \begin{bmatrix} e^{\lambda_m t} & te^{\lambda_m t} & 0 \\ 0 & e^{\lambda_m t} & 0 \\ 0 & 0 & e^{\lambda_m t} \end{bmatrix} \cancelto{I}{T^{-1}T} x^*(0)$$

If the initial condition is in the direction of t_1, i.e., $x^*(0) = [x_1^*(0), 0, 0]^T$ and $x_1^*(0) \neq 0$, the above equation yields $x(t) = x_1^*(0)t_1 e^{\lambda_m t}$. If $x(0)$ starts in the direction of t_2, i.e., $x^*(0) = [0, x_2^*(0), 0]^T$, then $x(t) = x_2^*(0)(t_1 te^{\lambda_m t} + t_2 e^{\lambda_m t})$. In this case, the response does not remain in the direction of t_2 but is confined in the subspace spanned by t_1 and t_2.

Exercise 1. Obtain eigenvalues of J and e^{Jt} by inspection:

$$J = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 & 0 \\ 0 & 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & 0 & -3 \end{bmatrix}.$$

1.4 Explicit Computation of the State Transition Matrix A^k

Everything in computing the similarity transform $A = T\Lambda T^{-1}$ or $A = TJT^{-1}$ applies to the discrete-time case. The state transition matrix in this case is

$$\boxed{A^k = T\Lambda^k T^{-1} \text{ or } A^k = TJ^kT^{-1}.}$$

You should be able to derive these results:

J	J^k
$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$	$\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}$
$\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$	$\begin{bmatrix} \lambda^k & k\lambda^{k-1} \\ 0 & \lambda^k \end{bmatrix}$
$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$	$\begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k\left(k-1\right)\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{bmatrix}$
$\begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$	$\begin{bmatrix} \lambda^k & k\lambda^{k-1} & 0 \\ 0 & \lambda^k & 0 \\ 0 & 0 & \lambda_3^k \end{bmatrix}$
$\begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}$	$r^k \begin{bmatrix} \cos k\theta & \sin k\theta \\ -\sin k\theta & \cos k\theta \end{bmatrix}$, $r = \sqrt{\sigma^2+\omega^2}$, $\theta = \tan^{-1}\frac{\omega}{\sigma}$

Exercise 2. Write down J^k for $J = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$ and $J = \begin{bmatrix} -10 & 1 & 0 & 0 & 0 \\ 0 & -10 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & -100 & 1 \\ 0 & 0 & 0 & -1 & -100 \end{bmatrix}$.

Exercise 3. Show that

$$J = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \Rightarrow J^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k\left(k-1\right)\lambda^{k-2} & \frac{1}{3!}k\left(k-1\right)\left(k-2\right)\lambda^{k-3} \\ 0 & \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k\left(k-1\right)\lambda^{k-2} \\ 0 & 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & 0 & \lambda^k \end{bmatrix}.$$

1.5 Transition Matrix via Inverse Transformation

We have now

	Continuous-time system	Discrete-time system
state equation	$\dot{x}(t) = Ax(t) + Bu(t)$, $x(0) = x_0$	$x(k+1) = Ax(k) + Bu(k)$, $x(0) = x_0$
solution	$x(t) = \underbrace{e^{At}x(0)}_{\text{free response}} + \underbrace{\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau}_{\text{forced response}}$	$x(k) = \underbrace{A^kx(0)}_{\text{free response}} + \underbrace{\sum_{j=0}^{(k-1)} A^{(k-1-j)}Bu(j)}_{\text{forced response}}$
transition matrix	e^{At}	A^k

We also know from Laplace transform, that

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ X(s) &= \underbrace{\left(sI-A\right)^{-1}x(0)}_{\text{free response}} + \underbrace{\left(sI-A\right)^{-1}BU(s)}_{\text{free response}} \end{aligned}$$

Comparing $x(t)$ and $X(s)$ gives

$$\boxed{e^{At} = \mathcal{L}^{-1}\left\{\left(sI-A\right)^{-1}\right\}} \tag{16}$$

Example 4. Consider $A = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}$. We have

$$\begin{aligned} e^{At} &= \mathcal{L}^{-1} \begin{bmatrix} s-\sigma & -\omega \\ \omega & s-\sigma \end{bmatrix}^{-1} = \mathcal{L}^{-1} \left\{ \frac{1}{(s-\sigma)^2 + \omega^2} \begin{bmatrix} s-\sigma & \omega \\ -\omega & s-\sigma \end{bmatrix} \right\} \\ &= e^{\sigma t} \begin{bmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{bmatrix} \end{aligned}$$

Similarly, for the discrete time case, we have $X(z) = (zI - A)^{-1} zx(0) + (zI - A)^{-1} BU(s)$ and

$$\boxed{A^k = \mathcal{Z}^{-1} \left\{ (zI - A)^{-1} z \right\}} \tag{17}$$

Example 5. Consider $A = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix}$. We have

$$\begin{aligned} A^k &= \mathcal{Z}^{-1} \left\{ z \begin{bmatrix} z-\sigma & -\omega \\ \omega & z-\sigma \end{bmatrix}^{-1} \right\} = \mathcal{Z}^{-1} \left\{ \frac{z}{(z-\sigma)^2 + \omega^2} \begin{bmatrix} z-\sigma & \omega \\ -\omega & z-\sigma \end{bmatrix} \right\} \\ &= \mathcal{Z}^{-1} \left\{ \frac{z}{z^2 - 2r\cos\theta z + r^2} \begin{bmatrix} z - r\cos\theta & r\sin\theta \\ -r\sin\theta & z - r\cos\theta \end{bmatrix} \right\}, \ r = \sqrt{\sigma^2 + \omega^2}, \ \theta = \tan^{-1} \frac{\omega}{\sigma} \\ &= r^k \begin{bmatrix} \cos k\theta & \sin k\theta \\ -\sin k\theta & \cos k\theta \end{bmatrix} \end{aligned}$$

Example 6. Consider $A = \begin{bmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{bmatrix}$. We have

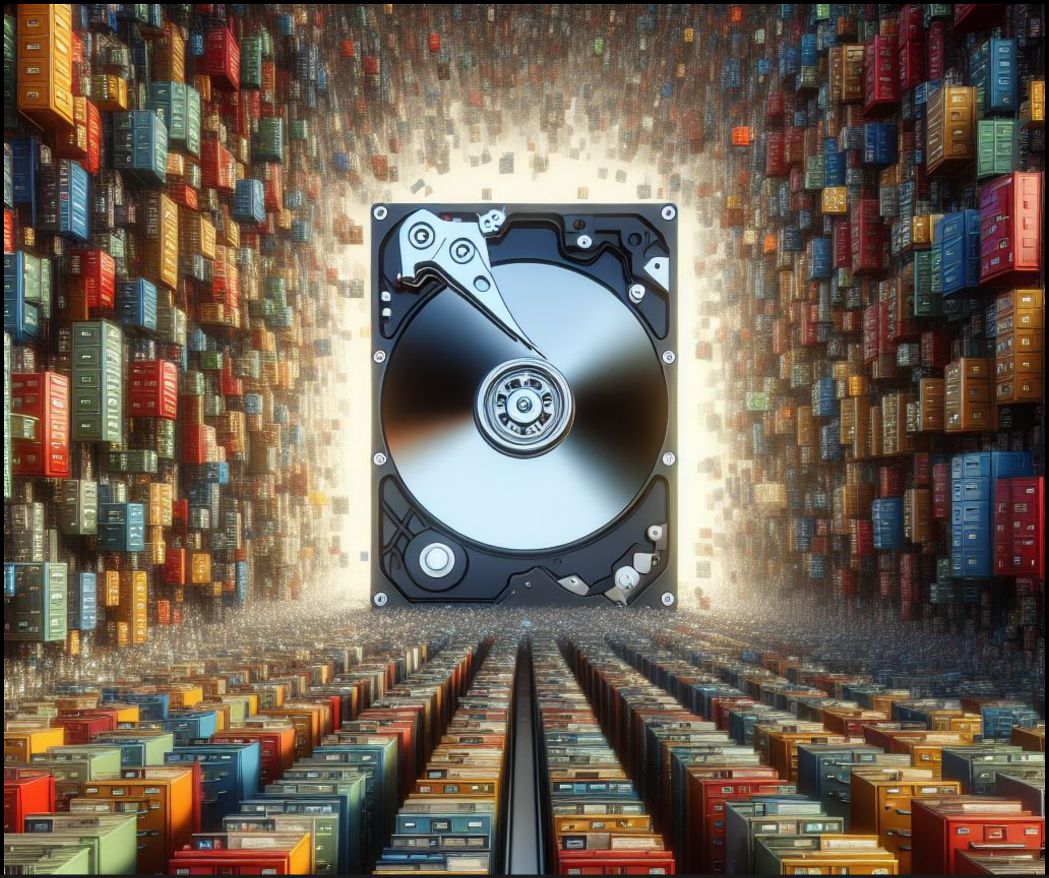
$$\begin{aligned} (zI - A)^{-1} z &= \begin{bmatrix} \frac{z(z-0.5)}{(z-0.8)(z-0.4)} & \frac{0.3z}{(z-0.8)(z-0.4)} \\ \frac{0.1z}{(z-0.8)(z-0.4)} & \frac{z(z-0.7)}{(z-0.8)(z-0.4)} \end{bmatrix} = \begin{bmatrix} \frac{0.75z}{z-0.8} + \frac{0.25z}{z-0.4} & \frac{0.75z}{z-0.8} - \frac{0.75z}{z-0.4} \\ \frac{0.25z}{z-0.8} - \frac{0.25z}{z-0.4} & \frac{0.25z}{z-0.8} + \frac{0.75z}{z-0.4} \end{bmatrix} \\ \Rightarrow A^k &= \begin{bmatrix} 0.75(0.8)^k + 0.25(0.4)^k & 0.75(0.8)^k - 0.75(0.4)^k \\ 0.25(0.8)^k - 0.25(0.4)^k & 0.25(0.8)^k + 0.75(0.4)^k \end{bmatrix} \end{aligned}$$

Discretization of State-Space System Models

Xu Chen

University of Washington

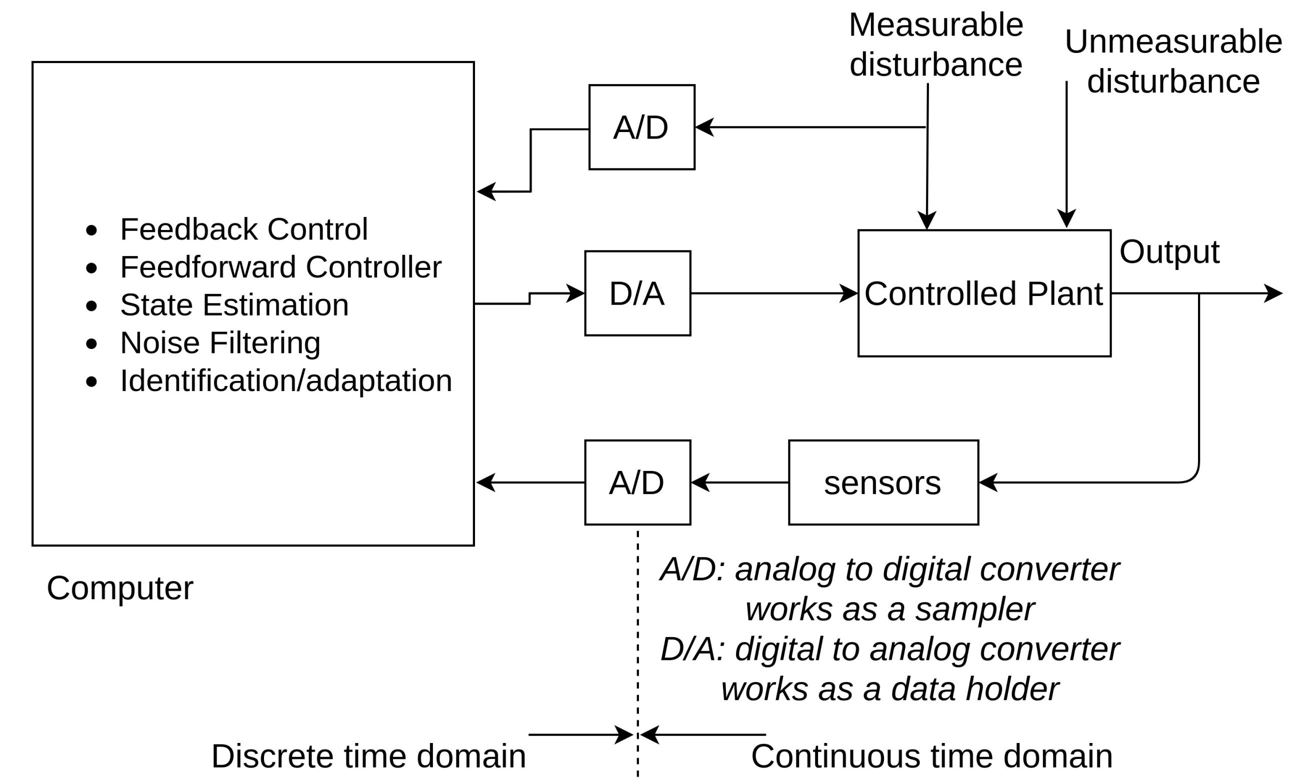
1TB vs 1,300 filing cabinets of paper



Inherent sampling in practice

$$\Delta t = \frac{1}{(\mathrm{rpm}/60) \times \text{sector number}}$$

Practical control systems



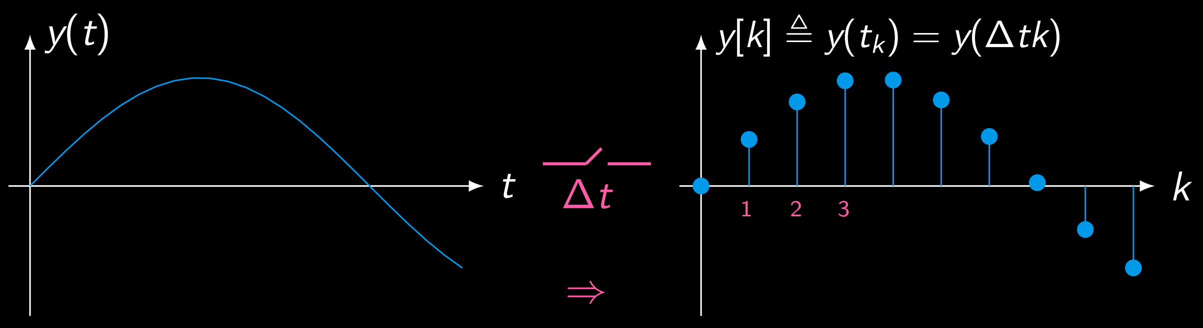
A/D: analog to digital converter works as a sampler

D/A: digital to analog converter works as a data holder

Sampler

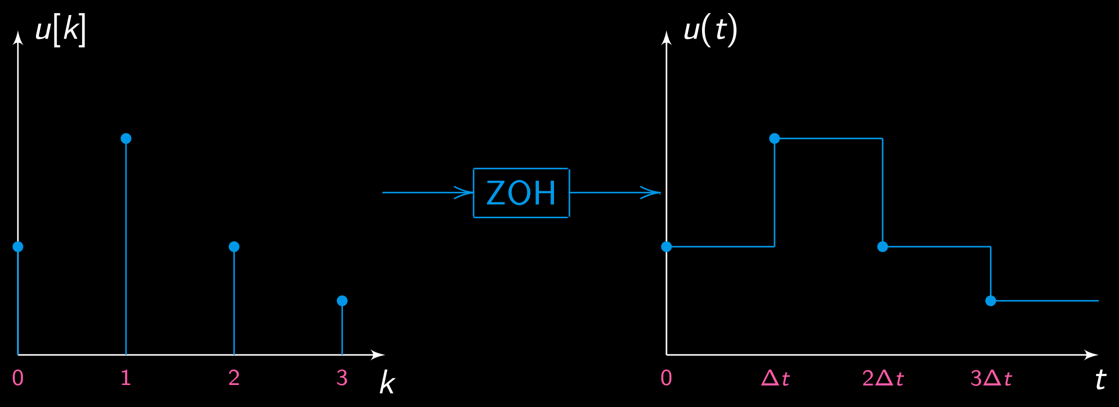
- sampler: converts a time function into a discrete sequence,

e.g.,



Signal holding

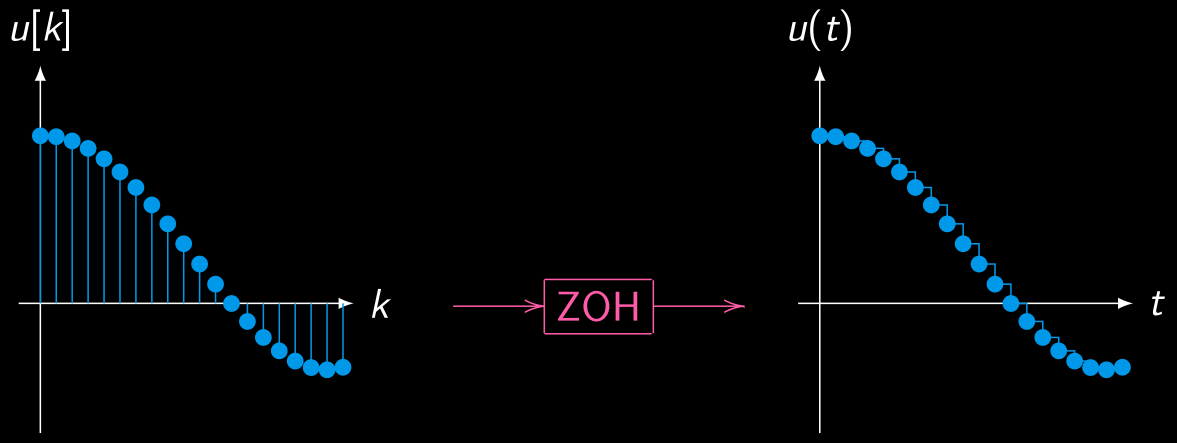
- Zero-order Hold (ZOH): converts a sequence into a “stair-case” time function, e.g.,



- $u(t) = u[k]$ for $t \in [k\Delta t, (k+1)\Delta t)$

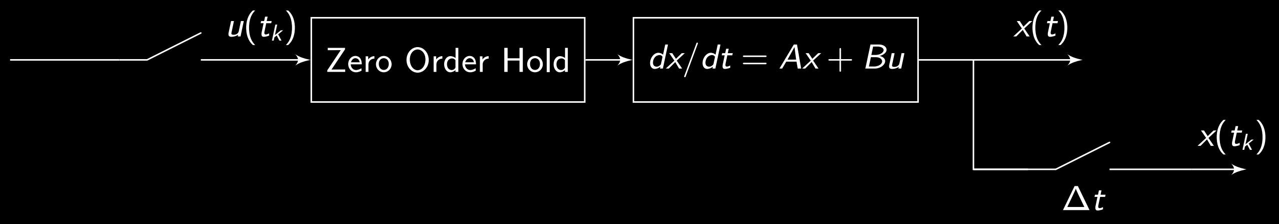
Signal holding

- more faithful presentation with fast sampling



Problem definition

continuous-time system preceded by a ZOH:

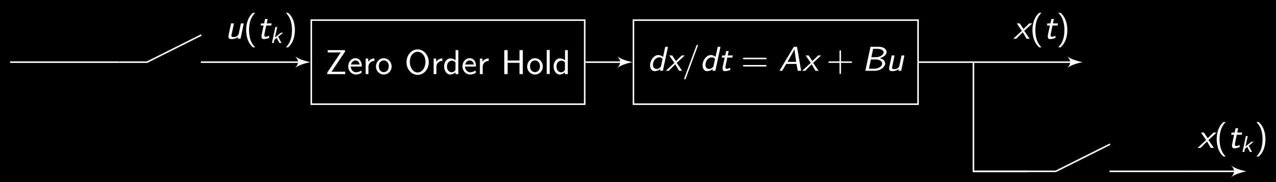


Zero Order Hold

$dx/dt = Ax + Bu$

- $u(t_k)$: discrete-time input
- $x(t)$: continuous-time output
- $x(t_k)$: sampled discrete-time output
- Δt: sampling time
- goal: to obtain the model between $u(t_k)$ and $x(t_k)$

Solution



$u(t_k)$

Zero Order Hold

$dx/dt = Ax + Bu$

$x(t)$

$x(t_k)$

- starting from t_k, the solution of $\dot{x} = Ax + Bu$ at time t_{k+1} is

$$x(t_{k+1}) = e^{A(t_{k+1}-t_k)}x(t_k) + \int_{t_k}^{t_{k+1}} e^{A(t_{k+1}-\tau_o)}Bu(\tau_o)d\tau_o$$

$$= e^{A(\overbrace{t_{k+1}-t_k}^{\Delta t})}x(t_k) + u(t_k)\underbrace{\int_{t_k}^{t_{k+1}} e^{A(\overbrace{t_{k+1}-\tau_o}^{\eta})}Bd\tau_o}_{=\int_{\Delta t}^{0} e^{A\eta}Bd(-\eta)=-\int_{\Delta t}^{0} e^{A\eta}Bd\eta}$$

- noting $-\int_{\Delta t}^{0} e^{A\eta}Bd\eta = \int_0^{\Delta t} e^{A\tau}Bd\tau$ and denoting t_k as k yield

$$\boxed{x[k+1] = A_d x[k] + B_d u[k],\ A_d = e^{A\Delta t},\ B_d = \int_0^{\Delta t} e^{A\tau}Bd\tau}$$

Mapping of eigenvalues

$$\boxed{x[k+1] = A_d x[k] + B_d u[k],\ A_d = e^{A\Delta t},\ B_d = \int_0^{\Delta t} e^{A\tau}Bd\tau}$$

- diagonalization / Jordan form: $A = T^{-1}\Lambda T$
- e^{At} has the same eigenvalues as $e^{\Lambda t}$
- $\Rightarrow$ eigenvalues of $A_d = e^{A\Delta t}$ are $e^{\lambda_i \Delta t}$'s where λ_i is an eigenvalue of A

Example

$$\dot{x}(t) = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{A} x(t) + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{B} u(t)$$

$$y(t) = \underbrace{\begin{bmatrix} \frac{1}{m} & 0 \end{bmatrix}}_{C} x(t)$$

discretization at a sampling time of $\Delta t \Rightarrow$

$$A_d = e^{A\Delta t} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \; B_d = \int_0^{\Delta t} e^{A\tau} B d\tau = \int_0^{\Delta t} \begin{bmatrix} \tau \\ 1 \end{bmatrix} d\tau = \begin{bmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{bmatrix}$$

$$C_d = C$$

Numerical example in Python

```python
import control
import numpy
m = 1
dt = 0.1
A = [[0, 1], [0, 0]]
B = [[0], [1]]
C = [[1/m, 0]]
D = 0

G_s = control.ss(A, B, C, D)
G_z = control.c2d(G_s, dt, 'zoh')
print(G_z.A)

# eigenvalues of continuous-time system
eigA, eigvecA = numpy.linalg.eig(A)
print(eigA)
# eigenvalues of discretized system
eigAd, eigvecAd = numpy.linalg.eig(G_z.A)
print(eigAd)
```

Spectral mapping theorem

- eigenvalues of $A_d = e^{AT}$ are $e^{\lambda_i T}$'s where λ_i is an eigenvalue of A
- more generally: take any $X \in \mathbb{C}^{n\times n}$ and a polynomial function $f(\cdot)$ (more generally, analytic functions)
- e.g.:

$$A = \begin{bmatrix} 99.8 & 2000 \\ -2000 & 99.8 \end{bmatrix} = 99.8I + 2000\overbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}}^{X}$$

- then

$$\text{eig}\,(f(X)) = f(\text{eig}\,(X))$$

- e.g.:

$$A = \begin{bmatrix} 99.8 & 2000 \\ -2000 & 99.8 \end{bmatrix} = 99.8I + 2000\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

$$\lambda(A) = 99.8 + 2000\lambda\left\{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right\} = 99.8 \pm 2000i$$

Spectral mapping theorem

$$A = \begin{bmatrix} 99.8 & 2000 \\ -2000 & 99.8 \end{bmatrix}$$

```
import numpy
A = [[99.8, 2000], [-2000, 99.8]]

eigA, eigvecA = numpy.linalg.eig(A)
print(eigA)
```

```
[99.8+2000.j 99.8-2000.j]
```

Essentials of Control Systems

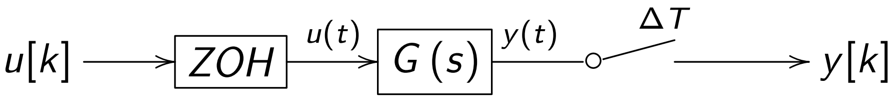
Discretization of Continuous-time Transfer-function Systems

Xu Chen

University of Washington

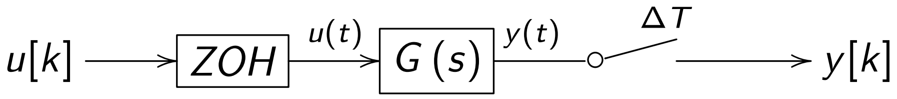
Overview

- Consider the discrete-time controller implementation scheme



  where $u[k]$ and $y[k]$ have the same sampling time.
- for this note, we use $[k]$ to distinguish DT signals from their CT counter parts
- Goal: to derive the transfer function from $u[k]$ to $y[k]$.
- Solution concept: let $u[k]$ be a discrete-time impulse (whose Z transform is 1) and obtain the Z transform of $y[k]$.

Solution



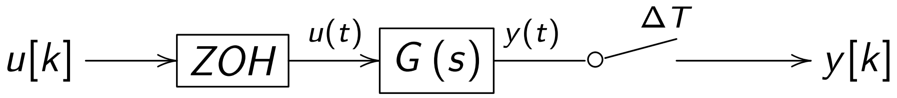
- $u[k]$ is a DT impulse $\Rightarrow$ after ZOH

$$u(t) = \begin{cases} 1, & 0 \le t < \Delta T \\ 0, & \text{otherwise} \end{cases} = 1(t) - 1(t-\Delta T) \Longrightarrow U(s) = \frac{1-e^{-s\Delta T}}{s}$$

- Hence

$$y(t) = \mathcal{L}^{-1}\left[G(s)\frac{1-e^{-s\Delta T}}{s}\right] = \mathcal{L}^{-1}\left[G(s)\frac{1}{s}\right] - \mathcal{L}^{-1}\left[G(s)\frac{e^{-s\Delta T}}{s}\right]$$

Solution



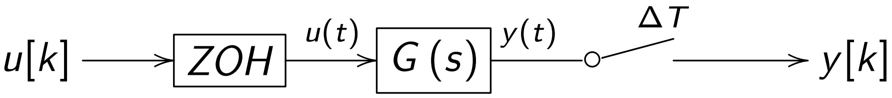
$$y(t) = \mathcal{L}^{-1}\left[G(s)\frac{1-e^{-s\Delta T}}{s}\right] = \mathcal{L}^{-1}\left[G(s)\frac{1}{s}\right] - \mathcal{L}^{-1}\left[G(s)\frac{e^{-s\Delta T}}{s}\right]$$

- Sampling $y(t)$ at ΔT and performing Z transform give:

$$G(z) = \mathcal{Z}\left\{ \underbrace{\overbrace{\mathcal{L}^{-1}\left[G(s)\frac{1}{s}\right]}^{\tilde{y}(t)}\Bigg|_{t=k\Delta T}}_{\triangleq \tilde{y}[k]} - \underbrace{\overbrace{\mathcal{L}^{-1}\left[G(s)\frac{e^{-s\Delta T}}{s}\right]}^{\tilde{y}(t-\Delta T)}\Bigg|_{t=k\Delta T}}_{=\tilde{y}[k-1]!!!} \right\}$$

$$= \mathcal{Z}\left\{\mathcal{L}^{-1}\left[G(s)\frac{1}{s}\right]\bigg|_{t=k\Delta T}\right\} - z^{-1}\mathcal{Z}\left\{\mathcal{L}^{-1}\left[G(s)\frac{1}{s}\right]\bigg|_{t=k\Delta T}\right\}$$

Solution



Fact

The zero order hold equivalent of $G(s)$ is

$$G(z) = (1 - z^{-1})\mathcal{Z}\left\{ \mathcal{L}^{-1}\left[G(s)\frac{1}{s} \right]\Big|_{t=k\Delta T} \right\}$$

where ΔT is the sampling time.

Example

Obtain the ZOH equivalent of

$$G(s) = \frac{a}{s+a}$$

Following the discretization procedures we have $\frac{G(s)}{s} = \frac{a}{s(s+a)} = \frac{1}{s} - \frac{1}{s+a}$ and hence

$$\mathcal{L}^{-1}\left\{ \frac{G(s)}{s} \right\} = 1(t) - e^{-at}1(t)$$

Sampling at ΔT gives $1[k] - e^{-ak\Delta T}1[k]$, whose Z transform is

$$\frac{z}{z-1} - \frac{z}{z - e^{-a\Delta T}} = \frac{z(1 - e^{-a\Delta T})}{(z-1)(z - e^{-a\Delta T})}$$

Hence the ZOH equivalent is

$$(1 - z^{-1})\frac{z(1 - e^{-a\Delta T})}{(z-1)(z - e^{-a\Delta T})} = \frac{1 - e^{-a\Delta T}}{z - e^{-a\Delta T}}$$

Matlab command

In MATLAB, the function *c2d.m* computes the ZOH equivalent of a continuous-time transfer function, as well as other discrete equivalents. For

$$G(s) = \frac{1}{s^2}$$

and $\Delta T = 1$, the following script

```
T=1;
numG=1; denG=[1 0 0];
G = tf(numG,denG);
Gd = c2d(G,T);
```

produces the correct ZOH equivalent.

Exercise

Exercise

Find the zero order hold equivalent of $G(s) = e^{-Ls}$, $2\Delta T < L < 3\Delta T$, where ΔT is the sampling time.

Linear Systems: Stability

Xu Chen

University of Washington

1. Definitions in Lyapunov stability analysis

2. Stability of LTI systems: method of eigenvalue/pole locations

3. Lyapunov's approach to stability
   - Relevant tools
   - Lyapunov stability theorems
   - Instability theorem
   - Discrete-time case

4. Recap

Finite dimensional vector norms

Let $v \in \mathbb{R}^n$. A norm is:

- a metric in vector space: a function that assigns a real-valued length to each vector in a vector space
- e.g., 2 (Euclidean) norm: $\|v\|_2 = \sqrt{v^T v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$

default in this set of notes: $\|\cdot\| = \|\cdot\|_2$

Equilibrium state

For an n-th order unforced system

$$\dot{x} = f\left(x, t\right), \ x(t_0) = x_0$$

an equilibrium state/point x_e is one such that

$$\boxed{f\left(x_e, t\right) = 0, \ \forall t}$$

- the condition must be satisfied by all $t \geq 0$
- if a system starts at equilibrium state, it stays there

Equilibrium state of a linear system

For a linear system

$$\dot{x}(t) = A(t)x(t),\ x(t_0) = x_0$$

- origin $x_e = 0$ is always an equilibrium state
- when $A(t)$ is singular, multiple equilibrium states exist

Lyapunov's definition of stability

- The equilibrium state 0 of $\dot{x} = f(x, t)$ is *stable in the sense of Lyapunov (s.i.L)* if for all $\epsilon > 0$, and t_0, there exists $\delta(\epsilon, t_0) > 0$ such that $\|x(t_0)\|_2 < \delta$ gives $\|x(t)\|_2 < \epsilon$ for all $t \geq t_0$

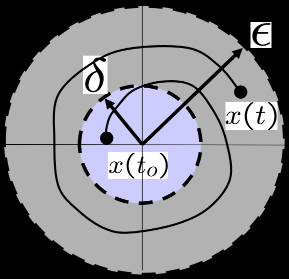


Figure: Stable s.i.L: $\|x(t_0)\| < \delta \Rightarrow \|x(t)\| < \epsilon\ \forall t \geq t_0$.

Asymptotic stability

The equilibrium state 0 of $\dot{x} = f(x, t)$ is asymptotically stable if

- it is stable in the sense of Lyapunov, and
- for all $\epsilon > 0$ and t_0, there exists $\delta(\epsilon, t_0) > 0$ such that $\|x(t_0)\|_2 < \delta$ gives $x(t) \to 0$

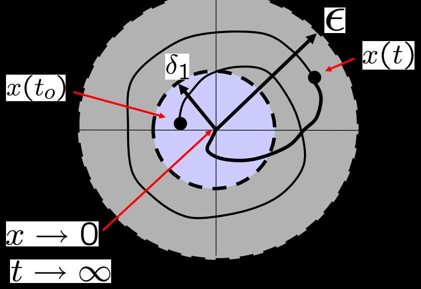


Figure: Asymptotically stable i.s.L: $\|x(t_0)\| < \delta \Rightarrow \|x(t)\| \to 0$.

1. Definitions in Lyapunov stability analysis

2. Stability of LTI systems: method of eigenvalue/pole locations

3. Lyapunov's approach to stability
   - Relevant tools
   - Lyapunov stability theorems
   - Instability theorem
   - Discrete-time case

4. Recap

Stability of LTI systems: method of eigenvalue/pole locations

the stability of the equilibrium point 0 for $\dot{x} = Ax$ or $x(k+1) = Ax(k)$ can be concluded immediately based on $\lambda(A)$:

- the response $e^{At}x(t_0)$ involves modes such as $e^{\lambda t}$, $te^{\lambda t}$, $e^{\sigma t}\cos\omega t$, $e^{\sigma t}\sin\omega t$
- the response $A^k x(k_0)$ involves modes such as λ^k, $k\lambda^{k-1}$, $r^k\cos k\theta$, $r^k\sin k\theta$
- $e^{\sigma t} \to 0$ if $\sigma < 0$; $e^{\lambda t} \to 0$ if $\lambda < 0$
- $\lambda^k \to 0$ if $|\lambda| < 1$; $r^k \to 0$ if $|r| = \left|\sqrt{\sigma^2+\omega^2}\right| = |\lambda| < 1$

Stability of the origin for $\dot{x} = Ax$

stability at 0	$\lambda_i(A)$
unstable	$\mathrm{Re}\{\lambda_i\} > 0$ for some λ_i or $\mathrm{Re}\{\lambda_i\} \le 0$ for all λ_i's but for a repeated λ_m on the imaginary axis with multiplicity m, $\mathrm{nullity}(A - \lambda_m I) < m$ (Jordan form)
stable i.s.L	$\mathrm{Re}\{\lambda_i\} \le 0$ for all λ_i's and $\forall$ repeated λ_m on the imaginary axis with multiplicity m, $\mathrm{nullity}(A - \lambda_m I) = m$ (diagonal form)
asymptotically stable	$\mathrm{Re}\{\lambda_i\} < 0$ $\forall \lambda_i$ (A is then called Hurwitz stable)

Example (Unstable moving mass)

$$\dot{x} = Ax,\ \ A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

- $\lambda_1 = \lambda_2 = 0,\ m = 2,$
  nullity $(A - \lambda_i I) =$ nullity $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = 1 < m$
- i.e., two repeated eigenvalues but needs a generalized eigenvector $\Rightarrow$ Jordan form after similarity transform
- verify by checking $e^{At} = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$: t grows unbounded

Example (Stable in the sense of Lyapunov)

$$\dot{x} = Ax,\ \ A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

- $\lambda_1 = \lambda_2 = 0,\ m = 2,$
  nullity $(A - \lambda_i I) =$ nullity $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 2 = m$
- verify by checking $e^{At} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Routh-Hurwitz criterion

- the Routh Test (by E.J. Routh, in 1877): a simple algebraic procedure to determine how many roots a given polynomial

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

  has in the closed right-half complex plane, without the need to explicitly solve for the roots
- German mathematician Adolf Hurwitz independently proposed in 1895 to approach the problem from a matrix perspective
- popular if stability is the only concern and no details on eigenvalues (e.g., speed of response) are needed

Routh-Hurwitz criterion

- the asymptotic stability of the equilibrium point 0 for $\dot{x} = Ax$ can also be concluded based on the Routh-Hurwitz criterion
- simply apply the Routh Test to $A(s) = \det(sI - A)$
- recap: the poles of transfer function $G(s) = C(sI - A)^{-1} B + D$ come from $\det(sI - A)$ in computing the inverse $(sI - A)^{-1}$

The Routh Array

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

$$\begin{array}{c|ccccc}
s^n & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\
s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \\
s^{n-2} & q_{n-2} & q_{n-4} & q_{n-6} & \cdots & \\
s^{n-3} & q_{n-3} & q_{n-5} & q_{n-7} & \cdots & \\
\vdots & \vdots & \vdots & \vdots & & \\
s^1 & x_2 & x_0 & & & \\
s^0 & x_0 & & & &
\end{array}$$

- first two rows contain the coefficients of $A(s)$
- third row constructed from the previous two rows via

$$\begin{array}{c|cccc}
\cdot & a & b & x & \cdot \\
\cdot & c & d & y & \cdot \\
\cdot & \dfrac{bc-ad}{c} & \dfrac{xc-ay}{c} & \cdot & \\
\cdot & \cdot & \cdot & \cdot & \cdot
\end{array}$$

The Routh Array

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

$$\begin{array}{c|ccccc}
s^n & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\
s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \\
s^{n-2} & q_{n-2} & q_{n-4} & q_{n-6} & \cdots & \\
s^{n-3} & q_{n-3} & q_{n-5} & q_{n-7} & \cdots & \\
\vdots & \vdots & \vdots & \vdots & & \\
s^1 & x_2 & x_0 & & & \\
s^0 & x_0 & & & &
\end{array}$$

- **All roots of $A(s)$ are on the left half s-plane if and only if all elements of the first column of the Routh array are positive.**

The Routh Array

Example $(A(s) = 2s^4 + s^3 + 3s^2 + 5s + 10)$

$$\begin{array}{c|ccc}
s^4 & 2 & 3 & 10 \\
s^3 & 1 & 5 & 0 \\
s^2 & 3 - \frac{2\times 5}{1} = -7 & 10 & 0 \\
s^1 & 5 - \frac{1\times 10}{-7} & 0 & 0 \\
s^0 & 10 & 0 & 0
\end{array}$$

- two sign changes in the first column
- unstable and two roots in the right half side of s-plane

The Routh Array

special cases:

- If the 1st element in any one row of Routh's array is zero, one can replace the zero with a small number ϵ and proceed further.
- There are other possible complications, which we will not pursue further. See, e.g. "Automatic Control Systems", by Kuo, 7th ed., pp. 339-340.

Stability of the origin for $x(k+1) = f(x(k), k)$

- stability analysis follows analogously for nonlinear time-varying discrete-time systems of the form

$$x(k+1) = f(x(k), k), \ x(k_0) = x_0$$

- equilibrium point x_e:

$$f(x_e, k) = x_e, \ \forall k$$

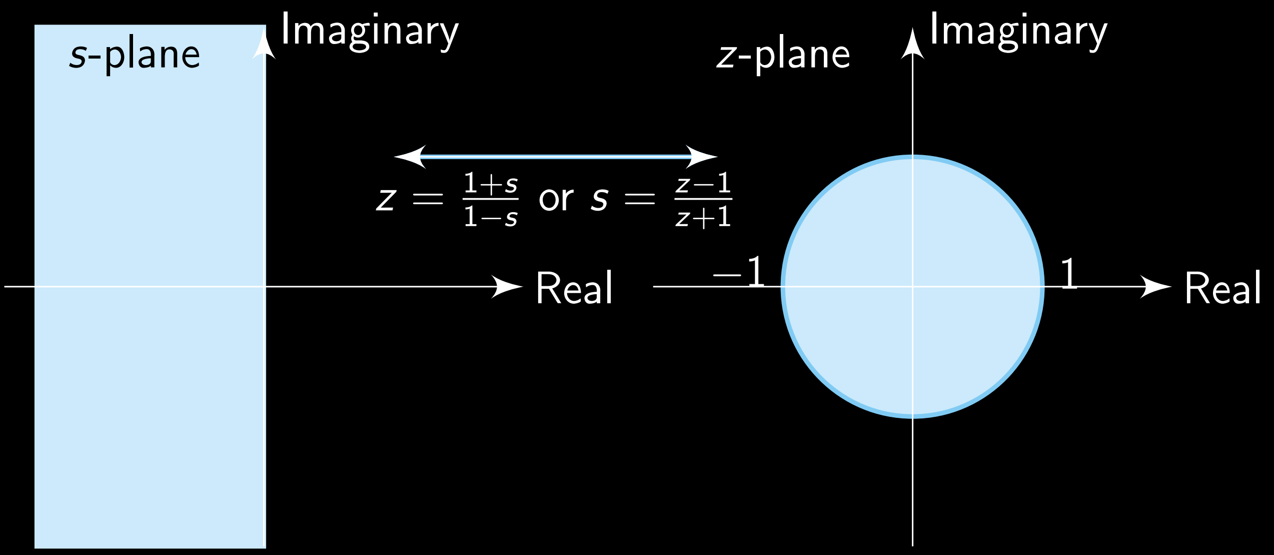
- without loss of generality, 0 is assumed an equilibrium point

Stability of the origin for $x(k+1) = Ax(k)$

stability at 0	$\lambda_i(A)$
unstable	$\lvert\lambda_i\rvert > 1$ for some λ_i or $\lvert\lambda_i\rvert \leq 1$ for all λ_i's but for a repeated λ_m on the unit circle with multiplicity m, nullity $(A - \lambda_m I) < m$ (Jordan form)
stable i.s.L	$\lvert\lambda_i\rvert \leq 1$ for all λ_i's but for any repeated λ_m on the unit circle with multiplicity m, nullity $(A - \lambda_m I) = m$ (diagonal form)
asymptotically stable	$\lvert\lambda_i\rvert < 1 \ \forall \lambda_i$ (such a matrix is called Schur stable)

Routh-Hurwitz criterion for DT LTI systems

- the stability domain $|\lambda_i| < 1$ is a unit disk
- Routh array validates stability in the left-half plane
- bilinear transformation maps the closed left half s-plane to the closed unit disk in z-plane



Routh-Hurwitz criterion for DT LTI systems

- Given $A(z) = z^n + a_1 z^{n-1} + \cdots + a_n$, procedures of Routh-Hurwitz test:
  - apply bilinear transform
    $$A(z)|_{z=\frac{1+s}{1-s}} = \left(\frac{1+s}{1-s}\right)^n + a_1\left(\frac{1+s}{1-s}\right)^{n-1} + \cdots + a_n =: \frac{A^*(s)}{(1-s)^n}$$
  - apply Routh test to
    $$A^*(s) = a_n^* s^n + a_{n-1}^* s^{n-1} + \cdots + a_0^* = A(z)|_{z=\frac{1+s}{1-s}}(1-s)^n$$

Routh-Hurwitz criterion for DT LTI systems

Example ($A(z) = z^3 + 0.8z^2 + 0.6z + 0.5$)

- $A^*(s) = A(z)|_{z=\frac{1+s}{1-s}} (1-s)^3 = (1+s)^3 + 0.8(1+s)^2(1-s) + 0.6(1+s)(1-s)^2 + 0.5(1-s)^3 = 0.3s^3 + 3.1s^2 + 1.7s + 2.9$
- Routh array

$$\begin{array}{c|cc} s^3 & 0.3 & 1.7 \\ s^2 & 3.1 & 2.9 \\ s & 1.7 - \frac{0.3\times 2.9}{3.1} = 1.42 & 0 \\ s^0 & 2.9 & 0 \end{array}$$

- all elements in first column are positive ⇒ roots of $A(z)$ are all in the unit circle

1. Definitions in Lyapunov stability analysis

2. Stability of LTI systems: method of eigenvalue/pole locations

3. Lyapunov's approach to stability
   - Relevant tools
   - Lyapunov stability theorems
   - Instability theorem
   - Discrete-time case

4. Recap

Stability of LTI systems: method of eigenvalue/pole locations

the stability of the equilibrium point 0 for $\dot{x} = Ax$ or $x(k+1) = Ax(k)$ can be concluded immediately based on $\lambda(A)$:

- the response $e^{At}x(t_0)$ involves modes such as $e^{\lambda t}$, $te^{\lambda t}$, $e^{\sigma t}\cos\omega t$, $e^{\sigma t}\sin\omega t$
- the response $A^k x(k_0)$ involves modes such as λ^k, $k\lambda^{k-1}$, $r^k\cos k\theta$, $r^k\sin k\theta$
- $e^{\sigma t} \to 0$ if $\sigma < 0$; $e^{\lambda t} \to 0$ if $\lambda < 0$
- $\lambda^k \to 0$ if $|\lambda| < 1$; $r^k \to 0$ if $|r| = \left|\sqrt{\sigma^2+\omega^2}\right| = |\lambda| < 1$

Lyapunov's approach to stability

The direct method of Lyapunov to stability problems:

- no need for explicit solutions to system responses
- an "energy" perspective
- fit for general dynamic systems (linear/nonlinear, time-invariant/time-varying)

Stability from an energy viewpoint: Example

Consider spring-mass-damper systems:

$$\dot{x}_1 = x_2 \qquad\qquad (x_1\text{: position; } x_2\text{ : velocity})$$

$$\dot{x}_2 = -\frac{k}{m}x_1 - \frac{b}{m}x_2,\ b > 0 \qquad\qquad (\text{Newton's law})$$

- $\lambda\left(A\right)$'s are in the left-half s-plane$\Rightarrow$ asymptotically stable
- total energy

$$\mathcal{E}\left(t\right) = \text{potential energy} + \text{kinetic energy} = \frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2$$

- energy dissipates / is dissipative:

$$\dot{\mathcal{E}}(t) = kx_1\dot{x}_1 + mx_2\dot{x}_2 = -bx_2^2 \leq 0$$

- $\dot{\mathcal{E}} = 0$ only when $x_2 = 0$. As $[x_1, x_2]^T = 0$ is the only equilibrium, the motion will not stop at $x_2 = 0$, $x_1 \neq 0$. Thus energy will keep decreasing toward 0 which is achieved at the origin.

Stability from an energy viewpoint: Generalization

Consider unforced, time-varying, nonlinear systems

$$\dot{x}(t) = f\left(x(t), t\right),\ \ x\left(t_0\right) = x_0$$
$$x\left(k+1\right) = f\left(x(k), k\right),\ \ x\left(k_0\right) = x_0$$

- assume the origin is an equilibrium state
- energy function $\Rightarrow$ Lyapunov function: a scalar function of x and t (or x and k)
- goal is to relate properties of the state through the Lyapunov function
- main tool: matrix formulation, linear algebra, positive definite functions

Relevant tools

Quadratic functions

- intrinsic in energy-like analysis, e.g.

$$\frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2 = \frac{1}{2}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} k & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- convenience of matrix formulation:

$$\frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2 + x_1x_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} \frac{k}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{m}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2 + x_1x_2 + c = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}^T \begin{bmatrix} \frac{k}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{m}{2} & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

- general quadratic functions in matrix form

$$Q(x) = x^T P x, \;\; P^T = P$$

Relevant tools

Symmetric matrices

- recall: a real square matrix A is
  - *symmetric* if $A = A^T$
  - *skew-symmetric* if $A = -A^T$
- examples:

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$$

- *Any* real square matrix can be decomposed as the sum of a *symmetric* matrix and a *skew-symmetric* matrix:

$$\text{e.g.} \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2.5 \\ 2.5 & 4 \end{bmatrix} + \begin{bmatrix} 0 & -0.5 \\ 0.5 & 0 \end{bmatrix}$$

$$\text{general case: } P = \frac{P + P^T}{2} + \frac{P - P^T}{2}$$

Relevant tools

Symmetric matrices

- a real square matrix $A \in \mathbb{R}^{n\times n}$ is *orthogonal* if $A^T A = AA^T = I$
- meaning that the columns of A form a orthonormal basis of $\mathbb{R}^n$

$$A = \left[\begin{array}{cccc} | & | & | & | \\ a_1 & a_2 & \dots & a_n \\ | & | & | & | \end{array}\right]$$

$$A^T A = \left[\begin{array}{cccc} a_1^T a_1 & a_1^T a_2 & \dots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \dots & a_2^T a_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \dots & a_n^T a_n \end{array}\right] = \left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{array}\right]$$

namely, $a_j^T a_j = 1$ and $a_j^T a_m = 0 \ \forall j \neq m$.

Theorem

The eigenvalues of symmetric matrices are all real.

Proof: $\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = A$.

Eigenvalue-eigenvector pair: $Au = \lambda u \Rightarrow \overline{u}^T Au = \lambda \overline{u}^T u$, where $\overline{u}$ is the complex conjugate of u. $\overline{u}^T Au$ is a real number, as

$$\begin{aligned} \overline{\overline{u}^T Au} &= u^T \overline{A}\overline{u} \\ &= u^T A\overline{u} \quad \because A \in \mathbb{R}^{n\times n} \\ &= u^T A^T \overline{u} \quad \because A = A^T \\ &= \lambda u^T \overline{u} \quad \because (Au)^T = (\lambda u)^T \\ &= \lambda \overline{u}^T u \quad \because u^T \overline{u} \in \mathbb{R} \\ &= \overline{u}^T Au \quad \because Au = \lambda u \end{aligned}$$

Also, $\overline{u}^T u \in \mathbb{R}$. Thus $\lambda = \frac{\overline{u}^T Au}{\overline{u}^T u}$ must also be a real number. □

Example

- $\begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$: $\lambda = \pm 2$
- $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$: $\lambda = 1 \pm 2$

```
import numpy as np #larger-scale Python example
N = 100
P = np.random.randint(-200,200,size=(N,N))
P_symm = (P + P.T)/2
lambdas, _ = np.linalg.eig(P_symm)
print(lambdas)
```

Theorem

The eigenvalues of skew-symmetric matrices are all imaginary or zero.

- $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$: $\lambda = \pm 2j$

```
import numpy as np
N = 100
P = np.random.randint(-200,200,size=(N,N))
P_symm = (P - P.T)/2
lambdas, _ = np.linalg.eig(P_symm)
print(lambdas)
```

Theorem

All eigenvalues of an orthogonal matrix have a magnitude of 1.

- $\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} : \lambda = 1 \pm 2j$

```
import numpy as np
from scipy.linalg import qr
n = 3
H = np.random.randn(n, n)
Q, _ = qr(H)
print (np.dot(Q,Q.T))
print (np.dot(Q.T,Q))
```

Important properties of symmetric matrices

Theorem

The eigenvalues of symmetric matrices are all real.

Theorem

The eigenvalues of skew-symmetric matrices are all imaginary or zero.

Theorem

All eigenvalues of an orthogonal matrix have a magnitude of 1.

matrix structure	analogy in complex plane
symmetric	real line
skew-symmetric	imaginary line
orthogonal	unit circle

The spectral theorem for symmetric matrices

When $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, we can do diagonalization $A = U\Lambda U^{-1}$. When A is symmetric, things are even better:

Theorem (Symmetric eigenvalue decomposition (SED))

$\forall : A \in \mathbb{R}^{n\times n},\ A^T = A$, *there always exist* $\lambda_i \in \mathbb{R}$ *and* $u_i \in \mathbb{R}^n$, *s.t.*

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U\Lambda U^T \tag{1}$$

- λ_i's: eigenvalues of A
- u_i: eigenvector associated to λ_i, normalized to have unity norms
- $U = [u_1, u_2, \cdots, u_n]$ is orthogonal: $U^T U = UU^T = I$
- $\Lambda = diagonal(\lambda_1, \lambda_2, \ldots, \lambda_n)$

Elements of proof for SED

Theorem

$\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = A$, *then eigenvectors of A, associated with different eigenvalues, are* ***orthogonal****.*

Proof.

Let $Au_i = \lambda_i u_i$ and $Au_j = \lambda_j u_j$. Then $u_i^T A u_j = u_i^T \lambda_j u_j = \lambda_j u_i^T u_j$. Also, $u_i^T A u_j = u_i^T A^T u_j = (Au_i)^T u_j = \lambda_i u_i^T u_j$. So $\lambda_i u_i^T u_j = \lambda_j u_i^T u_j$. But $\lambda_i \neq \lambda_j$. It must be that $u_i^T u_j = 0$. □

SED now follows:

- If A has distinct eigenvalues, then $U = [u_1, u_2, \cdots, u_n]$ is orthogonal after normalizing all the eigenvectors to unity norm.
- If A has $r(< n)$ distinct eigenvalues, we can *choose* multiple orthogonal eigenvectors for the eigenvalues with none-unity multiplicities.

Rethinking symmetric matrices

With the spectral theorem, next time we see a symmetric matrix A, we immediately know that

- λ_i is real for all i
- associated with λ_i, we can always find a real eigenvector
- $\exists$ an orthonormal basis $\{u_i\}_{i=1}^n$, which consists of the eigenvectors
- if $A \in \mathbb{R}^{2\times 2}$, then if you compute first λ_1, λ_2 and u_1, you won't need to go through the regular math to get u_2, but can simply solve for a u_2 that is orthogonal to u_1 with $\|u_2\| = 1$.

Example: $A = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}$

Computing the eigenvalues gives

$$\det \begin{bmatrix} 5-\lambda & \sqrt{3} \\ \sqrt{3} & 7-\lambda \end{bmatrix} = 35 - 12\lambda + \lambda^2 - 3 = (\lambda - 4)(\lambda - 8) = 0$$

$$\Rightarrow \lambda_1 = 4,\ \lambda_2 = 8$$

- first normalized eigenvector:

$$(A - \lambda_1 I)\, t_1 = 0 \Rightarrow \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

- A is symmetric $\Rightarrow$ eigenvectors are orthogonal to each other: *choose* $t_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$. No need to solve $(A - \lambda_2 I)\, t_2 = 0$!

Theorem (Eigenvalues of symmetric matrices)

If $A = A^T \in \mathbb{R}^{n\times n}$, then the eigenvalues of A satisfy

$$\lambda_{\max} = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^TAx}{\|x\|_2^2} \tag{2}$$

$$\lambda_{\min} = \min_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^TAx}{\|x\|_2^2} \tag{3}$$

Proof.

Perform SED to get $A = \sum_{i=1}^n \lambda_i u_i^T u_i$ where $\{u_i\}_{i=1}^n$ spans $\mathbb{R}^n$. Then any vector $x \in \mathbb{R}^n$ can be decomposed as $x = \sum_{i=1}^n \alpha_i u_i$. Thus

$$\max_{x\neq 0} \frac{x^TAx}{\|x\|_2^2} = \max_{\alpha_i} \frac{\left(\sum_i \alpha_i u_i\right)^T \sum_i \lambda_i \alpha_i u_i}{\sum_i \alpha_i^2} = \max_{\alpha_i} \frac{\sum_i \lambda_i \alpha_i^2}{\sum_i \alpha_i^2} = \lambda_{\max}$$

□

Positive definite matrices

- eigenvalues of symmetric matrices are real ⇒ we can order the eigenvalues
- a symmetric matrix P is called positive-definite if all its eigenvalues are positive
- equivalently:

Definition (Positive Definite Matrices)

A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **positive-definite**, written $P \succ 0$, if $x^TPx > 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$.
P is called **positive-semidefinite**, written $P \succeq 0$, if $x^TPx \geq 0$ for all $x \in \mathbb{R}^n$

- $P \succ 0$ $(P \succeq 0)$ ⇔ P can be decomposed as $P = N^TN$ where N is nonsingular (singular)

Negative definite matrices

Definition

A symmetric matrix $Q \in \mathbb{R}^{n\times n}$ is called **negative-definite**, written $Q \prec 0$, if $-Q \succ 0$, i.e., $x^T Qx < 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$.
Q is called **negative-semidefinite,** written $Q \preceq 0$, if $x^T Qx \leq 0$ for all $x \in \mathbb{R}^n$

- When A and B have compatible dimensions, $A \succ B$ means $A - B \succ 0$.

Positive definite matrices

- positive-definite matrices can have negative entries:

Example

$P = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ is positive-definite, as $P = P^T$ and take any $v = [x, y]^T$, we have

$$\begin{aligned} v^T Pv &= \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 2y^2 - 2xy \\ &= x^2 + y^2 + (x-y)^2 \geq 0 \end{aligned}$$

and the equality sign holds only when $x = y = 0$.

Positive definite matrices

- conversely, matrices whose entries are all positive are not necessarily positive-definite:

Example

$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is not positive-definite:

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 < 0$$

Positive definite matrices

Theorem

For a symmetric matrix P, $P \succ 0$ if and only if all the eigenvalues of P are positive.

Proof.

Since P is symmetric, we have

$$\lambda_{\max}(P) = \max_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{4}$$

$$\lambda_{\min}(P) = \min_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{5}$$

which gives $x^T A x \in [\lambda_{\min}\|x\|_2^2,\ \lambda_{\max}\|x\|_2^2]$. Thus $x^T A x > 0,\ x \neq 0 \Leftrightarrow \lambda_{\min} > 0$. □

Relevant tools

Checking positive definiteness of a matrix.

We often use the following necessary and sufficient conditions to check positive (semi-)definiteness:

- $P \succ 0$ $(P \succeq 0)$ $\Leftrightarrow$ the leading principle minors defined below are positive (nonnegative)

Definition

The leading principle minors of $P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$ are defined as

p_{11}, $\det \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$, $\det P$.

Relevant tools

Checking positive definiteness of a matrix.

Example

None of the following matrices are positive definite:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

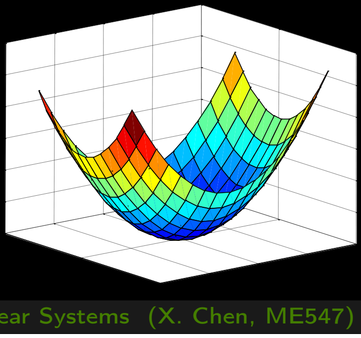
Relevant tools

Definition (Positive Definite Functions)

A continuous time function $W : \mathbb{R}^n \to \mathbb{R}_+$, called to be PD, satisfying

- $W(x) > 0$ for all $x \neq 0$
- $W(0) = 0$
- $W(x) \to \infty$ as $|x| \to \infty$ uniformly in x

In the 3D space, positive definite functions are "bowl-shaped", e.g., $W(x_1, x_2) = x_1^2 + x_2^2$.



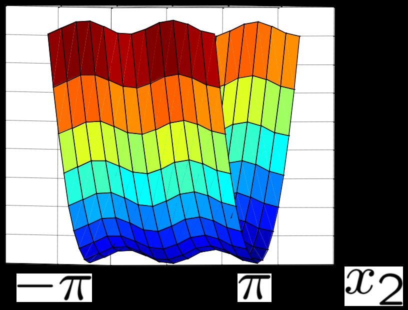
Relevant tools

Definition (Locally Positive Definite Functions)

A continuous time function $W : \mathbb{R}^n \to \mathbb{R}_+$, called to be LPD, satisfying

- $W(x) > 0$ for all $x \neq 0$ and $|x| < r$
- $W(0) = 0$

In the 3D space, locally positive definite functions are "bowl-shaped" locally, e.g., $W(x_1, x_2) = x_1^2 + \sin^2 x_2$ for $x_1 \in \mathbb{R}$ and $|x_2| < \pi$



Relevant tools

Exercise

Let $x = [x_1, x_2, x_3]^T$. Check the positive definiteness of the following functions

1. $V(x) = x_1^4 + x_2^2 + x_3^4$ (PD)
2. $V(x) = x_1^2 + x_2^2 + 3x_3^2 - x_3^4$ (LPD for $|x_3| < \sqrt{3}$)

1. Definitions in Lyapunov stability analysis
2. Stability of LTI systems: method of eigenvalue/pole locations
3. Lyapunov's approach to stability
   - Relevant tools
   - Lyapunov stability theorems
   - Instability theorem
   - Discrete-time case
4. Recap

Lyapunov stability theorems

- recall the spring mass damper example in matrix form

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- energy function is PD:
  $\mathcal{E}\left(t\right) = \text{potential energy} + \text{kinetic energy} = \frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2$
  and its derivative is NSD:

$$\begin{aligned} \dot{\mathcal{E}}(t) &= \left[\frac{\partial \mathcal{E}}{\partial x_1}, \frac{\partial \mathcal{E}}{\partial x_2}\right]\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = k_1 x_1 \dot{x}_1 + m x_2 \dot{x}_2 && (6) \\ &= k_1 x_1 x_2 + m x_2 \left(-\frac{k}{m}x_1 - \frac{b}{m}x_2\right) = \left[\frac{\partial \mathcal{E}}{\partial x_1}, \frac{\partial \mathcal{E}}{\partial x_2}\right] Ax && (7) \\ &= -bx_2^2 \end{aligned}$$

The notion of derivative along state trajectories

- Generalizing the concept to system $\dot{x} = f\left(x\right)$: let $V\left(x\right)$ be a general energy function, the energy dissipation w.r.t. time is

$$\frac{dV(x)}{dt} = \left[\frac{\partial V}{\partial x_1}, \frac{\partial V}{\partial x_2}, \dots, \frac{\partial V}{\partial x_n}\right]\begin{bmatrix} f_1\left(x\right) \\ \vdots \\ f_n\left(x\right) \end{bmatrix}$$

  also denoted as $L_f V(x)$, the Lie derivative of $V(x)$ w.r.t. $f(x)$.
- We concluded stability of the system by analyzing how energy will dissipate to zero along the trajectory of the state.

Theorem

The equilibrium point 0 *of* $\dot{x}(t) = f(x(t), t)\,,\ x(t_0) = x_0$ *is stable in the sense of Lyapunov if there exists a locally positive definite function* $V(x, t)$ *such that* $\dot{V}(x, t) \leq 0$ *for all* $t \geq t_0$ *and all* x *in a local region* $x : |x| < r$ *for some* $r > 0$*.*

- such a $V(x, t)$ is called a Lyapunov function
- i.e., $V(x)$ is PD and $\dot{V}(x)$ is negative semidefinite in a local region $|x| < r$

Theorem

The equilibrium point 0 *of* $\dot{x}(t) = f(x(t), t)\,,\ x(t_0) = x_0$ *is locally asymptotically stable if there exists a Lyapunov function* $V(x)$ *such that* $\dot{V}(x)$ *is locally negative definite.*

Theorem

The equilibrium point 0 *of* $\dot{x}(t) = f(x(t), t)\,,\ x(t_0) = x_0$ *is globally asymptotically stable if there exists a Lyapunov function* $V(x)$ *such that* $V(x)$ *is positive definite and* $\dot{V}(x)$ *is negative definite.*

Lyapunov stability concept for linear systems

- for linear system $\dot{x} = Ax$, a good Lyapunov candidate is the quadratic function $V(x) = x^T Px$ where $P = P^T$ and $P \succ 0$
- the derivative along the state trajectory is then

$$\begin{aligned}\dot{V}(x) &= \dot{x}^T Px + x^T P\dot{x}\\ &= (Ax)^T Px + x^T PAx\\ &= x^T\left(A^T P + PA\right)x\end{aligned}$$

- such a $V(x) = x^T Px$ is a Lyapunov function for $\dot{x} = Ax$ when $A^T P + PA \preceq 0$
- and the origin is stable in the sense of Lyapunov

Theorem (Lyapunov stability theorem for linear systems)

For $\dot{x} = Ax$ with $A \in \mathbb{R}^{n\times n}$, the origin is asymptotically stable if and only if for any symmetric positive definite matrix $Q \succ 0$, the Lyapunov equation

$$\boxed{A^T P + PA = -Q}$$

has a unique positive definite solution $P \succ 0$, $P^T = P$.

Proof.

“⇒”: $\frac{\dot{V}}{V} = -\frac{x^T Qx}{x^T Px} \le -\underbrace{\frac{(\lambda_Q)_{\min}}{(\lambda_P)_{\max}}}_{\triangleq\alpha} \Longrightarrow V(t) \le e^{-\alpha t}V(0)$. $Q \succ 0$ and $P \succ 0 \Rightarrow (\lambda_Q)_{\min} > 0$ and $(\lambda_P)_{\max} > 0$. Thus $\alpha > 0$; $V(t)$ decays exponentially to zero. $V(x) \succ 0 \Rightarrow V(x) = 0$ only at $x = 0$. Therefore, $x \to 0$ as $t \to \infty$, regardless of the initial condition. □

Proof.

"$\Leftarrow$": if 0 of $\dot{x} = Ax$ is asymptotically stable, then all eigenvalues of A have negative real parts. For any Q, the Lyapunov equation has a unique solution P. Note $x(t) = e^{At}x_0 \to 0$ as $t \to \infty$. We have

$$\cancelto{0}{x^T(\infty)Px(\infty)} - x^T(0)Px(0) = \int_0^\infty \frac{d}{dt}x^T(t)Px(t)\,dt = \int_0^\infty x^T(t)\left(A^TP + PA\right)x(t)\,dt$$

$$\Rightarrow x(0)^T Px(0) = \int_0^\infty x^T(t)Qx(t)\,dt = \int_0^\infty x(0)e^{A^Tt}Qe^{At}x(0)\,dt$$

If $Q \succ 0$, there exists a nonsingular N matrix: $Q = N^TN$. Thus

$$x(0)^T Px(0) = \int_0^\infty \|Ne^{At}x(0)\|^2 dt \geq 0$$

$$x(0)^T Px(0) = 0 \text{ only if } x_0 = 0$$

Thus $P \succ 0$. Furthermore

$$\boxed{P = \int_0^\infty e^{A^Tt}Qe^{At}dt}$$

□

Lyapunov stability theorems

Example

$\dot{x} = Ax$, $A = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}$. The Lyapunov equation is

$$\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}^T \underbrace{\begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}}_{P} + \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix}\begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} = -\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{Q}$$

We need

$$\begin{cases} -2p_{11} - 2p_{12} = -1 \\ -p_{12} - p_{22} + p_{11} = 0 \\ 2p_{12} = -1 \end{cases} \Rightarrow \begin{cases} p_{11} = 1 \\ p_{22} = 3/2 \\ p_{12} = -1/2 \end{cases}$$

Leading principle minors: $p_{11} > 0$, $p_{11}p_{22} - p_{12}^2 > 0$
$\Rightarrow P \succ 0 \Rightarrow$asymptotically stable

Essense of the Lyapunov Eq.

Observations:

- $A^T P + PA$ is a linear operation on P: e.g.,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \; Q = \begin{bmatrix} | & | \\ q_1 & q_2 \\ | & | \end{bmatrix}, \; P = \begin{bmatrix} | & | \\ p_1 & p_2 \\ | & | \end{bmatrix}$$

$$A^T \begin{bmatrix} | & | \\ p_1 & p_2 \\ | & | \end{bmatrix} + \begin{bmatrix} | & | \\ p_1 & p_2 \\ | & | \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = - \begin{bmatrix} | & | \\ q_1 & q_2 \\ | & | \end{bmatrix}$$

$$\begin{aligned} A^T p_1 + a_{11} p_1 + a_{21} p_2 &= -q_1 \\ A^T p_2 + a_{12} p_1 + a_{22} p_2 &= -q_2 \end{aligned}$$

Essense of the Lyapunov Eq.

Observations: with now

$$A^T P + PA = Q \Leftrightarrow \begin{cases} A^T p_1 + a_{11} p_1 + a_{21} p_2 & = -q_1 \\ A^T p_2 + a_{12} p_1 + a_{22} p_2 & = -q_2 \end{cases}$$

- can stack the columns of $A^T P + PA$ and Q to yield

$$\begin{bmatrix} A^T & 0 \\ 0 & A^T \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} + \begin{bmatrix} a_{11} I & a_{21} I \\ a_{12} I & a_{22} I \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = - \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$$

$$\underbrace{\left\{ \begin{bmatrix} A^T & 0 \\ 0 & A^T \end{bmatrix} + \begin{bmatrix} a_{11} I & a_{21} I \\ a_{12} I & a_{22} I \end{bmatrix} \right\}}_{L_A} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = - \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$$

The Lyapunov Eq.: Existence of solution

$$L_A(P) = A^T P + PA$$

- L_A is invertible if and only if $\lambda_i + \lambda_j \neq 0$ for all eigenvalues of A:
  - let $A^T u_i = \lambda_i u_i$ and $A^T u_j = \lambda_j u_j$
  - $L_A\left(u_i u_j^T\right) = u_i u_j^T A + A^T u_i u_j^T = u_i\left(\lambda_j u_j\right)^T + \lambda_i u_i u_j^T = \left(\lambda_i + \lambda_j\right) u_i u_j^T$
  - so $\lambda_i + \lambda_j$ is an eigenvalue of the operator $L_A(\cdot)$
  - if $\lambda_i + \lambda_j \neq 0$, the operator is invertible

The Lyapunov operator: eigenvalues

$$L_A = \begin{bmatrix} A^T & 0 \\ 0 & A^T \end{bmatrix} + \begin{bmatrix} a_{11}I & a_{21}I \\ a_{12}I & a_{22}I \end{bmatrix}$$

- can simply write $L_A = \underbrace{I \otimes A^T + A^T \otimes I}_{\text{mirror symmetric}}$ using the Kronecker product notation $B \otimes C = \begin{bmatrix} b_{11}C & b_{11}C & \dots & b_{11}C \\ b_{21}C & b_{22}C & \dots & b_{2n}C \\ \vdots & \vdots & \dots & \vdots \\ b_{m1}C & b_{m2}C & \dots & b_{mn}C \end{bmatrix}$

The Lyapunov operator: eigenvalues

$$L_A = \begin{bmatrix} A^T & 0 \\ 0 & A^T \end{bmatrix} + \begin{bmatrix} a_{11}I & a_{21}I \\ a_{12}I & a_{22}I \end{bmatrix}$$

- e.g., $A = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}$

$$L_A = I \otimes A^T + A^T \otimes I = \begin{bmatrix} A^T + a_{11}I & a_{21}I \\ a_{12}I & A^T + a_{22}I \end{bmatrix}$$

$$= \left[\begin{array}{cc|cc} -1-1 & -1 & -1 & 0 \\ 1 & 0-1 & 0 & -1 \\ \hline 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 \end{array}\right] = \left[\begin{array}{cc|cc} -2 & -1 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ \hline 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

Example: $A = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix}$, $\lambda_{1,2} = -0.5 \pm i\sqrt{3}/2$

$$L_A = I \otimes A^T + A^T \otimes I = \left[\begin{array}{cc|cc} -2 & -1 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ \hline 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

The eigenvalues of L_A are -1, -1, $-1-\sqrt{3}$, $-1+\sqrt{3}$, which are precisely $\lambda_1 + \lambda_1$, $\lambda_1 + \lambda_2$, $\lambda_2 + \lambda_1$, $\lambda_2 + \lambda_2$.

```
import numpy as np
A = [[-1,1],[-1,0]]; I2=np.eye(2); AT=np.transpose(A)
L_A=np.kron(I2,AT)+np.kron(AT,I2)
eigLA,_=np.linalg.eig(L_A)
eigA,_=np.linalg.eig(A)
print(eigLA)
print(eigA)
```

Procedures of Lyapunov's direct method

1. Given A, select an arbitrary positive-definite symmetric matrix Q (e.g., I).
2. Find the solution matrix P to the Lyapunov equation $A^TP + PA = -Q$.
3. If a solution P cannot be found, the origin is not asymptotically stable.
4. If a solution is found:
   - if P is positive-definite, then A is Hurwitz stable and the origin is asymptotically stable;
   - if P is not positive-definite, then A has at least one eigenvalue with a positive real part and the origin is an unstable equilibrium.

It suffices to select $Q = I$

For linear systems we can let $Q = I$ and check whether the resulting P is positive definite. If it is, then we can assert the asymptotic stability:

- take any $Q \succ 0$. there exists $Q = N^TN$, where N is invertible, yielding

$$A^TP + PA = -I$$

$$\Updownarrow$$

$$\underbrace{N^TA^TN^{-T}}_{\tilde{A}^T}\underbrace{N^TPN}_{\tilde{P}} + \underbrace{N^TPN}_{\tilde{P}}\underbrace{N^{-1}AN}_{\tilde{A}} = -N^TN$$

  - $\tilde{A} = N^{-1}AN$ and A are similar matrices and have the same eigenvalues.
  - $\tilde{P} = N^TPN$ and P have the same definiteness. If we can find a positive definite solution P then the $\tilde{P}$ will also be positive definite. Vise versa.

Instability theorem

- for nonlinear systems, Lyapunov function can be nontrivial to find
- failure to find a Lyapunov function does not imply instability

Theorem

The equilibrium state 0 *of* $\dot{x} = f(x)$ *is unstable if there exists a function* $W(x)$ *such that*

- $\dot{W}(x)$ *is PD locally:* $\dot{W}(x) > 0 \; \forall \, |x| < r$ *for some* r *and* $\dot{W}(0) = 0$
- $W(0) = 0$
- *there exist states* x *arbitrarily close to the origin such that* $W(x) > 0$

Discrete-time case: key concept of Lyapunov

For the discrete-time system

$$x(k+1) = Ax(k)$$

we consider a quadratic Lyapunov function candidate

$$V(x) = x^T P x, \; P = P^T \succ 0$$

and compute $\Delta V(x)$ along the trajectory of the state

$$V(x(k+1)) - V(x(k)) = x^T(k) \underbrace{\left[A^T P A - P\right]}_{\triangleq -Q} x(k)$$

Asymptotic stability desires $\Delta V(x)$ to be negative.

DT Lyapunov stability theorem for linear systems

Theorem

For system $x(k+1) = Ax(k)$ with $A \in \mathbb{R}^{n\times n}$, the origin is asymptotically stable if and only if $\exists$ $Q \succ 0$, such that the discrete-time Lyapunov equation

$$\boxed{A^T PA - P = -Q}$$

has a unique positive definite solution $P \succ 0$, $P^T = P$.

The DT Lyapunov Eq.

$$\boxed{A^T PA - P = -Q}$$

- Solution to the DT Lyapunov equation, when asymptotic stability holds (A is Schur stable), comes from:

$$\cancelto{0}{V(x(\infty))} - V(x(0)) = \sum_{k=0}^{\infty} x^T(k)\left[A^T PA - P\right]x(k)$$
$$= -\sum_{k=0}^{\infty} x^T(0)\left(A^T\right)^k QA^k x(0)$$
$$\Rightarrow P = \sum_{k=0}^{\infty} \left(A^T\right)^k QA^k$$

- can show that the DT Lyapunov operator $L_A = A^T PA - P$ is invertible if and only if $\forall i, j$ $(\lambda_A)_i (\lambda_A)_j \neq 1$

DT Lyapunov analysis with MATLAB

Example

$$x(k+1) = Ax(k),\ A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.275 & -0.225 & -0.1 \end{bmatrix}$$

```
% MATLAB
A=[ 0 1 0; 0 0 1; 0.275 -0.225 -0.1]
Q = eye(3)
P = dlyap(A',Q) % check function definition in Matlab help
eig(P)
```

DT Lyapunov analysis with Python

Example

$$x(k+1) = Ax(k),\ A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.275 & -0.225 & -0.1 \end{bmatrix}$$

```
#Python
import control as ct
import numpy as np
from numpy.linalg import eig
A = np.array([[0,1,0],[0,0,1],[0.275,-0.225,-0.1]])
Q = np.identity(3)
P = ct.dlyap(A.transpose(),Q)
w,v = eig(P)
print(w)
```

Recap

- Internal stability
  - Stability in the sense of Lyapunov: ε, δ conditions
  - Asymptotic stability
- Stability analysis of linear time invariant systems ($\dot{x} = Ax$ or $x(k+1) = Ax(k)$)
  - Based on the eigenvalues of A
    - Time response modes
    - Repeated eigenvalues on the imaginary axis
  - Routh's criterion
    - No need to solve the characteristic equation
    - Discrete time case: bilinear transform ($z = \frac{1+s}{1-s}$)

Recap

- Lyapunov equations

  Theorem: All eigenvalues of A have negative real parts iff for any given $Q \succ 0$, the Lyapunov equation

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

  has a unique solution P and $P \succ 0$.

  Given Q, the Lyapunov equation $A^T P + PA = -Q$ has a unique solution when $\lambda_{A,i} + \lambda_{A,j} \neq 0$ for all i and j.

  Theorem: All eigenvalues of A are inside the unit circle iff for any given $Q \succ 0$, the Lyapunov equation

  $$A^T PA - P = -Q$$

  has a unique solution P and $P \succ 0$.

  Given Q, the Lyapunov equation $A^T PA - P = -Q$ has a unique solution when $\lambda_{A,i}\lambda_{A,j} \neq 1$ for all i and j.

Recap

- P is positive definite if and only if any one of the following conditions holds:
  1. All the eigenvalues of P are positive.
  2. All the leading principle minors of P are positive.
  3. There exists a nonsingular matrix N such that $P = N^T N$.

Controllability and Observability

Xu Chen

University of Washington

The concept of controllability and observability

Controllability:

- inputs do not act directly on the states but via state dynamics:

$$\dot{x}(t) = Ax(t) + Bu(t) \text{ or } x(k+1) = Ax(k) + Bu(k) \tag{1}$$

- can the inputs drive the system to any value in the state space in finite time?

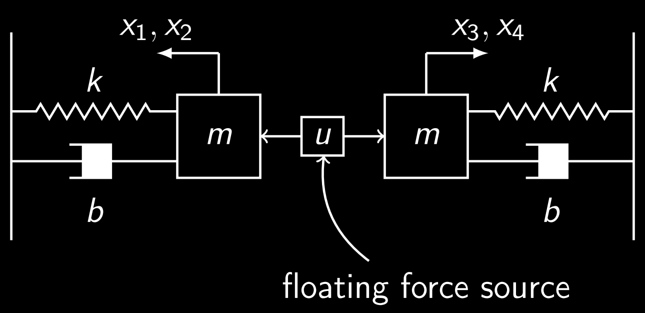
Observability:

- states are not all measured directly but instead impact the output via the output equation:

$$y = Cx + Du$$

- can we infer fully the initial state from the outputs and the inputs? (can then reveal the full state trajectory through (1))

The concept of controllability and observability



$$\dot{x}_1 = x_2$$
$$\dot{x}_3 = x_4$$

- assume $x(0) = 0$
- because of symmetry, we always have

$$x_1(t) = x_3(t), \; x_2(t) = x_4(t), \; \forall t \geq 0$$

- state cannot be arbitrarily steered $\Rightarrow$ uncontrollable

Controllability definition in discrete time

Definition

A discrete-time linear system $x(k+1) = A(k)x(k) + B(k)u(k)$ is called controllable at $k = 0$ if $\exists$ a finite time k_1 such that $\forall$ initial state $x(0)$ and target state x_1, there exists a control sequence $\{u(k); k = 0, 1, \ldots, k_1\}$ that will transfer the system from $x(0)$ at $k = 0$ to x_1 at $k = k_1$.

Controllability of LTI systems

$$x(k+1) = Ax(k) + Bu(k) \Rightarrow x(n) = A^n x(0) + \sum_{k=0}^{n-1} A^{n-1-k} Bu(k)$$

$$\Rightarrow \boxed{x(n) - A^n x(0) = \underbrace{\left[B, AB, A^2B, \dots, A^{n-1}B\right]}_{P_d} \underbrace{\begin{bmatrix} u(n-1) \\ u(n-2) \\ \vdots \\ u(0) \end{bmatrix}}_{\mathrm{u}_n}}$$

- given any $x(n)$ and $x(0)$ in $\mathbb{R}^n$, u_n can be solved if the columns of P_d span $\mathbb{R}^n$
- equivalently, system is controllable if P_d has rank n (full row rank)

Controllability of LTI systems Cont'd

$$x(k+1) = Ax(k) + Bu(k) \Rightarrow$$

$$\boxed{x(n) - A^n x(0) = \underbrace{\left[B, AB, A^2B, \dots, A^{n-1}B\right]}_{P_d} \underbrace{\begin{bmatrix} u(n-1) \\ u(n-2) \\ \vdots \\ u(0) \end{bmatrix}}_{\mathrm{u}_n}}$$

- also, no need to go beyond n: adding A^nB, $A^{n+1}B$, $\dots$ does not increase the rank of P_d (Cayley Halmilton Theorem):

$$x(k_1) - A^{k_1} x(0) = \underbrace{\left[\begin{array}{cccc|cc} B & AB & \dots & A^{n-1}B & \dots & A^{k_1-1}B \end{array}\right]}_{\text{rank=rank}(P_d)} \begin{bmatrix} u(k_1-1) \\ u(k_1-2) \\ \vdots \\ u(0) \end{bmatrix}$$

Theorem (Cayley Halmilton Theorem)

Let $A \in \mathbb{R}^{n\times n}$. A^n is linearly dependent with $\{I,\ A,\ A^2, \cdots A^{n-1}\}$

Proof.

Consider characteristic polynomial

$$\begin{aligned} p(\lambda) &= \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 = \det(\lambda I - A) \\ &= (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_p)^{m_p} \end{aligned}$$

$$\begin{aligned} \Rightarrow p(A) &= A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I \\ &= (A - \lambda_1 I)^{m_1} \dots (A - \lambda_p I)^{m_p}, \quad m_1 + m_2 + \cdots + m_p = n \end{aligned}$$

$\forall$ eigenvector or generalized eigenvector t_i, say, associated to λ_i:
$p(A)\,t_i = (A - \lambda_1 I)^{m_1} \dots \underline{(A - \lambda_p I)^{m_p}\, t_i} =$
$(A - \lambda_1 I)^{m_1} \dots \underline{(A - \lambda_p I)^{m_p - 1} (\lambda_i t_i - \lambda_p t_i)} = (\lambda_i - \lambda_1)^{m_1} \dots (\lambda_i - \lambda_p)^{m_p}\, t_i = 0$

- therefore $p(A)\,[t_1, t_2, \dots, t_n] = 0$
- but $T = [t_1, t_2, \dots, t_n]$ is invertible. Hence $p(A) = 0$
  $\Rightarrow A^n = -c_0 I - c_1 A - \cdots - c_{n-1}A^{n-1}$ □

Arthur Cayley: 1821-1895, British mathematician

- algebraic theory of curves and surfaces, group theory, linear algebra, graph theory, invariant theory, ...
- extraordinarily prolific career: ˜1,000 math papers

William Hamilton: 1805-1865, Irish mathematician

- optics and classical mechanics in physics, dynamics, algebra, quaternions, ...
- quaternions: extending complex numbers to higher spatial dimensions: 4D case

  $$i^2 = j^2 = k^2 = ijk = -1$$

  now used in computer graphics, control theory, orbital mechanics, e.g., spacecraft attitude-control systems

Theorem (Controllability Theorem)

The n-dimensional r-input LTI system with $x(k+1) = Ax(k) + Bu(k)$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times r}$ is controllable if and only if either one of the following is satisfied:

1. *the $n \times nr$ controllability matrix*

$$\boxed{P_d = \left[B, AB, A^2B, \ldots, A^{n-1}B\right]}$$

   has rank n

2. *the controllability gramian*

$$\boxed{W_{cd} = \sum_{k=0}^{k_1} A^k BB^T \left(A^T\right)^k}$$

   is nonsingular for some finite k_1

Proof: from controllability matrix to gramian

Recall

$$\boxed{x(n) - A^n x(0) = \underbrace{\left[B, AB, A^2B, \ldots, A^{n-1}B\right]}_{P_d} \mathrm{u}_n} \qquad (2)$$

- P_d is full row rank$\Rightarrow P_d P_d^T = \underbrace{\sum_{k=0}^{n} A^k BB^T \left(A^T\right)^k}_{W_{cd} \text{ at } k_1 = n}$ is nonsingular
- a (least-square) solution to (2) is

$$\mathrm{u}_n = P_d^T \left(P_d P_d^T\right)^{-1} \left[x(n) - A^n x(0)\right]$$

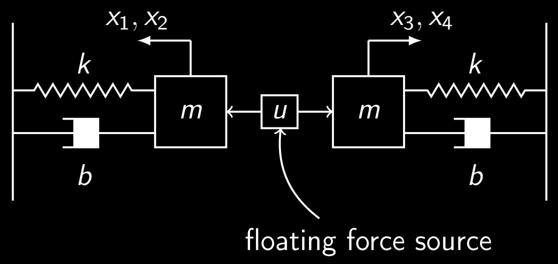
Example

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$P_d = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & \lambda_2 + \lambda_2 \\ 1 & \lambda_2 & \lambda_2^2 \end{bmatrix} \Rightarrow \text{rank}(P_d) = 2 < 3 \Rightarrow \text{uncontrollable}$$

Intuition: $\dot{x}_1 = \lambda_1 x_1$ is not impacted by the control input at all.

Example

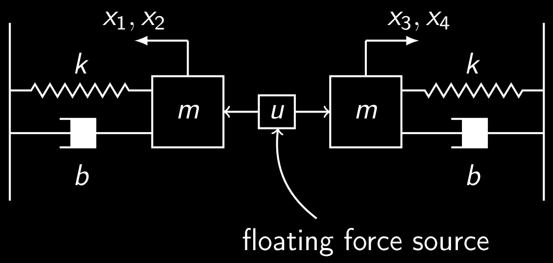


$$\dot{x}_1 = x_2$$
$$\dot{x}_3 = x_4$$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} 0.4 & 0.4 & 0 & 0 \\ -0.9 & -0.07 & 0 & 0 \\ 0 & 0 & 0.4 & 0.4 \\ 0 & 0 & -0.9 & -0.07 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.4 \\ 0.3 \\ 0.4 \end{bmatrix} u(k)$$

$$\text{rank}(P_d) = \text{rank} \left[\begin{array}{cccc} \overbrace{}^{B} & \overbrace{}^{AB} & \overbrace{}^{A^2B} & \overbrace{}^{A^3B} \\ 0.3 & 0.28 & -0.0072 & -0.0953 \\ 0.4 & -0.298 & -0.2311 & 0.0227 \\ 0.3 & 0.28 & -0.0072 & -0.0953 \\ 0.4 & -0.298 & -0.2311 & 0.0227 \end{array} \right] = 2 \Rightarrow \text{uncontrollable}$$

Example



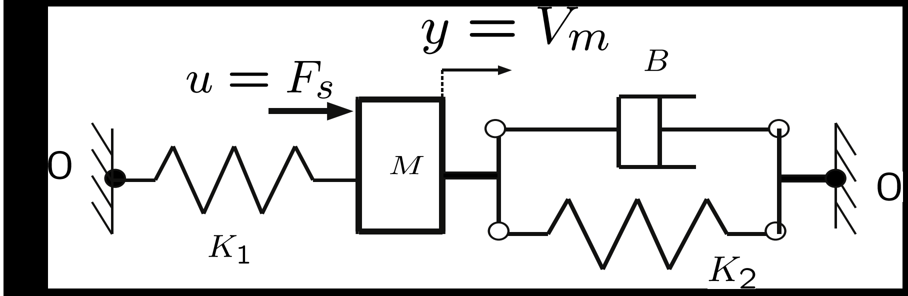
floating force source

$$\dot{x}_1 = x_2$$
$$\dot{x}_3 = x_4$$

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \\ x_4(k+1) \end{bmatrix} = \begin{bmatrix} 0.4 & 0.4 & 0 & 0 \\ -0.9 & -0.07 & 0 & 0 \\ 0 & 0 & 0.4 & 0.4 \\ 0 & 0 & -0.9 & -0.07 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \\ x_4(k) \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.4 \\ 0.3 \\ 0.4 \end{bmatrix} u(k)$$

```
import numpy as np
import control as ct
A = np.array([[0.4, 0.4, 0, 0], [-0.9, -0.07, 0, 0], [0, 0, 0.4, 0.4], [0, 0, -0.9,
-0.07]])
B = np.array([[0.3], [0.4], [0.3], [0.4]])
P = ct.ctrb(A,B)
print(np.linalg.matrix_rank(P))
```

Example



$$\frac{d}{dt}\begin{bmatrix} v_m \\ F_{k_1} \\ F_{k_2} \end{bmatrix} = \begin{bmatrix} -b/m & -1/m & -1/m \\ k_1 & 0 & 0 \\ k_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_m \\ F_{k_1} \\ F_{k_2} \end{bmatrix} + \begin{bmatrix} 1/m \\ 0 \\ 0 \end{bmatrix} F$$

$$P = \begin{bmatrix} 1/m & -b/m^2 & b^2/m^3 - k_1/m^2 - k_2/m^2 \\ 0 & k_1/m & -bk_1/m^2 \\ 0 & k_2/m & -bk_2/m^2 \end{bmatrix} \Rightarrow \text{rank}(P) = 2$$

$\Rightarrow$uncontrollable

Analysis: controllability and controllable canonical form

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix}, \; B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- controllability matrix

$$P_d = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -a_2 \\ 1 & -a_2 & -a_1 + a_2^2 \end{bmatrix}$$

  has full row rank
- *system in controllable canonical form is controllable*

Recap

General LTI state-space models:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \text{ or } x(k+1) = Ax(k) + Bu(k) \\ y &= Cx + Du \end{aligned}$$

	continuous time	discrete time
Lyapunov Eq.	$A^TP + PA = -Q$	$A^TPA - P = -Q$
unique sol. cond.	$\lambda_i(A) + \lambda_j(A) \neq 0$ $\forall\ i, j$	$\lvert\lambda_i(A)\rvert\,\lvert\lambda_j(A)\rvert < 1$ $\forall\ i, j$
solution	$P = \int_0^\infty e^{A^Tt}Qe^{At}dt$ (if A is Hurwitz stable)	$P = \sum_{k=0}^\infty \left(A^T\right)^k QA^k$ (if A is Schur stable)

Analysis: controllability gramian and Lyapunov Eq.

$$W_{cd} = \sum_{k=0}^{k_1} A^k B B^T \left(A^T\right)^k$$

- If A is Schur, k_1 can be set to ∞

$$W_{cd} = \sum_{k=0}^{\infty} A^k \underbrace{BB^T}_{Q} \left(A^T\right)^k$$

which can be solved via the Lyapunov Eq.

$$\boxed{A W_{cd} A^T - W_{cd} = -BB^T}$$

Analysis: controllability and similarity transformation

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) + Du(k) \end{cases} \xRightarrow{x=Tx^*} \begin{cases} x^*(k+1) = \overbrace{T^{-1}AT}^{\tilde{A}} x^*(k) + \overbrace{T^{-1}B}^{\tilde{B}} u(k) \\ y(k) = CTx^*(k) + Du(k) \end{cases}$$

- controllability matrix

$$\begin{aligned} P_d^* &= \left[\tilde{B}, \tilde{A}\tilde{B}, \dots, \tilde{A}^{n-1}\tilde{B}\right] \\ &= \left[T^{-1}B, T^{-1}AB, \dots, T^{-1}A^{n-1}B\right] = T^{-1}P_d \end{aligned}$$

hence $\underline{(A, B)\ \text{controllable} \Leftrightarrow \left(T^{-1}AT, T^{-1}B\right)\ \text{controllable}}$

- **The controllability property is invariant under any coordinate transformation.**

* Popov-Belevitch-Hautus (PBH) controllability test

- the full rank condition of the controllability matrix

$$\boxed{P_d = \left[B, AB, A^2B, \ldots, A^{n-1}B\right]}$$

  is equivalent to: *the matrix* $[A - \lambda I,\ B]$ *having full row rank at every eigenvalue,* λ*, of* A
- to see this: if $[A - \lambda I,\ B]$ is not full row rank then there exists nonzero vector (a left eigenvector) such that

$$v^T[A - \lambda I\ B] = 0$$

$$\Leftrightarrow v^T A = \lambda v^T$$
$$v^T B = 0$$

  i.e., the input vector B is orthogonal to a left eigenvector of A.

Example

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 1 \\ 0 & 0 & \lambda_2 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$\left[A - \lambda_1 I, \quad B\right] =$

$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_2 - \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_2 - \lambda_1 & 1 \end{bmatrix}$ not full row rank $\Rightarrow$uncontrollable

Intuition: $\dot{x}_1 = \lambda_1 x_1$ is not impacted by the control input at all.

1. Concepts
2. DT controllability
   - Controllability and controllable canonical form
   - Controllability and Lyapunov Eq.
3. DT observability
   - Observability and observable canonical form
4. CT cases
5. The degrees of controllability and observability
6. Transforming controllable systems into controllable canonical forms
7. Transforming observable systems into observable canonical forms

Observability of LTI systems

Definition

A discrete-time linear system

$$\begin{aligned} x(k+1) &= A(k)\,x(k) + B(k)\,u(k) \\ y(k) &= C(k)\,x(k) + D(k)\,u(k) \end{aligned}$$

is called observable at $k = 0$ if $\exists$ a finite time k_1 such that $\forall$ initial state $x(0)$, the knowledge of input $\{u(k); k = 0, 1, \ldots, k_1\}$ and $\{y(k); k = 0, 1, \ldots, k_1\}$ suffice to determine the state $x(0)$. Otherwise, the system is said to be unobservable at time $k = 0$.

Observability of LTI systems

let us start with the unforced system

$$\begin{aligned} x(k+1) &= Ax(k), \ A \in \mathbb{R}^n \\ y(k) &= Cx(k), \ y \in \mathbb{R}^m \end{aligned}$$

$x(k) = A^k x(0)$ and $y(k) = Cx(k)$ give

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n-1) \end{bmatrix}}_{\mathrm{Y}_n} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{Q_d : nm \times n} x(0)$$

- ▶ if the linear matrix equation has a nonzero solution $x(0)$, the system is observable

Observability of LTI systems

generalizing to
$x(k+1) = Ax(k) + Bu(k), \ y(k) = Cx(k) + Du(k)$:

$$x(k) = A^k x(0) + \sum_{j=0}^{k-1} A^{k-1-j} B u(j)$$

$$y(k) = \underbrace{CA^k x(0)}_{y_{\text{free}}(k)} + \underbrace{C \sum_{j=0}^{k-1} A^{k-1-j} B u(j) + Du(k)}_{y_{\text{forced}}(k)}$$

$$\underbrace{\begin{bmatrix} y(0) - y_{\text{forced}}(0) \\ y(1) - y_{\text{forced}}(1) \\ \vdots \\ y(n-1) - y_{\text{forced}}(n-1) \end{bmatrix}}_{\bar{\mathrm{Y}}_n\text{: available from measurements and inputs}} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{Q_d : nm \times n} x(0)$$

Observability of LTI systems

$$\underbrace{\begin{bmatrix} y(0) - y_{\text{forced}}(0) \\ y(1) - y_{\text{forced}}(1) \\ \vdots \\ y(n-1) - y_{\text{forced}}(n-1) \end{bmatrix}}_{\bar{y}_n} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{Q_d} x(0)$$

- $x(0)$ can be solved if Q_d has rank n (full column rank):
  - if Q_d is square, $x(0) = Q_d^{-1}\bar{y}_n$
  - if Q_d is a tall matrix, pick n linearly independent rows from Q_d

Observability of LTI systems Cont'd

$$\underbrace{\begin{bmatrix} y(0) - y_{\text{forced}}(0) \\ y(1) - y_{\text{forced}}(1) \\ \vdots \\ y(n-1) - y_{\text{forced}}(n-1) \end{bmatrix}}_{\bar{y}_n} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{Q_d} x(0)$$

- also, no need to go beyond n in Q_d: adding CA^n, CA^{n+1}, ... does not increase the column rank of Q_d (Cayley Halmilton Theorem)

Theorem (Observability Theorem)

System $x(k+1) = Ax(k) + Bu(k)\,,\; y(k) = Cx(k) + Du(k)$, $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times n}$ is observable if and only if either one of the following is satisfied:

1. *the observability matrix* $Q_d = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}_{(mn)\times n}$ *has full column rank*
2. *the observability gramian*

   $\boxed{W_{od} = \sum_{k=0}^{k_1} \left(A^T\right)^k C^T C A^k}$ *is nonsingular for some finite* k_1
3. ** PBF test: The matrix* $\begin{bmatrix} A - \lambda I \\ C \end{bmatrix}$ *has full column rank at every eigenvalue,* λ*, of* A*.*

Proof: from observability matrix to gramian

$$Q_d = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad W_{od} = \sum_{k=0}^{k_1} \left(A^T\right)^k C^T C A^k$$

- Q_d is full column rank$\Rightarrow Q_d^T Q_d = \underbrace{\sum_{k=0}^{n} \left(A^T\right)^k C^T C A^k}_{W_{od} \text{ at } k_1=n}$ is nonsingular

Observability check

- Analogous to the case in controllability, the observability property is invariant under any coordinate transformation:

$$(A, C) \text{ is observable} \iff (T^{-1}AT, CT) \text{ is observable}$$

- If A is Schur, k_1 can be set to ∞ in the observability gramian

$$W_{od} = \sum_{k=0}^{\infty} \left(A^T\right)^k C^T C A^k$$

and we can compute by solving the Lyapunov equation

$$A^T W_{od} A - W_{od} = -C^T C$$

The solution is nonsingular if and only if the system is observable. In fact, $W_{od} \succeq 0$ by definition $\Rightarrow$ "nonsingular" can be replaced with "positive definite".

Observability and observable canonical form

$$A = \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix}, \; C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

- observability matrix

$$Q_d = \begin{bmatrix} C \\ CA \\ CA^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -a_2 & 1 & 0 \\ a_2^2 - a_1 & -a_2 & 1 \end{bmatrix}$$

has full column rank

- *system in observable canonical form is observable*

* PBH test for observability

The matrix $\begin{bmatrix} A-\lambda I \\ C \end{bmatrix}$ has full column rank at every eigenvalue, λ, of A.

- if not full rank then there exists a nonzero eigenvector v:

$$\begin{aligned} Av &= \lambda v \\ Cv &= 0 \\ \Rightarrow CAv &= \lambda Cv = 0 \\ &\vdots \\ CA^{n-1}v &= 0 \end{aligned} \quad \Rightarrow \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} v = 0 \Rightarrow \text{unobservable}$$

- the reverse direction is analogous
- **interpretation**: some non-zero initial condition $x_0 = v$ will generate zero output, which is not distinguishable from the origin.

1. Concepts
2. DT controllability
   - Controllability and controllable canonical form
   - Controllability and Lyapunov Eq.
3. DT observability
   - Observability and observable canonical form
4. CT cases
5. The degrees of controllability and observability
6. Transforming controllable systems into controllable canonical forms
7. Transforming observable systems into observable canonical forms

Theorem (Controllability of continuous-time systems)

The n-dimensional r-input LTI system with $\dot{x} = Ax + Bu$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times r}$ is controllable if and only if either one of the following is satisfied

1. *the $n \times nr$ controllability matrix*

$$P = \left[B, AB, A^2B, \dots, A^{n-1}B\right]$$

   has rank n

2. *the controllability gramian*

$$\boxed{W_{cc} = \int_0^t e^{A\tau}BB^Te^{A^T\tau}d\tau}$$

   is nonsingular for any $t > 0$

Theorem (Observability of continuous-time systems)

System $\dot{x} = Ax + Bu$, $y = Cx + Du$, $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times n}$ is observable if and only if either one of the following is satisfied

1. *the $(mn) \times n$ observability matrix*

$$Q = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \text{ has rank } n \text{ (full column rank)}$$

2. *the observability gramian*

$$\boxed{W_{oc} = \int_0^t e^{A^T\tau}C^TCe^{A\tau}d\tau} \text{ is nonsingular for any } t > 0$$

Summary: computing the gramians

	Controllability Gramian	Observability Gramian
continuous time	$\int_0^t e^{A\tau}BB^T\left(e^{A\tau}\right)^T d\tau$	$\int_0^t \left(e^{A\tau}\right)^T C^T C e^{A\tau} d\tau$
Lyapunov eq. if $t \to \infty$ & A is Hurwitz stable	$AW_c + W_c A^T = -BB^T$	$A^T W_o + W_o A = -C^T C$
discrete time	$\sum_{k=0}^{k_1} A^k BB^T \left(A^T\right)^k$	$\sum_{k=0}^{k_1} (A^T)^k C^T C A^k$
Lyapunov eq. if $k_1 \to \infty$ & A is Schur stable	$AW_{cd}A^T - W_{cd} = -BB^T$	$A^T W_{od} A - W_{od} = -C^T C$

- duality: (A, B) is controllable if and only if $\left(\overline{A}, \overline{C}\right) = \left(A^T, B^T\right)$ is observable

Exercise

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 1 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix},\ B = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$$

- exercise: show that the system is not observable.
- in fact, by similarity transform $\overline{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} x$, we get

$$\bar{A} = \left[\begin{array}{cc|c} -2 & 0 & 0 \\ 0 & 0 & 0 \\ \hline 1 & 2 & 0 \end{array}\right],\ \bar{B} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\bar{C} = \left[\begin{array}{cc|c} 1 & 1 & 0 \end{array}\right]$$

where the third state is not observable

1. Concepts

2. DT controllability
   - Controllability and controllable canonical form
   - Controllability and Lyapunov Eq.

3. DT observability
   - Observability and observable canonical form

4. CT cases

5. The degrees of controllability and observability

6. Transforming controllable systems into controllable canonical forms

7. Transforming observable systems into observable canonical forms

The degree of controllability

consider two systems

$$S_1: \ x(k+1) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k)$$

$$S_2: \ x(k+1) = \begin{bmatrix} 0 & 0.01 \\ 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k)$$

- both systems are controllable:

$$P_{d_1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad P_{d_2} = \begin{bmatrix} 0 & 0.01 \\ 1 & 1 \end{bmatrix}$$

- however, P_{d_2} is nearly singular $\Rightarrow$ S_2 not “easy” to control
- e.g., to move from $x(0) = [0,0]^T$ to $x(1) = [1,1]^T$ in two steps:

$$S_1: \{u(0), u(1)\} = \{1, 1\} \qquad S_2: \{u(0), u(1)\} = \{100, -99\}$$

$\Rightarrow$ more energy for S_2!

The degree of controllability: multi-input case

consider two systems

$$S_1: \; x(k+1) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(k) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} u(k)$$

$$S_2: \; x(k+1) = \begin{bmatrix} 0 & 0.01 \\ 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} u(k)$$

- both systems are controllable:

$$P_{d_1} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \qquad P_{d_2} = \begin{bmatrix} 0 & 0.01 & 0.01 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

- degree of controllability reflected in the controllability Gramian:

$$W_{cd1} = P_{d1} P_{d1}^T = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \; W_{cd2} = \begin{bmatrix} 2 \times 0.01^2 & 0.02 \\ 0.02 & 3 \end{bmatrix}$$

W_{cd2} is almost singular (eigenvalues at 0.0001 and 3.0001)

The degree of controllability: multi-input case

- for general stable and controllable systems $\Sigma = (A, B, C, D)$, W_{cd} is computed from the Lyapunov Equation $AW_{cd}A^T - W_{cd} = -BB^T$
- if W_{cd} have eigenvalues close to zero, then the system is more difficult to control in the sense that it requires more energy in the input to steer the states in the state space

The degree of observability

consider two systems

$$S_1 : x(k+1) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(k) \qquad y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k)$$

$$S_2 : x(k+1) = \begin{bmatrix} 1 & 0.01 \\ 0 & 0 \end{bmatrix} x(k) \qquad y(k) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(k)$$

- both systems are observable:

$$Q_{d_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad Q_{d_2} = \begin{bmatrix} 1 & 0 \\ 1 & 0.01 \end{bmatrix}$$

- however, Q_{d_2} is nearly singular, hinting that S_2 is not “easy” to observe
- e.g., to infer $x(0) = [2, 1]^T$, the two measurements $y(0) = 2$ and $y(1) = CAx(0) = 2.001$ are nearly identical in S_2!

The degree of observability: multi-output case

- for general stable and controllable systems $\Sigma = (A, B, C, D)$, the observability matrix Q_d is not square
- the degree of observability is reflected in the eigenvalues of the observability Gramian W_{od}
- for stable systems, W_{od} is computed from the Lyapunov Equation $A^T W_{od} A - W_{od} = -C^T C$
- if W_{od} have eigenvalues close to zero, then the system is more difficult to observe

Balanced state-space realizations

we know now

- the controllability and observability Gramians represent the degrees of controllability and observability
- easily controllable systems may not be easily observable
- easily observable systems may not be easily controllable

$\Rightarrow$ there exists realizations that balance the two degrees of controllability and observability

Balanced state-space realizations

consider a stable system $\Sigma = (A, B, C, D)$ in a minimal[1] realization

- minimal realization $\Rightarrow$ Σ is controllable and observable
- stable $\Rightarrow$ can compute the Gramians from Lyapunov Equations
- if W_{cd} and W_{od} are equal and diagonal, then Σ is called a balanced realization
- i.e., there exists a diagonal matrix $M = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$ such that

$$M = AMA^T + BB^T$$
$$M = A^TMA + C^TC$$

[1] i.e., $\dim A$ is the minimal order of the system

1. Concepts
2. DT controllability
   - Controllability and controllable canonical form
   - Controllability and Lyapunov Eq.
3. DT observability
   - Observability and observable canonical form
4. CT cases
5. The degrees of controllability and observability
6. Transforming controllable systems into controllable canonical forms
7. Transforming observable systems into observable canonical forms

Transforming single-input controllable system into ccf

Let $x = M\tilde{x}$, where $M = \begin{bmatrix} | & | & & | \\ m_1 & m_2 & \dots & m_n \\ | & | & & | \end{bmatrix}$, then

$$\dot{\tilde{x}} = M^{-1}\dot{x} = M^{-1}\left(Ax + Bu\right) = M^{-1}AM\tilde{x} + \underbrace{M^{-1}B}_{\tilde{B}}u$$

If system is controllable, we show how to transform the state equation into the controllable canonical form.

- goal 1: $\tilde{B}$ be in controllable canonical form$\Leftrightarrow$

$$M^{-1}B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \Rightarrow B = [m_1, m_2, \dots, m_n]\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = m_n$$

Transforming SI controllable system into ccf

Let $x = M\tilde{x}$, where $M = [m_1, m_2, \ldots, m_n]$, then

$$\dot{\tilde{x}} = M^{-1}\dot{x} = M^{-1}(Ax + Bu) = \underbrace{M^{-1}AM}_{\tilde{A}}\tilde{x} + M^{-1}Bu$$

- goal 2: $\tilde{A}$ be in controllable canonical form$\Leftrightarrow$

$$A[m_1, m_2, \ldots, m_n] =$$

$$[m_1, m_2, \ldots, m_n]\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ 0 & \ldots & 0 & 0 & 1 \\ -a_0 & -a_1 & \ldots & \ldots & -a_{n-1} \end{bmatrix}$$

Transforming SI controllable system into ccf

Let $x = M\tilde{x}$, where $M = [m_1, m_2, \ldots, m_n]$, then

$$\dot{\tilde{x}} = M^{-1}\dot{x} = M^{-1}(Ax + Bu) = M^{-1}AM\tilde{x} + M^{-1}Bu$$

- solving goals 1 and 2 yields

$$\begin{aligned} m_n &= B \\ m_{n-1} &= Am_n + a_{n-1}m_n \\ m_{n-2} &= Am_{n-1} + a_{n-2}m_n \\ m_{i-1} &= Am_i + a_{i-1}m_n, \quad i = n, \ldots, 2 \\ &\vdots \end{aligned}$$

- when implementing, obtain a_0, a_1, $\ldots$, a_{n-1} first by calculating $\det(sI - A) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$

Transforming single-output (SO) observable system into ocf

Let $x = R^{-1}\tilde{x}$, where $R = \left[r_1^T, r_2^T, \dots, r_n^T\right]^T$ (r_i is a row vector).

$$\dot{\tilde{x}} = R\dot{x} = R\left(Ax + Bu\right) = \underbrace{RAR^{-1}}_{\tilde{A}}\tilde{x} + RBu$$

$$y = Cx = \underbrace{CR^{-1}}_{\tilde{C}}\tilde{x}$$

If system is observable, we show how to transform the state equation into the observable canonical form.

- goal 1: $\tilde{C}$ be in observable canonical form$\Leftrightarrow$

$$CR^{-1} = \begin{bmatrix}1 & 0 & \dots & 0\end{bmatrix} \Rightarrow C = r_1$$

Transforming SO observable system into ocf

Let $x = R^{-1}\tilde{x}$, where $R = \left[r_1^T, r_2^T, \dots, r_n^T\right]^T$ (r_i is a row vector).

$$\dot{\tilde{x}} = R\dot{x} = R\left(Ax + Bu\right) = \underbrace{RAR^{-1}}_{\tilde{A}}\tilde{x} + RBu$$

$$y = Cx = \underbrace{CR^{-1}}_{\tilde{C}}\tilde{x}$$

- goal 2: $\tilde{A}$ be in observable canonical form$\Leftrightarrow$

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} A = \begin{bmatrix} -a_{n-1} & 1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ & 0 & \ddots & \ddots & 0 \\ -a_1 & \vdots & \ddots & \ddots & 1 \\ -a_0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

Transforming SO observable system into ocf

Let $x = R^{-1}\tilde{x}$, where $R = \left[r_1^T, r_2^T, \dots, r_n^T\right]^T$ (r_i is a row vector).

$$\dot{\tilde{x}} = R\dot{x} = R\left(Ax + Bu\right) = \underbrace{RAR^{-1}}_{\tilde{A}}\tilde{x} + RBu$$

$$y = Cx = \underbrace{CR^{-1}}_{\tilde{C}}\tilde{x}$$

- solving goals 1 and 2 yields

$$\begin{aligned} r_1 &= C \\ r_2 &= r_1 A + a_{n-1} r_1 \\ r_3 &= r_2 A + a_{n-2} r_1 \\ r_{i+1} &= r_i A + a_{n-i} r_1, \;\; i = 1, \dots, n-1 \\ &\vdots \end{aligned}$$

- when implementing, obtain a_0, a_1, $\dots$, a_{n-1} first by calculating $\det\left(sI - A\right)$

Transforming SO observable system into ocf

Example: $x\left(k+1\right) = \begin{bmatrix} 1 & 0.01 \\ 0 & 0 \end{bmatrix} x\left(k\right) y\left(k\right) = \begin{bmatrix} 1 & 0 \end{bmatrix} x\left(k\right)$

$$\det\left(A - \lambda I\right) = \lambda^2 - \lambda \Rightarrow a_1 = -1, \;\; a_0 = 0$$

$$r_1 = C = [1, 0]$$

$$r_2 = r_1 C + a_1 r_1 = [1, 0]A + \left(-1\right)[1, 0]$$

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 0.01 \end{bmatrix}, R^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix}$$

$$\tilde{C} = CR^{-1} = [1, 0] \Longleftarrow \text{ocf!}$$

$$\tilde{A} = RAR^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \Longleftarrow \text{ocf!}$$

ME 547: Linear Systems

Controllable and Observable Subspaces Kalman Canonical Decomposition

Xu Chen

University of Washington

1. Controllable subspace
2. Observable subspace
3. Separating the uncontrollable subspace
   - Discrete-time version
   - Continuous-time version
   - Stabilizability
4. Separating the unobservable subspace
   - Discrete-time version
   - Detectability
   - Continuous-time version
5. Transfer-function perspective
6. Kalman decomposition

Controllable subspace: Introduction

Example

$$\bar{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \ \bar{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} x_1(k+1) & = x_1(k) + u(k) \\ x_2(k+1) & = 0 \end{cases}$$

$$\bar{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \bar{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} x_1(k+1) & = x_1(k) + x_2(k) + u(k) \\ x_2(k+1) & = x_2(k) \end{cases}$$

- there exists controllable and uncontrollable states: x_1 controllable and x_2 uncontrollable
- how to compute the dimensions of the two for general systems?
- how to separate them?

Controllable subspace: Assumptions

Consider an uncontrollable LTI system

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k), \ A \in \mathbb{R}^{n \times n} \\ y(k) &= Cx(k) + Du(k) \end{aligned}$$

Let the controllability matrix

$$P = \left[B, AB, A^2B, \dots, A^{n-1}B\right]$$

have rank $n_1 < n$.

Controllable subspace

- The controllable subspace χ_C is the set of all vectors $x \in \mathbb{R}^n$ that can be reached from the origin.
- From

$$x(n) - A^n x(0) = \underbrace{\left[B, AB, A^2B, \dots, A^{n-1}B\right]}_{P} \begin{bmatrix} u(n-1) \\ u(n-2) \\ \vdots \\ u(0) \end{bmatrix}$$

χ_C is the range space of P: $\chi_C = \mathcal{R}(P)$

1. Controllable subspace
2. Observable subspace
3. Separating the uncontrollable subspace
    - Discrete-time version
    - Continuous-time version
    - Stabilizability
4. Separating the unobservable subspace
    - Discrete-time version
    - Detectability
    - Continuous-time version
5. Transfer-function perspective
6. Kalman decomposition

Observable subspace: Introduction

Example

$$\bar{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \ \bar{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \Leftrightarrow \begin{cases} x_1(k+1) & = x_1(k) + u(k) \\ x_2(k+1) & = x_1(k) + x_2(k) \\ y(k) & = x_1(k) \end{cases}$$

$$\bar{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

- exists observable and unobservable states: x_1 observable and x_2 unobservable
- how to separate the two?
- how to separate controllable but observable states, controllable but unobservable states, etc?

Observable subspace: Assumptions

Consider an unobservable LTI system

$$x(k+1) = Ax(k) + Bu(k), \ A \in \mathbb{R}^{n\times n}$$
$$y(k) = Cx(k) + Du(k)$$

Let the observability matrix

$$Q = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

have rank $n_2 < n$.

Unobservable subspace

- The unobservable subspace χ_{uo} is the set of all nonzero initial conditions $x(0) \in \mathbb{R}^n$ that produce a zero free response.
- From

$$\underbrace{\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n-1) \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}_{Q} x(0)$$

χ_{uo} is the null space of Q: $\chi_{uo} = \mathcal{N}(Q)$

1. Controllable subspace

2. Observable subspace

3. Separating the uncontrollable subspace
   Discrete-time version
   Continuous-time version
   Stabilizability

4. Separating the unobservable subspace
   Discrete-time version
   Detectability
   Continuous-time version

5. Transfer-function perspective

6. Kalman decomposition

Separating the uncontrollable subspace

- recall 1: similarity transform $x = Mx^*$ preserves controllability

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) + Du(k) \end{cases} \Rightarrow \begin{cases} x^*(k+1) = M^{-1}AMx^*(k) + M^{-1}Bu(k) \\ y(k) = CMx^*(k) + Du(k) \end{cases}$$

- recall 2: the uncontrollable system structure at introduction

$$\bar{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \bar{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} x_1(k+1) & = x_1(k) + x_2(k) + u(k) \\ x_2(k+1) & = x_2(k) \end{cases}$$

- decoupled structure for generalized systems

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + Du(k)$$

$\bar{x}_{uc}$ impacted by neither u nor $\bar{x}_c$.

Theorem (Kalman canonical form (controllability))

Let $x \in \mathbb{R}^n$, $x(k+1) = Ax(k) + Bu(k)$, $y(k) = Cx(k) + Du(k)$ be uncontrollable with rank of the controllability matrix, $\operatorname{rank}(P) = n_1 < n$. Let $M = \begin{bmatrix} M_c & M_{uc} \end{bmatrix}$, where $M_c = [m_1, \ldots, m_{n_1}]$ consists of n_1 linearly independent columns of P, and $M_{uc} = [m_{n_1+1}, \ldots, m_n]$ are added columns to complete the basis and yield a nonsingular M. Then $x = M\bar{x}$ transforms the system equation to

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + Du(k)$$

Furthermore, $(\bar{A}_c, \bar{B}_c)$ is controllable, and

$$C(zI - A)^{-1}B + D = \bar{C}_c(zI - \bar{A}_c)^{-1}\bar{B}_c + D$$

Theorem (Kalman canonical form (controllability))

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \overbrace{\begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix}}^{M^{-1}B} u(k)$$

intuition: the "B" matrix after transformation

- columns of $B \in$ column space of P, which is equivalent to $\mathcal{R}(M_c)$
- columns of M_{uc} and M_c are linearly independent $\Rightarrow$ columns of $B \notin \mathcal{R}(M_{uc})$
- thus

$$B = \begin{bmatrix} M_c & M_{uc} \end{bmatrix} \begin{bmatrix} \overbrace{*}^{\text{denote as } \bar{B}_c} \\ 0 \end{bmatrix} \Rightarrow M^{-1}B = \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix}$$

Theorem (Kalman canonical form (controllability))

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \overbrace{\begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix}}^{M^{-1}AM} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u(k)$$

intuition: the "A" matrix after transformation

- range space of M_c is "A-invariant":

  $$\text{columns of } AM_c \in \left\{AB, A^2B, \dots, A^nB\right\} \in \mathcal{R}(M_c)$$

  where columns of $A^nB \in \mathcal{R}(P) = \mathcal{R}(M_c)$ ($\because$ Cayley Halmilton Thm)
- i.e., $AM_c = M_c\bar{A}_c$ for some $\bar{A}_c \Rightarrow$

$$A[M_c, M_{uc}] = [M_c, M_{uc}] \underbrace{\begin{bmatrix} \bar{A}_c & \overbrace{*}^{\triangleq \bar{A}_{\mathbf{12}}} \\ 0 & \overbrace{*}^{\triangleq \bar{A}_{uc}} \end{bmatrix}}_{\bar{A}} \Rightarrow M^{-1}AM = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix}$$

Theorem (Kalman canonical form (controllability))

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \overbrace{\begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix}}^{M^{-1}AM} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \overbrace{\begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix}}^{M^{-1}B} u(k)$$

$\underline{(\bar{A}_c, \bar{B}_c) \textit{ is controllable}}$

- controllability matrix after similarity transform

$$\begin{aligned} \bar{P} &= \left[\begin{array}{cccc|cc} \bar{B}_c & \bar{A}_c\bar{B}_c & \dots & \bar{A}_c^{n_1-1}\bar{B}_c & \dots & \bar{A}_c^{n-1}\bar{B}_c \\ 0 & 0 & \dots & 0 & \dots & 0 \end{array}\right] \\ &= \left[\begin{array}{c|ccc} \bar{P}_c & \bar{A}_c^{n_1}\bar{B}_c & \dots & \bar{A}_c^{n-1}\bar{B}_c \\ 0 & 0 & \dots & 0 \end{array}\right] \end{aligned}$$

- similarity transform does not change controllability$\Rightarrow$ rank$(\bar{P}) = $ rank$(P) = n_1$
- thus rank$(\bar{P}_c) = n_1 \Rightarrow (\bar{A}_c, \bar{B}_c)$ is controllable

Theorem (Kalman canonical form (controllability))

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + Du(k)$$

$\underline{C(zI-A)^{-1}B + D = \bar{C}_c(zI-\bar{A}_c)^{-1}\bar{B}_c + D}$

we can check that

$$\begin{aligned} &\begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} zI-\bar{A}_c & -\bar{A}_{12} \\ 0 & zI-\bar{A}_{uc} \end{bmatrix}^{-1} \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} + D \\ =&\begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} \left(zI-\bar{A}_c\right)^{-1} & * \\ 0 & \left(zI-\bar{A}_{uc}\right)^{-1} \end{bmatrix} \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} + D \\ =&\bar{C}_c\left(zI-\bar{A}_c\right)^{-1}\bar{B}_c + D \end{aligned}$$

Matlab commands

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \overbrace{\begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix}}^{M^{-1}AM} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \overbrace{\begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix}}^{M^{-1}B} u(k)$$

$x = M\bar{x}$ where $M = \begin{bmatrix} M_c & M_{uc} \end{bmatrix}$

- $M_c = [m_1, \dots, m_{n_1}]$ consists of all the linearly independent columns of P: **Mc = orth(P)**
- $M_{uc} = [m_{n_1+1}, \dots, m_n]$ are added columns to complete the basis and yield a nonsingular M
  - from linear algebra: the orthogonal complement of the range space of P is the null space of P^T:

    $$\mathbb{R}^n = \mathcal{R}(P) \oplus \mathcal{N}\left(P^T\right)$$

  - hence **Muc = null(P')** (the transpose is important here)

The techniques apply to CT systems

Theorem (Kalman canonical form (controllability))

Let a n-dimensional state-space system $\dot{x} = Ax + Bu$, $y = Cx + Du$ be uncontrollable with the rank of the controllability matrix $rank(P) = n_1 < n$. Let $M = \begin{bmatrix} M_c & M_{uc} \end{bmatrix}$ where $M_c = [m_1, \dots, m_{n_1}]$ consists of n_1 linearly independent columns of P, $M_{uc} = [m_{n_1+1}, \dots, m_n]$ are added columns to complete the basis for $\mathbb{R}^n$ and yield a nonsingular M. Then the similarity transformation $x = M\bar{x}$ transforms the system equation to

$$\frac{d}{dt}\begin{bmatrix} \bar{x}_c \\ \bar{x}_{uc} \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c \\ \bar{x}_{uc} \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u$$

$$y = \begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c \\ \bar{x}_{uc} \end{bmatrix} + Du$$

Example

$$\frac{d}{dt}\begin{bmatrix} v_m \\ F_{k_1} \\ F_{k_2} \end{bmatrix} = \begin{bmatrix} -b/m & -1/m & -1/m \\ k_1 & 0 & 0 \\ k_2 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_m \\ F_{k_1} \\ F_{k_2} \end{bmatrix} + \begin{bmatrix} 1/m \\ 0 \\ 0 \end{bmatrix} F$$

Let $m = 1, b = 1$

$$P = \begin{bmatrix} 1 & -1 & 1-k_1-k_2 \\ 0 & k_1 & -k_1 \\ 0 & k_2 & -k_2 \end{bmatrix}, \; M = \begin{bmatrix} 1 & -1 & 0 \\ 0 & k_1 & 0 \\ 0 & k_2 & 1 \end{bmatrix}, \; M^{-1} = \begin{bmatrix} 1 & 1/k_1 & 0 \\ 0 & 1/k_1 & 0 \\ 0 & -k_2/k_1 & 1 \end{bmatrix}$$

$$\bar{A} = M^{-1}AM = \left[\begin{array}{cc|c} 0 & -(k_1+k_2) & 1 \\ 1 & -1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right], \; \bar{B} = M^{-1}B = \left[\begin{array}{c} 1 \\ 0 \\ \hline 0 \end{array}\right]$$

Stabilizability

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix}\begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix}\begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + Du(k)$$

The system is *stabilizable* if

- all its unstable modes, if any, are controllable
- i.e., the uncontrollable modes are stable ($\bar{A}_{uc}$ is Schur, namely, all eigenvalues are in the unit circle)

1. Controllable subspace

2. Observable subspace

3. Separating the uncontrollable subspace
   - Discrete-time version
   - Continuous-time version
   - Stabilizability

4. Separating the unobservable subspace
   - Discrete-time version
   - Detectability
   - Continuous-time version

5. Transfer-function perspective

6. Kalman decomposition

Separating the unobservable subspace

- recall 1: similarity transform $x = O^{-1}x^*$ preserves observability

$$\begin{cases} x\left(k+1\right) = Ax\left(k\right) + Bu\left(k\right) \\ y\left(k\right) = Cx\left(k\right) + Du\left(k\right) \end{cases} \Rightarrow \begin{cases} x^*\left(k+1\right) = OAO^{-1}x^*\left(k\right) + OBu\left(k\right) \\ y\left(k\right) = CO^{-1}x^*\left(k\right) + Du\left(k\right) \end{cases}$$

- an unobservable system structure

$$\bar{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \ \bar{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Leftrightarrow \begin{cases} x_1(k+1) & = x_1(k) + u(k) \\ x_2(k+1) & = x_1(k) + x_2(k) \\ y(k) & = x_1(k) \end{cases}$$

$$\bar{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

- decoupled structure for generalized systems

$$\begin{bmatrix} \bar{x}_o(k+1) \\ \bar{x}_{uo}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_o & 0 \\ \bar{A}_{21} & \bar{A}_{uo} \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_o \\ \bar{B}_{uo} \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_o & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + Du(k)$$

the "observed" $\bar{x}_o$ doesn't reflect $\bar{x}_{uc}$ $\left(\bar{x}_o(k+1) = \bar{A}_o\bar{x}_o\left(k\right) + \bar{B}_o u\left(k\right)\right)$

Theorem (Kalman canonical form (observability))

Let $x \in \mathbb{R}^n$, $x(k+1) = Ax(k) + Bu(k)$, $y(k) = Cx(k) + Du(k)$ be unobservable with rank of the observability matrix, $\operatorname{rank}(Q) = n_2 < n$. Let $O = \begin{bmatrix} O_o \\ O_{uo} \end{bmatrix}$ where O_o consists of n_2 linearly independent rows of Q, and $O_{uo} = \left[o^T_{n_1+1}, \ldots, o^T_n\right]^T$ are added rows to complete the basis and yield a nonsingular O. Then $\bar{x} = Ox$ transforms the system equation to

$$\begin{bmatrix} \bar{x}_o(k+1) \\ \bar{x}_{uo}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_o & 0 \\ \bar{A}_{21} & \bar{A}_{uo} \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_o \\ \bar{B}_{uo} \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_o & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + Du(k)$$

Furthermore, $(\bar{A}_o, \bar{O}_o)$ is observable, and

$$C(zI - A)^{-1}B + D = \bar{C}_o(zI - \bar{A}_o)^{-1}\bar{B}_o + D$$

Theorem (Kalman canonical form)

Case for observability

$$\begin{bmatrix} \bar{x}_o(k+1) \\ \bar{x}_{uo}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_o & 0 \\ \bar{A}_{21} & \bar{A}_{uo} \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_o \\ \bar{B}_{uo} \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_o & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + Du(k)$$

v.s. case for controllability

$$\begin{bmatrix} \bar{x}_c(k+1) \\ \bar{x}_{uc}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_c & \bar{A}_{12} \\ 0 & \bar{A}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_c \\ 0 \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_c & \bar{C}_{uc} \end{bmatrix} \begin{bmatrix} \bar{x}_c(k) \\ \bar{x}_{uc}(k) \end{bmatrix} + Du(k)$$

Intuition: duality between controllability and observability

$$(A, B) \text{ unconrollable} \Leftrightarrow \left(A^T, B^T\right) \text{ unobservable}$$

Detectability

$$\begin{bmatrix} \bar{x}_o(k+1) \\ \bar{x}_{uo}(k+1) \end{bmatrix} = \begin{bmatrix} \bar{A}_o & 0 \\ \bar{A}_{21} & \bar{A}_{uo} \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + \begin{bmatrix} \bar{B}_o \\ \bar{B}_{uo} \end{bmatrix} u(k)$$

$$y(k) = \begin{bmatrix} \bar{C}_o & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_o(k) \\ \bar{x}_{uo}(k) \end{bmatrix} + Du(k)$$

The system is *detectable* if

- all its unstable modes, if any, are observable
- i.e., the unobservable modes are stable ($\bar{A}_{uo}$ is Schur)

Continuout-time version

Theorem (Kalman canonical form (observability))

Let a n-dimensional state-space system $\dot{x} = Ax + Bu$, $y = Cx + Du$ be unobservable with the rank of the observability matrix $\text{rank}\,(Q) = n_2 < n$. Then there exists similarity transform $\bar{x} = Ox$ that transforms the system equation to

$$\frac{d}{dt} \begin{bmatrix} \bar{x}_o \\ \bar{x}_{uo} \end{bmatrix} = \begin{bmatrix} \bar{A}_o & 0 \\ \bar{A}_{21} & \bar{A}_{uo} \end{bmatrix} \begin{bmatrix} \bar{x}_o \\ \bar{x}_{uo} \end{bmatrix} + \begin{bmatrix} \bar{B}_o \\ \bar{B}_{uo} \end{bmatrix} u$$

$$y = \begin{bmatrix} \bar{C}_o & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_o \\ \bar{x}_{uo} \end{bmatrix} + Du$$

Furthermore, $(\bar{A}_o, \bar{C}_o)$ is observable, and $C(sI - A)^{-1}B + D = \bar{C}_o(sI - \bar{A}_o)^{-1}\bar{B}_o + D$.

1. Controllable subspace
2. Observable subspace
3. Separating the uncontrollable subspace
   - Discrete-time version
   - Continuous-time version
   - Stabilizability
4. Separating the unobservable subspace
   - Discrete-time version
   - Detectability
   - Continuous-time version
5. Transfer-function perspective
6. Kalman decomposition

Transfer-function perspective

$$\text{uncontrollable system: } C(zI-A)^{-1}B+D=\bar{C}_c(zI-\bar{A}_c)^{-1}\bar{B}_c+D$$

$$\text{unobservable system: } C(zI-A)^{-1}B+D=\bar{C}_o(zI-\bar{A}_o)^{-1}\bar{B}_o+D$$

where $A\in\mathbb{R}^{n\times n}$, $\bar{A}_c\in\mathbb{R}^{n_1\times n_1}$, $\bar{A}_o\in\mathbb{R}^{n_2\times n_2}$

- Order reduction exists

$$G(z)=C(zI-A)^{-1}B+D=\frac{B(z)}{A(z)},\ A(z)=\det\left(zI-A\right)\ \text{order}:n$$

$$G(z)=\bar{C}_c(zI-\bar{A}_c)^{-1}\bar{B}_c+D=\frac{\bar{B}_c(z)}{\bar{A}_c(z)},\ \bar{A}_c(z)=\det\left(zI-\bar{A}_c\right)\ \text{order}:n_1$$

- **$\Rightarrow A(z)$ and $B(z)$ are not co-prime | pole-zero cancellation exists**
- same applies to unobservable systems

Example

Consider

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u$$
$$y = \begin{bmatrix} c_1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- The transfer function is

$$G(s) = \frac{s + c_1}{s^2 + 3s + 2} = \frac{s + c_1}{(s+1)(s+2)}$$

- System is in controllable canonical form and is controllable.
- observability matrix

$$Q = \begin{bmatrix} c_1 & 1 \\ -2 & c_1 - 3 \end{bmatrix}, \ \det Q = (c_1 - 1)(c_1 - 2)$$

$\Rightarrow$unobservable if $c_1 = 1$ or 2

1. Controllable subspace
2. Observable subspace
3. Separating the uncontrollable subspace
   - Discrete-time version
   - Continuous-time version
   - Stabilizability
4. Separating the unobservable subspace
   - Discrete-time version
   - Detectability
   - Continuous-time version
5. Transfer-function perspective
6. Kalman decomposition

Kalman decomposition

an extended example:

$$A = \left[\begin{array}{c|c|c|c} A_{11} & 0 & A_{13} & 0 \\ \hline A_{21} & A_{22} & A_{23} & A_{24} \\ \hline 0 & 0 & A_{33} & 0 \\ 0 & 0 & A_{43} & A_{44} \end{array}\right], \quad B = \left[\begin{array}{c} B_1 \\ B_2 \\ \hline 0 \\ 0 \end{array}\right]$$

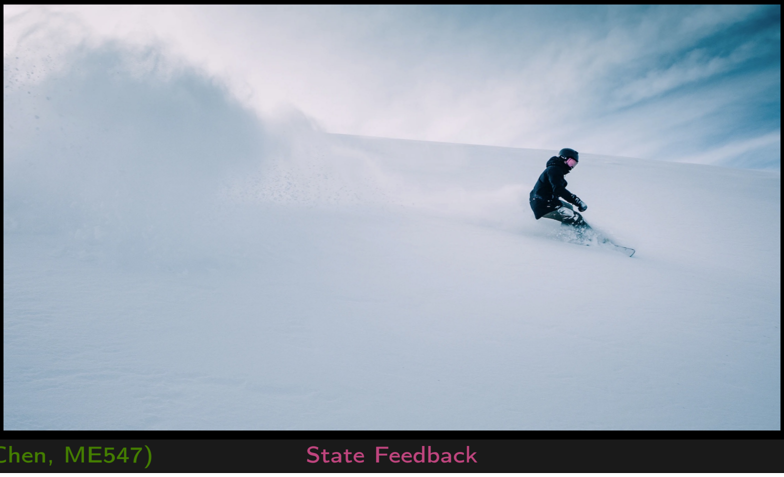
$$C = \left[\begin{array}{cccc} C_1 & 0 & C_3 & 0 \end{array}\right]$$

- A_{ij}, C_i and B_i are nonzero
- The A_{11} mode is controllable and observable. The A_{22} mode is controllable but not observable. The A_{33} mode is not controllable but observable. The A_{44} mode is not controllable and not observable.

State Feedback Control

Xu Chen

University of Washington



Motivation

- At the center of designing control systems is the idea of feedback.
- In such transfer-function approaches as lead-lag and root locus methods, the primal goal is to achieve a proper map of closed-loop poles with output feedback.

Key questions:

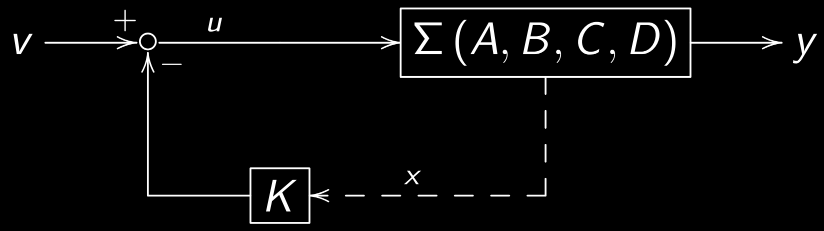
- How much freedom do we have for state-space systems?
- Are there fundamental system properties that yield higher achievable performance?
- How to implement the design algorithms?

General feedback structure

Consider an n-dimensional state-space system

$$\Sigma : \begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad x(t_0) = x_0$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.

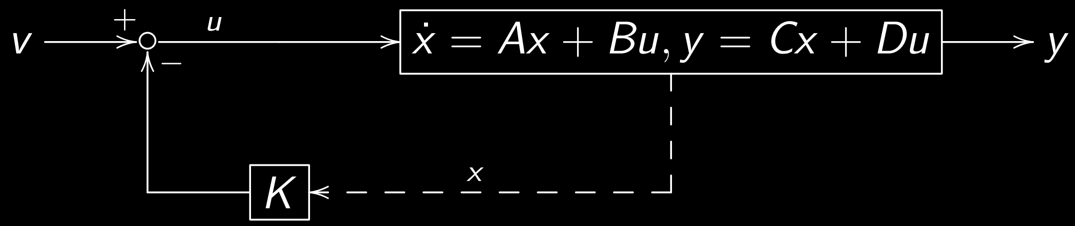


state-feedback law:

$$u = -Kx + v \qquad (1)$$

- v: new input
- $K \in \mathbb{R}^{m \times n}$: n-number of states, m-number of inputs

Goal



- closed-loop system:

$$\Sigma_{cl} : \begin{cases} \dot{x}(t) = (A - BK)\,x(t) + Bv(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad x(t_0) = x_0 \qquad (2)$$

- key closed-loop property: eigenvalues of $A - BK$.
- How freely can we place the eigenvalues of $A_{cl} = A - BK$?

Eigenvalue placement by state feedback

Fact: If $\Sigma = (A, B, C, D)$ is in controllable canonical form, we can completely change all the eigenvalues of $A - BK$ by choice of state-feedback gain matrix K.

- Problem setup: single-input single-output system in c.c.f.

$$H(s) = \frac{\beta_{n-1}s^{n-1} + \cdots + \beta_1 s + \beta_0}{s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0} + d, \quad \Sigma = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -\alpha_0 & \dots & \dots & -\alpha_{n-2} & -\alpha_{n-1} \end{bmatrix}, \ B = \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$$C = \begin{bmatrix} \beta_0 & \beta_1 & \dots & \dots & \beta_{n-1} \end{bmatrix}, \ D = d$$

$$\det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0 \tag{3}$$

Eigenvalue placement by state feedback: c.c.f.

- Goal: achieve desired closed-loop eigenvalue locations $p_1, \cdots, p_n$, i.e.

$$\det(sI - (A - BK)) = (s - p_1)(s - p_2)\cdots(s - p_n) \tag{4}$$

$$= s^n + \gamma_{n-1}s^{n-1} + \cdots + \gamma_1 s + \gamma_0 \tag{5}$$

- Let $K = [k_0, k_1, \dots, k_{n-1}]$. The structured A and B give

$$BK = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} [k_0, k_1, \dots, k_{n-1}] = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \vdots \\ \vdots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 0 \\ k_0 & \dots & \dots & k_{n-2} & k_{n-1} \end{bmatrix}$$

$$A - BK = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \vdots \\ \vdots & \dots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & 1 \\ -\alpha_0 - k_0 & \dots & \dots & -\alpha_{n-2} - k_{n-2} & -\alpha_{n-1} - k_{n-1} \end{bmatrix}$$

Eigenvalue placement by state feedback: c.c.f.

A
$\begin{bmatrix} 0 & 1 & 0 & \dots \\ \vdots & \ddots & 1 & \vdots \\ 0 & \dots & 0 & 1 \\ -\alpha_0 & \dots & \dots & -\alpha_{n-1} \end{bmatrix}$
$\det\left(sI - A\right)$
$s^n + \alpha_{n-1}s^{n-1} + \dots + \alpha_1 s + \alpha_0$

Eigenvalue placement by state feedback: c.c.f.

A	$A - BK$
$\begin{bmatrix} 0 & 1 & 0 & \dots \\ \vdots & \ddots & 1 & \vdots \\ 0 & \dots & 0 & 1 \\ -\alpha_0 & \dots & \dots & -\alpha_{n-1} \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ -\alpha_0 - k_0 & \dots & \dots & -\alpha_{n-1} - k_{n-1} \end{bmatrix}$
$\det\left(sI - A\right)$	
$s^n + \alpha_{n-1}s^{n-1} + \dots + \alpha_1 s + \alpha_0$	

Eigenvalue placement by state feedback: c.c.f.

A	$A-BK$
$\begin{bmatrix} 0 & 1 & 0 & \dots \\ \vdots & \ddots & 1 & \vdots \\ 0 & \dots & 0 & 1 \\ -\alpha_0 & \dots & \dots & -\alpha_{n-1} \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \\ -\alpha_0-k_0 & \dots & \dots & -\alpha_{n-1}-k_{n-1} \end{bmatrix}$
$\det\left(sI-A\right)$	$\det\left(sI-(A-BK)\right)$
$s^n+\alpha_{n-1}s^{n-1}+\cdots+\alpha_1 s+\alpha_0$	$s^n+\left(\alpha_{n-1}+k_{n-1}\right)s^{n-1}+\cdots+\left(\alpha_0+k_0\right)$

Eigenvalue placement by state feedback: c.c.f.

Goal (recap): achieve desired closed-loop eigenvalue locations $p_1, \cdots, p_n$, i.e.

$$\begin{aligned}\det\left(sI-(A-BK)\right) &= (s-p_1)(s-p_2)\cdots(s-p_n)\\ &= s^n+\gamma_{n-1}s^{n-1}+\cdots+\gamma_1 s+\gamma_0\end{aligned}$$

$$\det\left(sI-(A-BK)\right)=s^n+\underbrace{\left(\alpha_{n-1}+k_{n-1}\right)}_{\text{target: }\gamma_{n-1}}s^{n-1}+\cdots+\underbrace{\left(\alpha_0+k_0\right)}_{\text{target: }\gamma_0}$$

- hence

$$\begin{gathered}k_0=\gamma_0-\alpha_0\\ \vdots\\ k_{n-1}=\gamma_{n-1}-\alpha_{n-1}\end{gathered}$$

Eigenvalue placement by state feedback: c.c.f.

Eigenvalue-placement Algorithm

```
1  determine desired eigenvalue locations $p_1, \cdots, p_n$
2  calculate desired closed-loop characteristic polynomial
       $(s-p_1)(s-p_2)\cdots(s-p_n) = s^n + \gamma_{n-1}s^{n-1} + \cdots + \gamma_1 s + \gamma_0$
3  calculate open-loop characteristic polynomial
       $\det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$
4  define the matrices:
   $K = [\gamma_0 - \alpha_0, \ldots, \gamma_{n-1} - \alpha_{n-1}]$
```

Powerful result: if the system is in controllable canonical form, we can arbitrarily place the closed-loop eigenvalues by state feedback!

General eigenvalue placement by state feedback

- What if the given state-space realization $\Sigma = (A, B, C, D)$ is not in the required form?
- We can then transform it to c.c.f. via a similarity transformation.
- **Powerful fact**: if system $\Sigma = (A, B, C, D)$ is controllable, then we can arbitrarily place the closed-loop eigenvalues via state feedback.

Discrete-time case

- the eigenvalue assignment of discrete-time systems is analogous:
  - system dynamics:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) \end{aligned}$$

  - controller: $u(k) = -Kx(k) + v(k)$
  - closed-loop dynamics:

$$x(k+1) = Ax(k) - BKx(k) + Bv(k) = (A - BK)x(k) + Bv(k)$$

- arbitrary closed-loop eigenvalue assignment if system is controllable

Numerical example

$$\begin{aligned} x(k+1) &= \begin{bmatrix} 1 & 1 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} u(k) \\ y(k) &= \begin{bmatrix} 2 & 0 & 0 \end{bmatrix} x(k) \end{aligned}$$

```
%MATLAB
A = [1,1,-2;0,1,1;0,0,1];
B = [1;0;1];
p = [0;0.1;0.2];
K = place(A, B, p)
```

```
#Python
import control as ct
import numpy as np
A = np.array([[1,1,-2],[0,1,1],[0,0,1]])
B = np.array([[1],[0],[1]])
p = [0,0.1,0.2]
K = ct.place(A, B, p)
print(K)
```

The case with output feedback

- if the full state is not measurable, state feedback control is not feasible
- consider output feedback

$$\begin{cases} \dot{x} &= Ax + Bu \\ y &= Cx \\ u &= -Fy + v \end{cases} \Rightarrow \dot{x} = Ax - BFy + Bv = (A - BFC)\,x + Bv$$

- $A - BFC$ not as structured as $A - BK$
- arbitrary closed-loop eigenvalue assignment not feasible

The case with output feedback

Example

Controllable mass-spring-damper system

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u$$

$$\overset{u^* \triangleq \frac{u}{m}}{=} \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u^*$$

- arbitrary closed-loop eigenvalue assignment if $u^* = -k_1 x_1 - k_2 x_2$, namely $U^*(s) = -k_1 X_1(s) - k_2 X_2(s) = -(k_1 + k_2 s)\, X_1(s) \Rightarrow$ a proportional plus derivative (PD) control law
- if with only proportional control, $u^* = -k_1 x_1$, arbitrary closed-loop eigenvalue assignment is not possible

Observer and Observer State Feedback

Xu Chen

University of Washington

Introduction

- full state feedback is usually not available
- the state estimation problem
  - deterministic case: observer design
  - stochastic case: the most frequent option is Kalman filter

Open-loop observer

$$\frac{d}{dt}x(t) = Ax(t) + Bu(t),\ x(k+1) = Ax(k) + Bu(k)$$

- conceptually simplest scheme to estimate x:

$$\frac{d}{dt}\hat{x}(t) = A\hat{x}(t) + Bu(t),\ \hat{x}(k+1) = A\hat{x}(k) + Bu(k)$$

  with a best guess of initial estimate $\hat{x}(0) \overset{e.g.}{=} 0$.
- error dynamics: $e = x - \hat{x}$:

$$\dot{e}(t) = Ae(t),\ e(k+1) = Ae(k),\ e(0) = x_0 - \hat{x}(0)$$

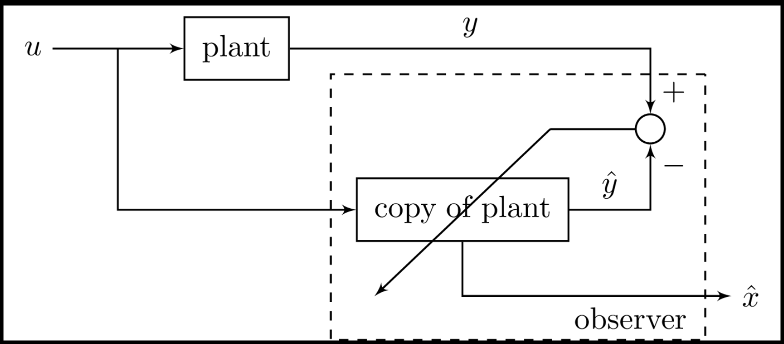
  - sensitive to input disturbances
  - if A is not Hurwitz/Schur stable, the error diverges
- open-loop observers look simple but do not work in practice

Luenberger (closed-loop) observer concept

- given system dynamics

$$\dot{x} = Ax + Bu,\ x(0) = x_0,\ A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times r}$$
$$y = Cx,\ y \in \mathbb{R}^{m\times n}$$

- in contrast to open-loop observers, the Luenberger observer adds correction based on output differences

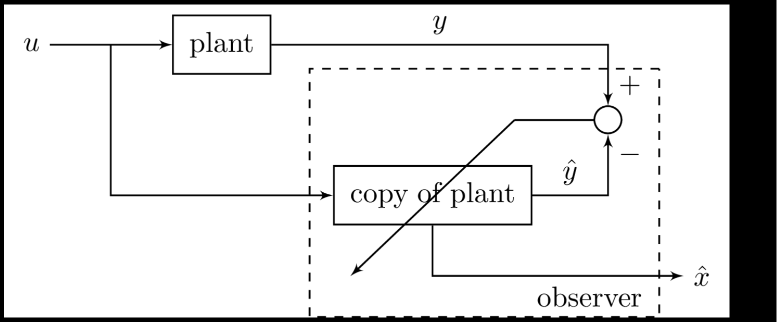


Luenberger (closed-loop) observer algorithm

plant:

$$\dot{x} = Ax + Bu, \; x(0) = x_0$$
$$y = Cx$$

observer concept



- observer realization:

$$\dot{\hat{x}} = A\hat{x} + Bu + L\left(y - \hat{y}\right) = A\hat{x} + Bu + L\left(y - C\hat{x}\right), \; \hat{x}(0) = 0$$
$$= \left(A - LC\right)\hat{x} + Ly + Bu$$

Luenberger (closed-loop) observer error dynamics

- system dynamics

$$\dot{x} = Ax + Bu, \; x(0) = x_0, \; A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times r}$$
$$y = Cx, \; y \in \mathbb{R}^{m\times n}$$

- Luenberger observer with correction:

$$\dot{\hat{x}} = A\hat{x} + Bu + L\left(y - \hat{y}\right) = A\hat{x} + Bu + L\left(y - C\hat{x}\right), \; \hat{x}(0) = 0$$
$$= \left(A - LC\right)\hat{x} + Ly + Bu$$

- error dynamics: $e = x - \hat{x}$:

$$\dot{e} = Ae - LCe = (A - LC)e, \; e(0) = x(0)$$

- if all eigenvalues of $A - LC$ are on the left half plane, then the error dynamics can be made asymptotically stable

Luenberger (closed-loop) observer

Theorem

If (A, C) is an observable pair, then all the eigenvalues of $A - LC$ can be arbitrarily assigned, provided that they are symmetric with respct to the real axis of the complex plane.

- we show the SISO case when A and C are in observable canonical form (if not, a similarity transform can help out):

$$A = \begin{bmatrix} -\alpha_{n-1} & 1 & 0 & \dots \\ \vdots & 0 & \ddots & \ddots \\ -\alpha_1 & \vdots & \ddots & 1 \\ -\alpha_0 & 0 & \dots & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_{n-1} \\ \vdots \\ \beta_1 \\ \beta_0 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \end{bmatrix}, \quad D = d$$

$$\det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$$

Observer eigenvalue placement: o.c.f.

- Luenberger observer with correction:

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu + L(y - \hat{y}) = A\hat{x} + Bu + L(y - C\hat{x}), \ \hat{x}(0) = 0 \\ &= (A - LC)\hat{x} + Ly + Bu \end{aligned}$$

- Goal: place eigenvalues of the observer at locations $\bar{p}_1, \cdots, \bar{p}_n$:

$$\begin{aligned} \det(sI - (A - LC)) &= (s - \bar{p}_1)(s - \bar{p}_2)\cdots(s - \bar{p}_n) \\ &= s^n + \bar{\gamma}_{n-1}s^{n-1} + \cdots + \bar{\gamma}_1 s + \bar{\gamma}_0 \end{aligned}$$

Observer eigenvalue placement: o.c.f.

- Goal: place eigenvalues of the observer at locations $\bar{p}_1, \cdots, \bar{p}_n$:

$$\det(sI - (A - LC)) = (s - \overline{p}_1)(s - \overline{p}_2)\cdots(s - \overline{p}_n)$$
$$= s^n + \overline{\gamma}_{n-1}s^{n-1} + \cdots + \overline{\gamma}_1 s + \overline{\gamma}_0$$

- Let $L = [l_0, l_1, \ldots, l_{n-1}]^T$. The unique structures of A and C give

$$LC = \begin{bmatrix} l_0 \\ \vdots \\ l_{n-2} \\ l_{n-1} \end{bmatrix} \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix} = \begin{bmatrix} l_0 & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & \vdots \\ l_{n-2} & \ddots & \ddots & 0 \\ l_{n-1} & 0 & \ldots & 0 \end{bmatrix}$$

$$A - LC = \begin{bmatrix} -\alpha_{n-1} - l_0 & 1 & 0 & \ldots & 0 \\ -\alpha_{n-2} - l_1 & 0 & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ -\alpha_1 - l_{n-2} & \vdots & \ddots & 0 & 1 \\ -\alpha_0 - l_{n-1} & 0 & \ldots & 0 & 0 \end{bmatrix}$$

Observer eigenvalue placement: o.c.f.

- A and $A - LC$ have the same structure:

$$A = \begin{bmatrix} -\alpha_{n-1} & 1 & 0 & \ldots \\ \vdots & 0 & \ddots & \ddots \\ -\alpha_1 & \vdots & \ddots & 1 \\ -\alpha_0 & 0 & \ldots & 0 \end{bmatrix}, \quad A - LC = \begin{bmatrix} -\alpha_{n-1} - l_0 & 1 & 0 & \ldots \\ \vdots & 0 & \ddots & \ddots \\ -\alpha_1 - l_{n-2} & \vdots & \ddots & 1 \\ -\alpha_0 - l_{n-1} & 0 & \ldots & 0 \end{bmatrix}$$

- Recall: $\det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$.
- Thus

$$\det(sI - (A - LC)) = s^n + \underbrace{(\alpha_{n-1} + l_0)}_{\text{target: } \overline{\gamma}_{n-1}} s^{n-1} + \cdots + \underbrace{(\alpha_0 + l_{n-1})}_{\text{target: } \overline{\gamma}_0}$$

- Hence

$$l_0 = \overline{\gamma}_{n-1} - \alpha_{n-1}$$
$$\vdots$$
$$l_{n-1} = \overline{\gamma}_0 - \alpha_0$$

General observer eigenvalue placement

- What if (A, B, C, D) is not in the observable canonical form?
- We can transform it to o.c.f. via a similarity transform:

$$\begin{cases}\dot{x} = Ax + Bu \\ y = Cx\end{cases} \quad \xrightarrow{x=R^{-1}x_{ob}} \quad \begin{cases}\dot{x}_{ob} &= \underbrace{RAR^{-1}}_{A_o} x_{ob} + \underbrace{RB}_{B_o} u \\ y &= C_o x_{ob} = CR^{-1}x_{ob}\end{cases}$$

- use previous formulas to design $\tilde{L}$ in:

$$\dot{\hat{x}}_{ob} = \left(A_o - \tilde{L}C_o\right)\hat{x}_{ob} + \tilde{L}y + B_o u \qquad \text{(analysis form)}$$

  correspondingly in the original state space (via $\hat{x}_{ob} = R\hat{x}$):

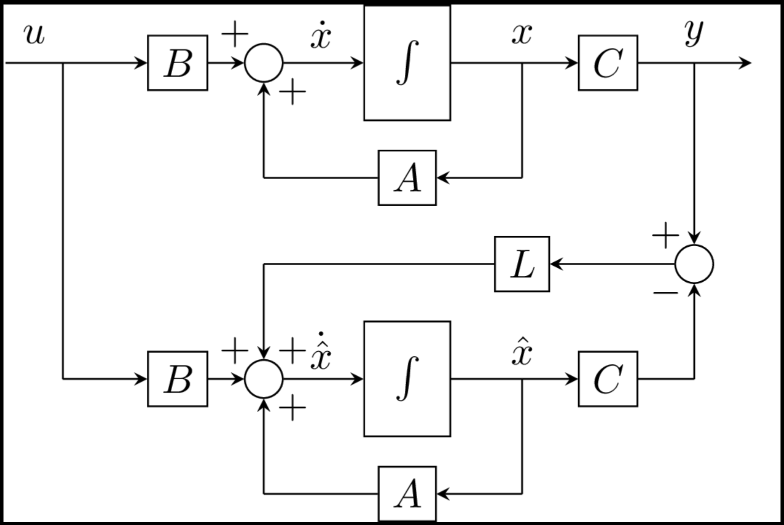
$$R\dot{\hat{x}} = \left(RAR^{-1} - \tilde{L}CR^{-1}\right)R\hat{x} + \tilde{L}y + RBu$$

$$\Rightarrow \dot{\hat{x}} = (A - \overbrace{R^{-1}\tilde{L}}^{L}\, C)\hat{x} + Ly + Bu \qquad \text{(implementation form)}$$

- **Powerful fact**: if system $\Sigma = (A, B, C, D)$ is observable, then we can arbitrarily place the observer eigenvalues.

Luenberger observer summary

- observer dynamics: $\dot{\hat{x}} = A\hat{x} + Bu + L\left(y - C\hat{x}\right), \ \hat{x}(0) = 0$
- block diagram



Luenberger observer summary

- system dynamics

$$\begin{aligned}\dot{x} &= Ax + Bu,\ x(0) = x_0,\ A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times r}\\ y &= Cx,\ y \in \mathbb{R}^{m\times 1}\end{aligned}$$

- observer dynamics

$$\begin{aligned}\dot{\hat{x}} &= A\hat{x} + Bu + L\left(y - C\hat{x}\right),\ \hat{x}(0) = 0\\ &= \left(A - LC\right)\hat{x} + LCx + Bu\end{aligned}$$

- augmented system

$$\begin{bmatrix}\dot{x}\\ \dot{\hat{x}}\end{bmatrix} = \begin{bmatrix}A & 0\\ LC & A - LC\end{bmatrix}\begin{bmatrix}x\\ \hat{x}\end{bmatrix} + \begin{bmatrix}B\\ B\end{bmatrix}u$$

Luenberger observer summary

- augmented system

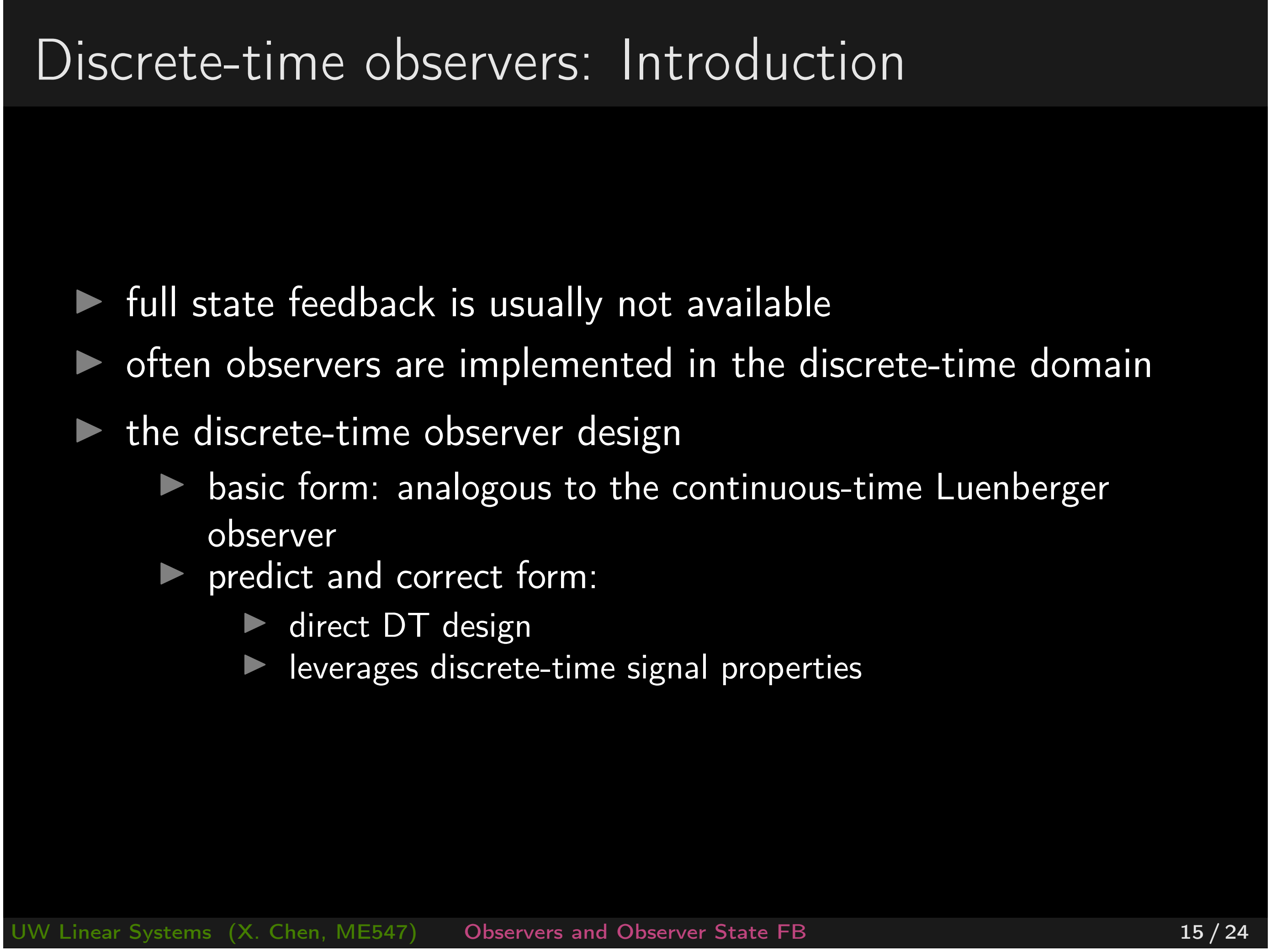
$$\begin{aligned}\begin{bmatrix}\dot{x}\\ \dot{\hat{x}}\end{bmatrix} &= \begin{bmatrix}A & 0\\ LC & A - LC\end{bmatrix}\begin{bmatrix}x\\ \hat{x}\end{bmatrix} + \begin{bmatrix}B\\ B\end{bmatrix}u\\ y &= Cx\end{aligned}$$

- to see the distribution of eigenvalues, note the error dynamics $\dot{e} = (A - LC)e \Rightarrow$

$$\begin{bmatrix}\dot{x}\\ \dot{e}\end{bmatrix} = \begin{bmatrix}A & 0\\ 0 & A - LC\end{bmatrix}\begin{bmatrix}x\\ e\end{bmatrix} + \begin{bmatrix}B\\ 0\end{bmatrix}u$$

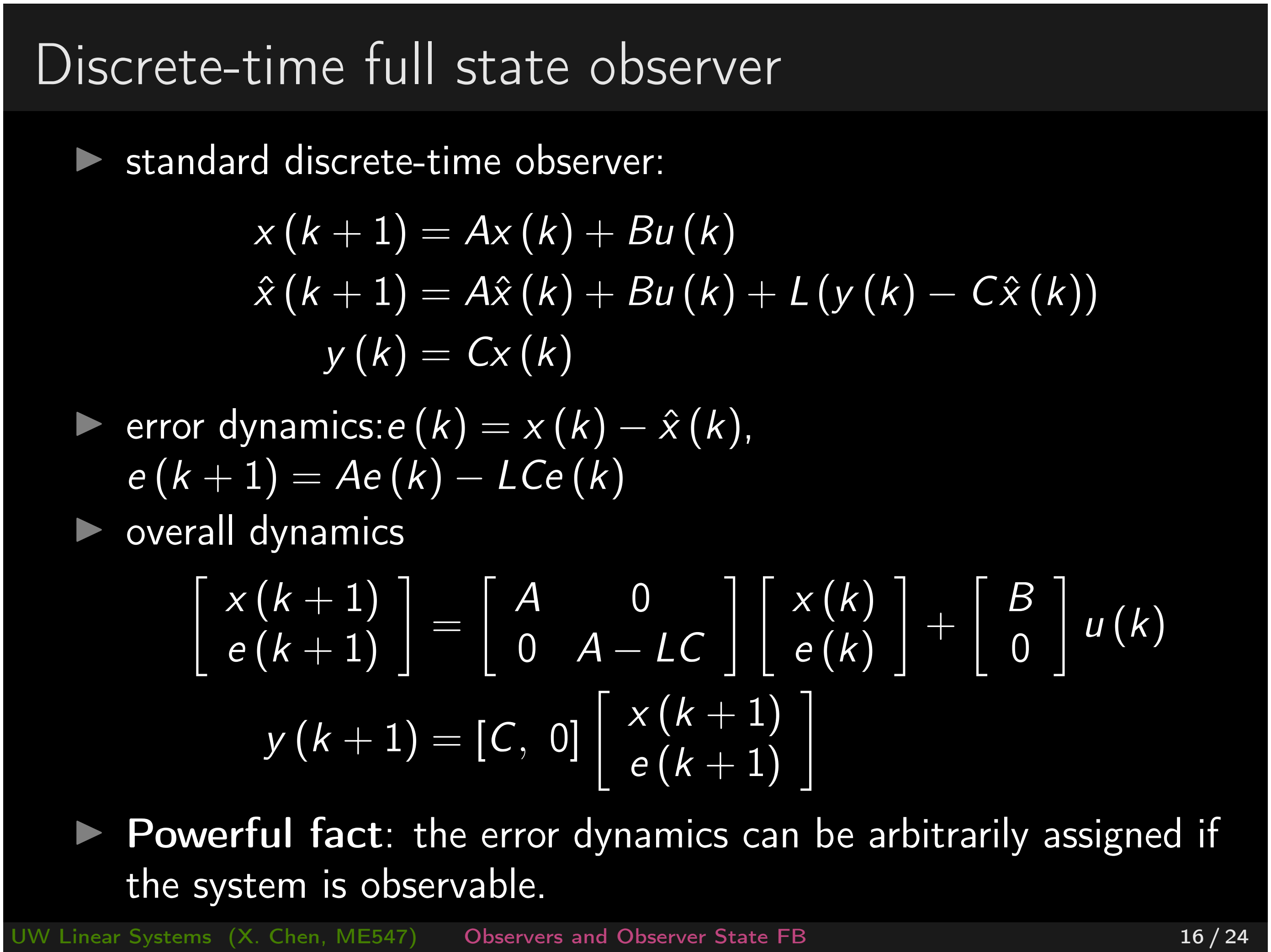
$\Rightarrow$eigenvalues are separated into: $\lambda\left(A\right)$ and observer eigenvalues

- underlying similarity transform: $\begin{bmatrix}x\\ e\end{bmatrix} = \begin{bmatrix}I_n & 0\\ I_n & -I_n\end{bmatrix}\begin{bmatrix}x\\ \hat{x}\end{bmatrix}$



Discrete-time observers: Introduction

- full state feedback is usually not available
- often observers are implemented in the discrete-time domain
- the discrete-time observer design
  - basic form: analogous to the continuous-time Luenberger observer
  - predict and correct form:
    - direct DT design
    - leverages discrete-time signal properties



Discrete-time full state observer

- standard discrete-time observer:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ \hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + L(y(k) - C\hat{x}(k)) \\ y(k) &= Cx(k) \end{aligned}$$

- error dynamics:$e(k) = x(k) - \hat{x}(k)$, $e(k+1) = Ae(k) - LCe(k)$
- overall dynamics

$$\begin{bmatrix} x(k+1) \\ e(k+1) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & A-LC \end{bmatrix} \begin{bmatrix} x(k) \\ e(k) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(k)$$

$$y(k+1) = [C,\ 0] \begin{bmatrix} x(k+1) \\ e(k+1) \end{bmatrix}$$

- **Powerful fact**: the error dynamics can be arbitrarily assigned if the system is observable.

DT full state observer with predictor

- motivation: $\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + L(y(k) - C\hat{x}(k))$ doesn't use most recent measurement $y(k+1) = Cx(k+1)$
- discrete-time observer **with predictor**:

  predictor: $\hat{x}(k+1|k) = A\hat{x}(k|k) + Bu(k)$

  corrector: $\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + L(y(k+1) - C\hat{x}(k+1|k))$

  - $\hat{x}(k|k)$: estimate of $x(k)$ based on measurements up to time k
  - $\hat{x}(k|k-1)$: estimate based on measurements up to time $k-1$
  - $e(k) \triangleq x(k) - \hat{x}(k|k)$: estimation error
- error dynamics

$$\begin{aligned}
\hat{x}(k+1|k+1) &= (I - LC)\hat{x}(k+1|k) + Ly(k+1) \\
&= (I - LC)A\hat{x}(k|k) + (I - LC)Bu(k) + Ly(k+1) \\
\Rightarrow e(k+1) &= x(k+1) - Ly(k+1) - (I - LC)A\hat{x}(k|k) - (I - LC)Bu(k) \\
&= (A - LCA)e(k)
\end{aligned}$$

DT full state observer with predictor

$$e(k+1) = \left(A - L\underbrace{CA}_{\tilde{C}}\right)e(k), \quad e(0) = (I - LC)x_0$$

- the error dynamics can be arbitrarily assigned if the pair $\left(A,\ \tilde{C}\right) = (A,\ CA)$ is observable
- observability matrix

$$\tilde{Q}_d = \begin{bmatrix} \tilde{C} \\ \tilde{C}A \\ \vdots \\ \tilde{C}A^{n-1} \end{bmatrix} = \overbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}}^{Q_d} A$$

- if A is invertible, then $\tilde{Q}_d$ has the same rank as Q_d
- $\left(A,\ \tilde{C}\right)$ is observable if $(A,\ C)$ is observable and A is nonsingular (guaranteed if discretized from a CT system)

Example

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} b_2 \\ b_1 \\ b_0 \end{bmatrix} u(k),$$

$y(k) = x_1(k)$. Place all eigenvalues of an **observer with predictor** at the origin.

$$A - LCA = \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \begin{bmatrix} -a_2 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} (l_1 - 1)a_2 & 1 - l_1 & 0 \\ l_2 a_2 - a_1 & -l_2 & 1 \\ l_3 a_2 - a_0 & -l_3 & 0 \end{bmatrix}$$

$\det(A - LCA - \lambda I) = ((l_1 - 1)a_2 - \lambda)(l_2 + \lambda)\lambda +$
$(1 - l_1)(l_3 a_a - a_0) + l_3((l_1 - 1)a_2 - \lambda) + \lambda(1 - l_1)(l_2 a_2 - a_1)$
roots must be all 0 $\Rightarrow l_1 = 1,\ l_2 = l_3 = 0$.

1. Concepts

2. Continuous-time Luenberger observer

3. Discrete-time observers
   DT full state observer
   DT full state observer with predictor

4. Observer state feedback

Observer state feedback

given system dynamics:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned}$$

- state feedback control: arbitrary eigenvalue assignment if system controllable
- observer design: arbitrary observer eigenvalue assignment for state estimation if system observable
- when full states are not available, what's the performance if we combine both?

$$u = -K\hat{x} + v$$

Closed-loop dynamics

- full closed-loop system

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \\ \dot{\hat{x}} &= A\hat{x} + Bu + L\left(y - C\hat{x}\right) \\ u &= -K\hat{x} + v \end{aligned}$$

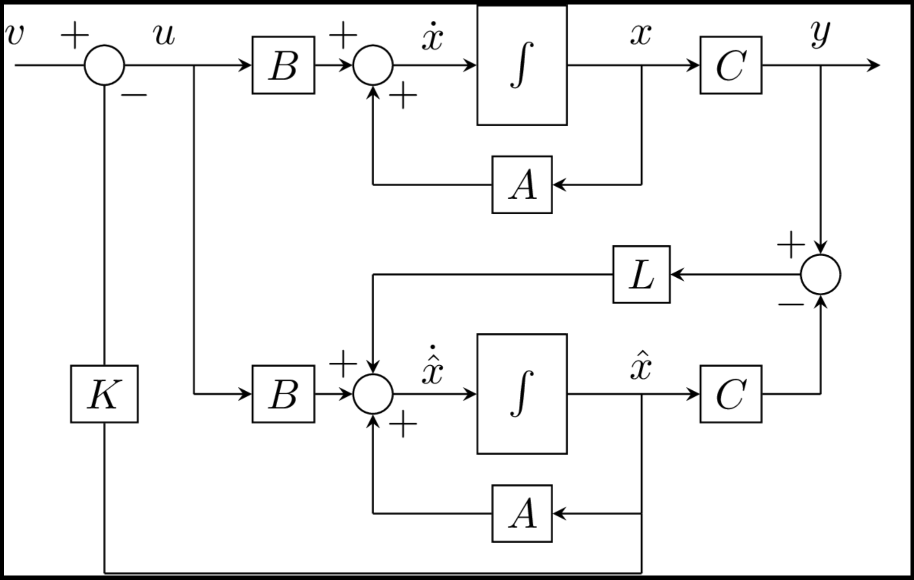
$$\Rightarrow \frac{d}{dt}\begin{bmatrix} x \\ \hat{x} \end{bmatrix} = \begin{bmatrix} A & -BK \\ LC & A - LC - BK \end{bmatrix}\begin{bmatrix} x \\ \hat{x} \end{bmatrix} + \begin{bmatrix} B \\ B \end{bmatrix} v$$

- using again similarity transform $\begin{bmatrix} x \\ e \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ I_n & -I_n \end{bmatrix}\begin{bmatrix} x \\ \hat{x} \end{bmatrix}$ gives

$$\frac{d}{dt}\begin{bmatrix} x \\ e \end{bmatrix} = \begin{bmatrix} A - BK & BK \\ 0 & A - LC \end{bmatrix}\begin{bmatrix} x \\ e \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} v$$

Block diagram

- $\dot{\hat{x}} = A\hat{x} + Bu + L(y - C\hat{x})$, $u = -K\hat{x} + v$



The separation theorem

- closed-loop dynamics

$$\frac{d}{dt}\begin{bmatrix} x \\ e \end{bmatrix} = \begin{bmatrix} A - BK & BK \\ 0 & A - LC \end{bmatrix}\begin{bmatrix} x \\ e \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} v$$

- **powerful result: separation theorem:** closed-loop eigenvalues consist of
  - eigenvalues of $A - BK$ from the state feedback control design
  - eigenvalues of $A - LC$ from the observer design
- can design K and L separately based on discussed tools
- if system is controllable and observable, we can arbitrarily assign the closed-loop eigenvalues
- rule of thumb: assign observer dynamics to be faster than state-feedback dynamics

ME 547: Linear Systems
Linear Quadratic Optimal Control

Xu Chen

University of Washington

Motivation

state feedback control:

- allows to arbitrarily assign the closed-loop eigenvalues for a controllable system
- the eigenvalue assignment has been manual thus far
- performance is implicit: we assign eigenvalues to induce proper error convergence

Motivation

state feedback control:

- allows to arbitrarily assign the closed-loop eigenvalues for a controllable system
- the eigenvalue assignment has been manual thus far
- performance is implicit: we assign eigenvalues to induce proper error convergence

linear quadratic (LQ) optimal regulation control, aka, LQ regulator (or LQR):

- no need to specify closed-loop poles
- performance is explicit: a performance index is defined ahead of time

1. Problem formulation

2. Solution to the finite-horizon LQ problem

3. From finite-horizon LQ to stationary LQ

Goal

Consider an n-dimensional state-space system

$$\begin{aligned} \dot{x}(t) &= Ax\left(t\right) + Bu\left(t\right), \;\; x\left(t_0\right) = x_0 \\ y\left(t\right) &= Cx\left(t\right) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.
LQ optimal control aims at minimizing the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

Goal

Consider an n-dimensional state-space system

$$\begin{aligned} \dot{x}(t) &= Ax\left(t\right) + Bu\left(t\right), \;\; x\left(t_0\right) = x_0 \\ y\left(t\right) &= Cx\left(t\right) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.
LQ optimal control aims at minimizing the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

- $S \succeq 0, Q \succeq 0, R \succ 0$: for a nonnegative cost and well-posed problem

Goal

Consider an n-dimensional state-space system

$$\begin{aligned} \dot{x}(t) &= Ax\left(t\right) + Bu\left(t\right),\ \ x\left(t_0\right) = x_0 \\ y\left(t\right) &= Cx\left(t\right) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.
LQ optimal control aims at minimizing the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

- $S \succeq 0, Q \succeq 0, R \succ 0$: for a nonnegative cost and well-posed problem
- $\frac{1}{2}x^T(t_f)Sx(t_f)$ penalizes the deviation of x from the origin at t_f

Goal

Consider an n-dimensional state-space system

$$\begin{aligned} \dot{x}(t) &= Ax\left(t\right) + Bu\left(t\right),\ \ x\left(t_0\right) = x_0 \\ y\left(t\right) &= Cx\left(t\right) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.
LQ optimal control aims at minimizing the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

- $S \succeq 0, Q \succeq 0, R \succ 0$: for a nonnegative cost and well-posed problem
- $\frac{1}{2}x^T(t_f)Sx(t_f)$ penalizes the deviation of x from the origin at t_f

- $x^T(t)Qx(t)$ $t \in (t_0, t_f)$ penalizes the transient

Goal

Consider an n-dimensional state-space system

$$\begin{aligned} \dot{x}(t) &= Ax\left(t\right) + Bu\left(t\right), \; x\left(t_0\right) = x_0 \\ y\left(t\right) &= Cx\left(t\right) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.
LQ optimal control aims at minimizing the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

- $S \succeq 0, Q \succeq 0, R \succ 0$: for a nonnegative cost and well-posed problem
- $\frac{1}{2}x^T(t_f)Sx(t_f)$ penalizes the deviation of x from the origin at t_f
- $x^T(t)Qx(t)$ $t \in (t_0, t_f)$ penalizes the transient
- often, $Q = C^TC \Rightarrow x^T(t)Qx(t) = y\left(t\right)^T y\left(t\right)$

Goal

Consider an n-dimensional state-space system

$$\begin{aligned} \dot{x}(t) &= Ax\left(t\right) + Bu\left(t\right), \; x\left(t_0\right) = x_0 \\ y\left(t\right) &= Cx\left(t\right) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, and $y \in \mathbb{R}^m$.
LQ optimal control aims at minimizing the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

- $S \succeq 0, Q \succeq 0, R \succ 0$: for a nonnegative cost and well-posed problem
- $\frac{1}{2}x^T(t_f)Sx(t_f)$ penalizes the deviation of x from the origin at t_f
- $x^T(t)Qx(t)$ $t \in (t_0, t_f)$ penalizes the transient
- often, $Q = C^TC \Rightarrow x^T(t)Qx(t) = y\left(t\right)^T y\left(t\right)$

- $u^T(t)Ru(t)$ penalizes large control efforts

Observations

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

- when the control horizon is made to be infinitely long, i.e., $t_f \to \infty$, the problem reduces to the infinite-horizon LQ problem

$$J = \frac{1}{2}\int_{t_0}^{\infty} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

Observations

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

- when the control horizon is made to be infinitely long, i.e., $t_f \to \infty$, the problem reduces to the infinite-horizon LQ problem

$$J = \frac{1}{2}\int_{t_0}^{\infty} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

  - terminal cost is not needed, as it will turn out, that the control will have to drive x to the origin. Otherwise J will go unbounded.

Observations

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

- when the control horizon is made to be infinitely long, i.e., $t_f \to \infty$, the problem reduces to the infinite-horizon LQ problem

$$J = \frac{1}{2}\int_{t_0}^{\infty} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

  - terminal cost is not needed, as it will turn out, that the control will have to drive x to the origin. Otherwise J will go unbounded.
  - often, we have $t_0 = 0$ and

$$J = \frac{1}{2}\int_{0}^{\infty} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

1. Problem formulation

2. Solution to the finite-horizon LQ problem

3. From finite-horizon LQ to stationary LQ

Solution to the finite-horizon LQ

Consider the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

with $\dot{x} = Ax + Bu,\ x\left(t_0\right) = x_0,\ S \succeq 0,\ R \succ 0$, and $Q = C^TC$.

Solution to the finite-horizon LQ

Consider the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

with $\dot{x} = Ax + Bu,\ x\left(t_0\right) = x_0,\ S \succeq 0,\ R \succ 0$, and $Q = C^TC$.

- do a Lyapunov-like construction: $V\left(t\right) \triangleq \frac{1}{2}x^T\left(t\right)P\left(t\right)x\left(t\right)$

Solution to the finite-horizon LQ

Consider the performance index

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

with $\dot{x} = Ax + Bu$, $x\left(t_0\right) = x_0$, $S \succeq 0$, $R \succ 0$, and $Q = C^TC$.

- do a Lyapunov-like construction: $V\left(t\right) \triangleq \frac{1}{2}x^T\left(t\right)P\left(t\right)x\left(t\right)$
- then

$$\begin{aligned}\frac{d}{dt}V\left(t\right) &= \frac{1}{2}\dot{x}^T\left(t\right)P\left(t\right)x\left(t\right) + \frac{1}{2}x^T\left(t\right)\dot{P}\left(t\right)x\left(t\right) + \frac{1}{2}x^T\left(t\right)P\left(t\right)\dot{x}\left(t\right)\\ &= \frac{1}{2}\left(Ax + Bu\right)^T Px + \frac{1}{2}x^T\frac{dP}{dt}x + \frac{1}{2}x^TP\left(Ax + Bu\right)\\ &= \frac{1}{2}\left\{x^T\left(t\right)\left(A^TP + \frac{dP}{dt} + PA\right)x\left(t\right) + u^TB^TPx + x^TPBu\right\}\end{aligned}$$

Solution to the finite-horizon LQ

with $\frac{d}{dt}V\left(t\right)$ from the last slide, we have

$$\begin{aligned}V\left(t_f\right) - V\left(t_0\right) &= \int_{t_0}^{t_f}\dot{V}dt\\ &= \frac{1}{2}\int_{t_0}^{t_f}\left(x^T\left(A^TP + PA + \frac{dP}{dt}\right)x + u^TB^TPx + x^TPBu\right)dt\end{aligned}$$

Solution to the finite-horizon LQ

with $\frac{d}{dt}V(t)$ from the last slide, we have

$$V(t_f) - V(t_0) = \int_{t_0}^{t_f} \dot{V} dt$$

$$= \frac{1}{2}\int_{t_0}^{t_f} \left(x^T \left(A^T P + PA + \frac{dP}{dt} \right) x + u^T B^T P x + x^T P B u \right) dt$$

- adding

$$J = \frac{1}{2} x^T(t_f) S x(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t) Q x(t) + u^T(t) R u(t) \right) dt$$

yields

$$J + V(t_f) - V(t_0) = \frac{1}{2} x^T(t_f) S x(t_f) +$$

$$\frac{1}{2}\int_{t_0}^{t_f} \left(x^T \left(A^T P + PA + Q + \frac{dP}{dt} \right) x + \underbrace{u^T B^T P x + x^T P B u}_{\text{products of } x \text{ and } u} + \underbrace{u^T R u}_{\text{quadratic}} \right) dt$$

Solution to the finite-horizon LQ

- "complete the squares" in $\underbrace{u^T B^T P x + x^T P B u}_{\text{products of } x \text{ and } u} + \underbrace{u^T R u}_{\text{quadratic}}$ (scalar case):

$$u^T B^T P x + x^T P B u + u^T R u \overset{\text{scalar case}}{=} R u^2 + 2 x P B u$$

$$= R u^2 + 2 \left(x P B R^{-1/2} \right) \underbrace{R^{1/2} u}_{\sqrt{R u^2}} + \left(R^{-1/2} B P x \right)^2 - \left(R^{-1/2} B P x \right)^2$$

$$= \left(R^{1/2} u + R^{-1/2} B P x \right)^2 - \left(R^{-1/2} B P x \right)^2$$

Solution to the finite-horizon LQ

- "complete the squares" in $\underbrace{u^TB^TPx + x^TPBu}_{\text{products of } x \text{ and } u} + \underbrace{u^TRu}_{\text{quadratic}}$ (scalar case):

$$
\begin{aligned}
u^TB^TPx + x^TPBu + u^TRu &\overset{\text{scalar case}}{=} Ru^2 + 2xPBu \\
&= Ru^2 + 2\left(xPBR^{-1/2}\right)\underbrace{R^{1/2}u}_{\sqrt{Ru^2}} + \left(R^{-1/2}BPx\right)^2 - \left(R^{-1/2}BPx\right)^2 \\
&= \left(R^{1/2}u + R^{-1/2}BPx\right)^2 - \left(R^{-1/2}BPx\right)^2
\end{aligned}
$$

- extending the concept to the general vector case:

$$
u^TB^TPx + x^TPBu + u^TRu = \underbrace{\|R^{\frac{1}{2}}u + R^{\frac{-1}{2}}B^TPx\|_2^2}_{\text{recall } \|\vec{a}\|_2^2 = \vec{a}^T\vec{a}} - x^TPBR^{-1}B^TPx
$$

Solution to the finite-horizon LQ

$$
J + V(t_f) - V(t_0) = \frac{1}{2}x^T(t_f)Sx(t_f) +
$$

$$
\frac{1}{2}\int_{t_0}^{t_f}\left(x^T\left(A^TP + PA + Q + \frac{dP}{dt}\right)x + \underbrace{u^TB^TPx + x^TPBu + u^TRu}_{\|R^{\frac{1}{2}}u + R^{\frac{-1}{2}}B^TPx\|_2^2 - x^TPBR^{-1}B^TPx}\right)dt
$$

⇓"completing the squares"

$$
J + \frac{1}{2}x^T(t_f)P(t_f)x(t_f) - \frac{1}{2}x^T(t_0)P(t_0)x(t_0) = \frac{1}{2}x^T(t_f)Sx(t_f) +
$$

$$
\frac{1}{2}\int_{t_0}^{t_f}\left(x^T\left(\frac{dP}{dt} + A^TP + PA + Q - PBR^{-1}B^TP\right)x + \|R^{\frac{1}{2}}u + R^{\frac{-1}{2}}B^TPx\|_2^2\right)dt
$$

Solution to the finite-horizon LQ

$$J + V(t_f) - V(t_0) = \frac{1}{2}x^T(t_f)Sx(t_f) +$$

$$\frac{1}{2}\int_{t_0}^{t_f}\left(x^T\left(A^TP + PA + Q + \frac{dP}{dt}\right)x + u^TB^TPx + x^TPBu + u^TRu\right)dt$$

⇓"completing the squares"

$$J + \underline{\frac{1}{2}x^T(t_f)P(t_f)x(t_f)} - \frac{1}{2}x^T(t_0)P(t_0)x(t_0) = \underline{\frac{1}{2}x^T(t_f)Sx(t_f)} +$$

$$\frac{1}{2}\int_{t_0}^{t_f}\left(x^T\underline{\underline{\left(\frac{dP}{dt} + A^TP + PA + Q - PBR^{-1}B^TP\right)}}x + \|\underline{\underline{\underline{R^{\frac{1}{2}}u + R^{\frac{-1}{2}}B^TPx}}}\|_2^2\right)dt$$

- the best that the control can do in minimizing the cost is to have

$$u(t) = -K(t)x(t) = \underline{\underline{\underline{-R^{-1}B^TP(t)x(t)}}}$$

$$-\frac{dP}{dt} = \underline{\underline{A^TP + PA - PBR^{-1}B^TP + Q}},\ \underline{P(t_f) = S}$$

  to yield the optimal cost $J^0 = \frac{1}{2}x_0^TP(t_0)x_0$

Observation 1

$$u(t) = -K(t)x(t) = -R^{-1}B^TP(t)x(t) \qquad \text{optimal control law}$$

$$-\frac{dP}{dt} = A^TP + PA - PBR^{-1}B^TP + Q,\ P(t_f) = S \qquad \text{the Riccati differential equation}$$

- the control $u(t) = -R^{-1}B^TP(t)x(t)$ is a state feedback law (the power of state feedback!)

Observation 1

$$u(t) = -K(t)\,x(t) = -R^{-1}B^T P(t)x(t) \qquad \text{optimal control law}$$

$$-\frac{dP}{dt} = A^T P + PA - PBR^{-1}B^T P + Q,\ P(t_f) = S \qquad \text{the Riccati differential equation}$$

- the control $u(t) = -R^{-1}B^T P(t)\,x(t)$ is a state feedback law (the power of state feedback!)
- the state feedback law is time-varying because of $P(t)$
- the closed-loop dynamics becomes

$$\dot{x}(t) = Ax(t) + Bu(t) = \underbrace{\left(A - BR^{-1}B^T P(t)\right)}_{\text{time-varying closed-loop dynamics}} x(t)$$

Observation 2

$$u(t) = -K(t)\,x(t) = -R^{-1}B^T P(t)x(t) \qquad \text{optimal state feedback control}$$

$$-\frac{dP}{dt} = A^T P + PA - PBR^{-1}B^T P + Q,\ P(t_f) = S \qquad \text{the Riccati differential equation}$$

- boundary condition of the Riccati equation is given at the final time t_f $\Rightarrow$ the equation must be integrated backward in time

Observation 2

$$u(t) = -K(t)x(t) = -R^{-1}B^TP(t)x(t) \qquad \text{optimal state feedback control}$$

$$-\frac{dP}{dt} = A^TP + PA - PBR^{-1}B^TP + Q,\ P(t_f) = S \qquad \text{the Riccati differential equation}$$

- boundary condition of the Riccati equation is given at the final time $t_f \Rightarrow$ the equation must be integrated backward in time
- *backward* integration of

  $$-\frac{dP}{dt} = A^TP + PA + Q - PBR^{-1}B^TP,\ P(t_f) = S$$

  is equivalent to the *forward* integration of

  $$\frac{dP^*}{dt} = A^TP^* + P^*A + Q - P^*BR^{-1}B^TP^*,\ P^*(0) = S \qquad (2)$$

  by letting $P(t) = P^*(t_f - t)$

Observation 2

$$u(t) = -K(t)x(t) = -R^{-1}B^TP(t)x(t) \qquad \text{optimal state feedback control}$$

$$-\frac{dP}{dt} = A^TP + PA - PBR^{-1}B^TP + Q,\ P(t_f) = S \qquad \text{the Riccati differential equation}$$

- boundary condition of the Riccati equation is given at the final time $t_f \Rightarrow$ the equation must be integrated backward in time
- *backward* integration of

  $$-\frac{dP}{dt} = A^TP + PA + Q - PBR^{-1}B^TP,\ P(t_f) = S$$

  is equivalent to the *forward* integration of

  $$\frac{dP^*}{dt} = A^TP^* + P^*A + Q - P^*BR^{-1}B^TP^*,\ P^*(0) = S \qquad (2)$$

  by letting $P(t) = P^*(t_f - t)$

- Eq. (2) can be solved by numerical integration, e.g., ODE45 in Matlab

Observation 3

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

$$J^0 = \frac{1}{2}x_0^T P(t_0)x_0$$

- the minimum value J^0 is a function of the initial state $x\left(t_0\right)$

Observation 3

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

$$J^0 = \frac{1}{2}x_0^T P(t_0)x_0$$

- the minimum value J^0 is a function of the initial state $x\left(t_0\right)$

- J (and hence J^0) is nonnegative $\Rightarrow$ $P\left(t_0\right)$ is at least positive semidefinite

Observation 3

$$J=\frac{1}{2}x^T(t_f)Sx(t_f)+\frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t)+u^T(t)Ru(t)\right)dt$$

$$J^0=\frac{1}{2}x_0^TP(t_0)x_0$$

- the minimum value J^0 is a function of the initial state $x\left(t_0\right)$
- J (and hence J^0) is nonnegative $\Rightarrow P\left(t_0\right)$ is at least positive semidefinite
- t_0 can be taken anywhere in $(0,t_f)\Rightarrow P\left(t\right)$ is at least positive semidefinite for any t

Example: LQR of a pure inertia system

Consider

$$\dot{x}=\begin{bmatrix}0&1\\0&0\end{bmatrix}x+\begin{bmatrix}0\\1\end{bmatrix}u,\ J=\frac{1}{2}x^T\left(t_f\right)Sx\left(t_f\right)+\frac{1}{2}\int_0^{t_f}\left(x^TQx+Ru^2\right)dt$$

where $S=\begin{bmatrix}1&0\\0&1\end{bmatrix},\ Q=\begin{bmatrix}1&0\\0&0\end{bmatrix},\ R>0$

Example: LQR of a pure inertia system

Consider

$$\dot{x} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} x + \begin{bmatrix}0\\ 1\end{bmatrix} u,\ J = \frac{1}{2}x^T\left(t_f\right)Sx\left(t_f\right) + \frac{1}{2}\int_0^{t_f}\left(x^TQx + Ru^2\right)dt$$

where $S = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ Q = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\ R > 0$

- we let $P\left(t\right) = P^*\left(t_f - t\right)$ and solve

$$\frac{dP^*}{dt} = A^TP^* + P^*A + Q - P^*BR^{-1}B^TP^*,\ P^*\left(0\right) = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$\Leftrightarrow \frac{dP^*}{dt} = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}P^* + P^*\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} - P^*\begin{bmatrix}0\\ 1\end{bmatrix}\frac{1}{R}\begin{bmatrix}0 & 1\end{bmatrix}P^*$$

Example: LQR of a pure inertia system

Consider

$$\dot{x} = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} x + \begin{bmatrix}0\\ 1\end{bmatrix} u,\ J = \frac{1}{2}x^T\left(t_f\right)Sx\left(t_f\right) + \frac{1}{2}\int_0^{t_f}\left(x^TQx + Ru^2\right)dt$$

where $S = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ Q = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\ R > 0$

- we let $P\left(t\right) = P^*\left(t_f - t\right)$ and solve

$$\frac{dP^*}{dt} = A^TP^* + P^*A + Q - P^*BR^{-1}B^TP^*,\ P^*\left(0\right) = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

$$\Leftrightarrow \frac{dP^*}{dt} = \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}P^* + P^*\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} - P^*\begin{bmatrix}0\\ 1\end{bmatrix}\frac{1}{R}\begin{bmatrix}0 & 1\end{bmatrix}P^*$$

- letting

$$P^* = \begin{bmatrix}p_{11}^* & p_{12}^*\\ p_{12}^* & p_{22}^*\end{bmatrix} \Rightarrow \begin{cases}\frac{d}{dt}p_{11}^* = 1 - \frac{1}{R}\left(p_{12}^*\right)^2\\ \frac{d}{dt}p_{12}^* = p_{11}^* - \frac{1}{R}p_{12}^*p_{22}^*\\ \frac{d}{dt}p_{22}^* = 2p_{12}^* - \frac{1}{R}\left(p_{22}^*\right)^2\end{cases} \Rightarrow \begin{matrix}p_{11}^*\left(0\right) = 1\\ p_{12}^*\left(0\right) = 0\\ p_{22}^*\left(0\right) = 1\end{matrix}$$

Example: LQR of a pure inertia system: analysis

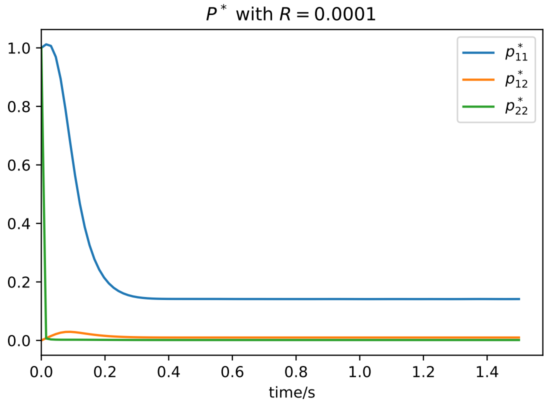


Figure: LQ example: $P^*(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $P(t) = P^*(t_f - t)$

Example: LQR of a pure inertia system: analysis

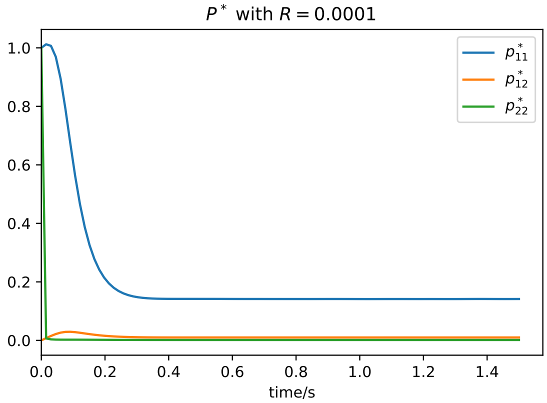


Figure: LQ example: $P^*(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $P(t) = P^*(t_f - t)$

- if the final time t_f is large, $P^*(t)$ forward converges to a stationary value

Example: LQR of a pure inertia system: analysis

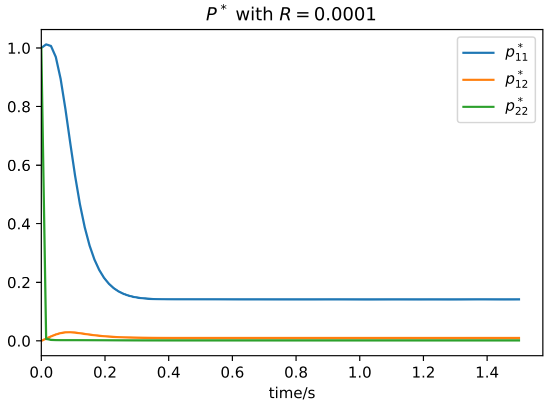


Figure: LQ example: $P^*(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $P(t) = P^*(t_f - t)$

- if the final time t_f is large, $P^*(t)$ forward converges to a stationary value
- i.e., $P(t)$ backward converges to a stationary value at $P(0)$

Example: LQR of a pure inertia system: analysis

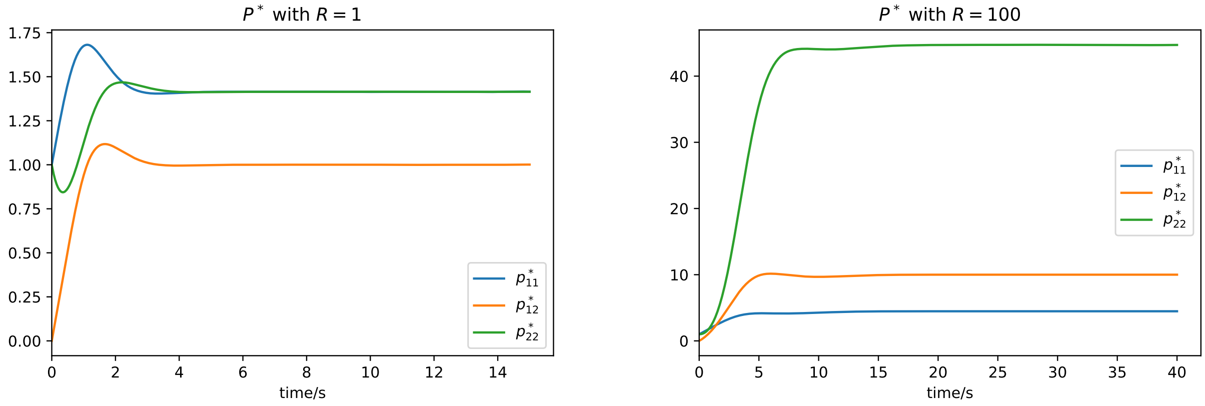


Figure: LQ example with different penalties on control. $P^*(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Example: LQR of a pure inertia system: analysis

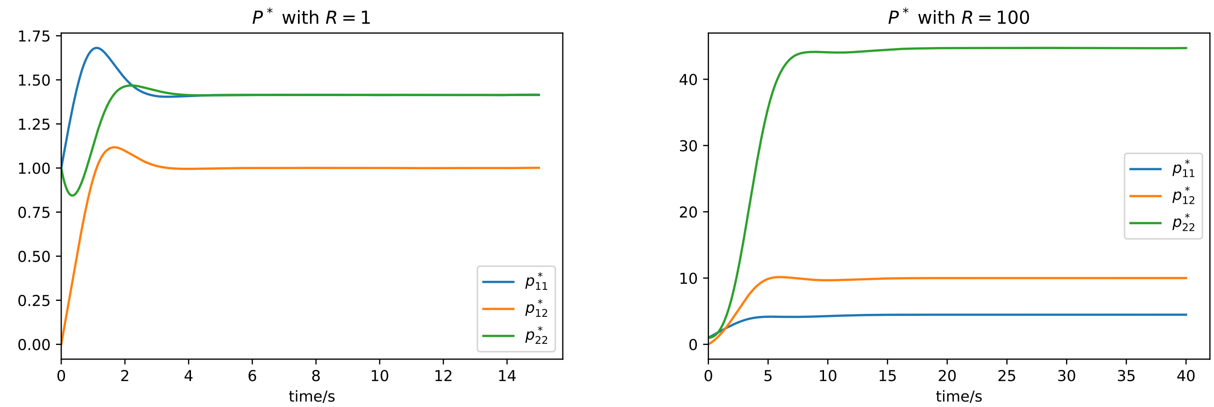


Figure: LQ example with different penalties on control. $P^*(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- a larger R results in a longer transient

Example: LQR of a pure inertia system: analysis

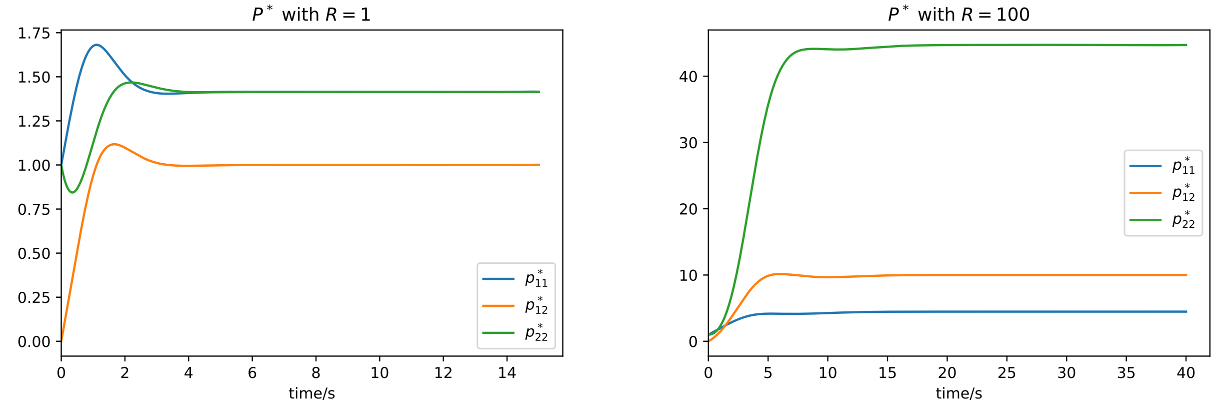
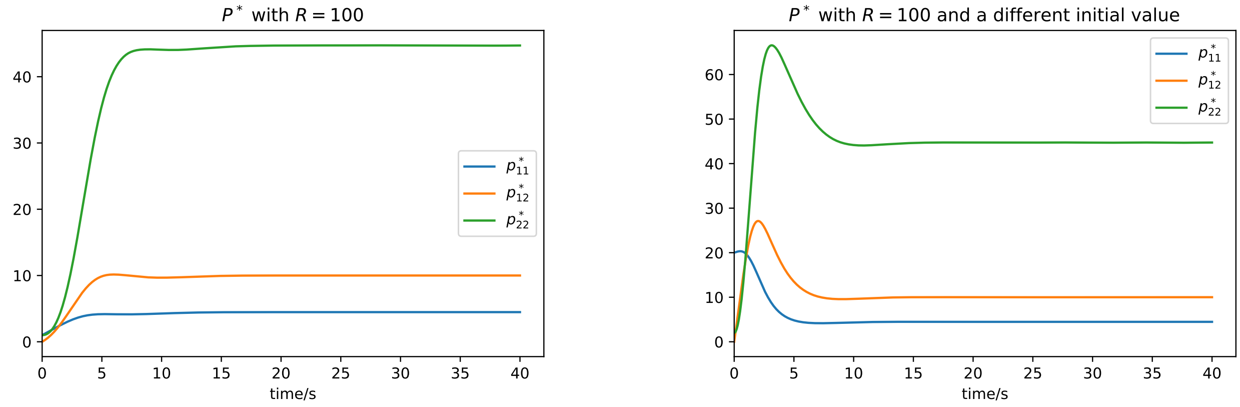


Figure: LQ example with different penalties on control. $P^*(0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- a larger R results in a longer transient

- i.e., a larger penalty on the control input yields a longer time to settle

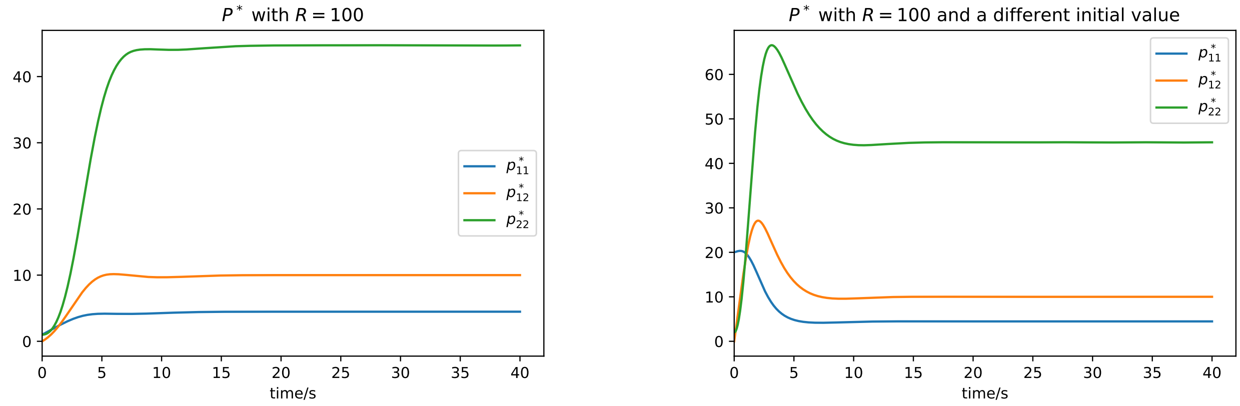
Example: LQR of a pure inertia system: analysis



(a) $P^*(0) = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ (b) $P^*(0) = \begin{bmatrix}20 & 0\\ 0 & 2\end{bmatrix}$

Figure: LQ with different boundary values in Riccati difference Eq.

Example: LQR of a pure inertia system: analysis



(a) $P^*(0) = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ (b) $P^*(0) = \begin{bmatrix}20 & 0\\ 0 & 2\end{bmatrix}$

Figure: LQ with different boundary values in Riccati difference Eq.

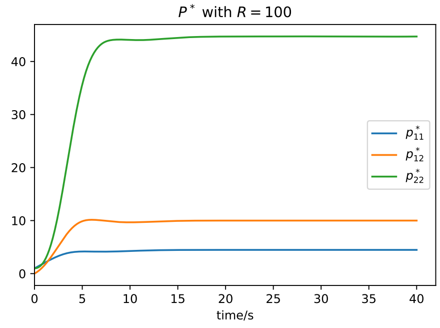
- for the same R, the initial value $P(t_f) = S$ becomes irrelevant as $t_f \to \infty$

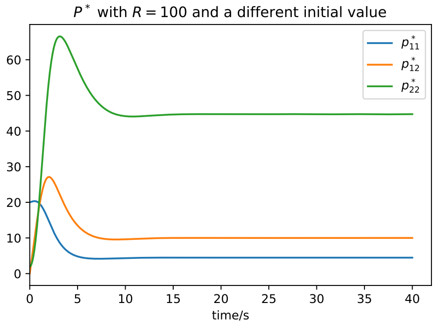
1. Problem formulation

2. Solution to the finite-horizon LQ problem

3. From finite-horizon LQ to stationary LQ

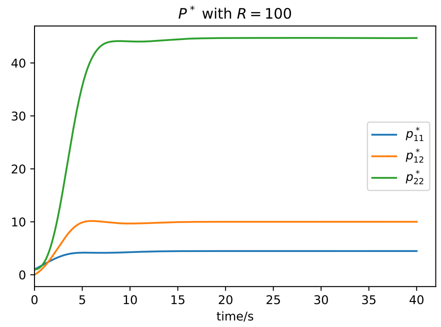
From LQ to stationary LQ

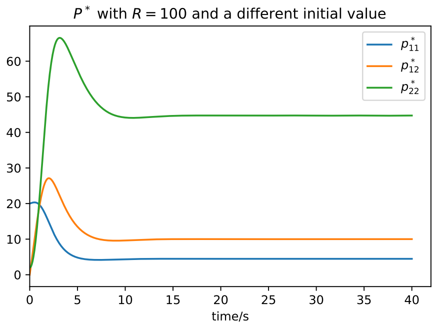




- in the example, we see that P in the Riccati differential Eq. converges to a stationary value given sufficient time

From LQ to stationary LQ





- in the example, we see that P in the Riccati differential Eq. converges to a stationary value given sufficient time
- when $t_f \to \infty$, LQ becomes the stationary LQ problem, under two additional conditions that we now discuss in details:
  - (A, B) is controllable/stabilizable
  - (A, C) is observable/detectable

Need for controllability/stabilizability

$$J = \frac{1}{2}x^T(t_f)Sx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t) + u^T(t)Ru(t)\right)dt$$

$$-\frac{dP}{dt} = A^TP + PA - PBR^{-1}B^TP + Q,\ P(t_f) = S \qquad \text{the Riccati differential equation}$$

$$J^0 = \frac{1}{2}x_0^TP(t_0)x_0$$

if (A, B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

Need for controllability/stabilizability

$$J=\frac{1}{2}x^T(t_f)Sx(t_f)+\frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t)+u^T(t)Ru(t)\right)dt$$

$$-\frac{dP}{dt}=A^TP+PA-PBR^{-1}B^TP+Q,\ P(t_f)=S \qquad \text{the Riccati differential equation}$$

$$J^0=\frac{1}{2}x_0^TP(t_0)x_0$$

if (A,B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

- for uncontrollable or unstabilizable systems, there can be unstable uncontrollable modes that cause J to be unbounded

Need for controllability/stabilizability

$$J=\frac{1}{2}x^T(t_f)Sx(t_f)+\frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t)+u^T(t)Ru(t)\right)dt$$

$$-\frac{dP}{dt}=A^TP+PA-PBR^{-1}B^TP+Q,\ P(t_f)=S \qquad \text{the Riccati differential equation}$$

$$J^0=\frac{1}{2}x_0^TP(t_0)x_0$$

if (A,B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

- for uncontrollable or unstabilizable systems, there can be unstable uncontrollable modes that cause J to be unbounded

- then if $J^0=\frac{1}{2}x_0^TP(0)x_0$ is unbounded, we will have $||P(0)||=\infty$

Need for controllability/stabilizability

if (A, B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

- e.g.: $\dot{x} = x + 0 \cdot u$, $x(0) = 1$, $Q = 1$ and R be any positive value
  - system is uncontrollable and the uncontrollable mode is unstable

Need for controllability/stabilizability

if (A, B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

- e.g.: $\dot{x} = x + 0 \cdot u$, $x(0) = 1$, $Q = 1$ and R be any positive value
  - system is uncontrollable and the uncontrollable mode is unstable

  - $x(t)$ will keep increasing to infinity

Need for controllability/stabilizability

if (A, B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

- e.g.: $\dot{x} = x + 0 \cdot u$, $x(0) = 1$, $Q = 1$ and R be any positive value
  - system is uncontrollable and the uncontrollable mode is unstable
  - $x(t)$ will keep increasing to infinity
  - $\Rightarrow J = \frac{1}{2}\int_0^\infty \left(x^T Q x + u^T R u\right) dt$ unbounded regardless of $u(t)$

Need for controllability/stabilizability

if (A, B) is controllable or stabilizable, then $P(t)$ is guaranteed to converge to a bounded and stationary value

- e.g.: $\dot{x} = x + 0 \cdot u$, $x(0) = 1$, $Q = 1$ and R be any positive value
  - system is uncontrollable and the uncontrollable mode is unstable
  - $x(t)$ will keep increasing to infinity
  - $\Rightarrow J = \frac{1}{2}\int_0^\infty \left(x^T Q x + u^T R u\right) dt$ unbounded regardless of $u(t)$
  - in this case, the Riccati equation is

$$-\frac{dP}{dt} = P + P + 1 = 2P + 1 \Leftrightarrow \frac{dP^*}{dt} = 2P^* + 1$$

forward integration of P^* (backward integration of P), will drive $P^*(\infty)$ and $P(0)$ to infinity

Need for observability/detectability

if (A, C) is observable or detectable, the optimal state feedback control system will be asymptotically stable

Need for observability/detectability

if (A, C) is observable or detectable, the optimal state feedback control system will be asymptotically stable

- *intuition*: if the system is observable, $y = Cx$ will relate to all states ⇒ regulating $x^T Qx = x^T C^T Cx$ will regulate all states

Need for observability/detectability

if (A, C) is observable or detectable, the optimal state feedback control system will be asymptotically stable

- *intuition*: if the system is observable, $y = Cx$ will relate to all states ⇒ regulating $x^TQx = x^TC^TCx$ will regulate all states
- *formally*: if (A, C) is observable (detectable), the solution of the Riccati equation will converge to a positive (semi)definite value P_+ (proof in course notes)

From LQ to stationary LQ

	LQ		stationary LQ
Cost	$J = \frac{1}{2}x^T(t_f)Sx(t_f)+$ $\frac{1}{2}\int_{t_0}^{t_f}\left(x^T(t)Qx(t)+u^T(t)Ru(t)\right)dt$	⇒	$J = \frac{1}{2}\int_{t_0}^{\infty}\left(x^TQx+u^TRu\right)dt$
Syst.	$\dot{x} = Ax + Bu$	⇒	$\dot{x} = Ax + Bu$ (A, B) controllable/stabilizable (A, C) observable/detectable
Key Eq.	Riccati Eq. (RE) $-\frac{dP}{dt} = A^TP + PA - PBR^{-1}B^TP$ $+Q, \quad P(t_f) = S$	⇒	Algebraic RE (ARE) $A^TP + PA - PBR^{-1}B^TP + Q = 0$
Opt. control & cost	$u(t) = -R^{-1}B^TP(t)x(t)$ $J^0 = \frac{1}{2}x_0^TP(t_0)x_0$	⇒ ⇒	$u(t) = -R^{-1}B^TP_+x(t)$ $J^0 = \frac{1}{2}x_0^TP_+x_0$

More formally: Solution of the infinite-horizon LQ

For

$$J = \frac{1}{2}\int_{t_0}^{\infty}\left(x\left(t\right)^T Qx\left(t\right) + u\left(t\right)^T Ru\left(t\right)\right)dt,\ \ Q = C^TC$$

with $\dot{x}(t) = Ax\left(t\right) + Bu\left(t\right),\ \ x\left(t_0\right) = x_0$ and $R \succ 0$:

- if (A, B) is **controllable** (stabilizable) and (A, C) is **observable** (detectable)

More formally: Solution of the infinite-horizon LQ

For

$$J = \frac{1}{2}\int_{t_0}^{\infty}\left(x\left(t\right)^T Qx\left(t\right) + u\left(t\right)^T Ru\left(t\right)\right)dt,\ \ Q = C^TC$$

with $\dot{x}(t) = Ax\left(t\right) + Bu\left(t\right),\ \ x\left(t_0\right) = x_0$ and $R \succ 0$:

- if (A, B) is **controllable** (stabilizable) and (A, C) is **observable** (detectable)
- then the optimal control input is given by

$$u(t) = -R^{-1}B^TP_+x(t)$$

More formally: Solution of the infinite-horizon LQ

For

$$J = \frac{1}{2}\int_{t_0}^{\infty}\left(x\left(t\right)^T Qx\left(t\right) + u\left(t\right)^T Ru\left(t\right)\right)dt,\ \ Q = C^T C$$

with $\dot{x}(t) = Ax\left(t\right) + Bu\left(t\right),\ \ x\left(t_0\right) = x_0$ and $R \succ 0$:

- if (A, B) is **controllable** (stabilizable) and (A, C) is **observable** (detectable)
- then the optimal control input is given by

$$u(t) = -R^{-1}B^T P_+ x(t)$$

- where $P_+ \left(= P_+^T\right)$ is the positive (semi)definite solution of the **algebraic Riccati equation** (ARE)

$$A^T P + PA - PBR^{-1}B^T P + Q = 0$$

More formally: Solution of the infinite-horizon LQ

For

$$J = \frac{1}{2}\int_{t_0}^{\infty}\left(x\left(t\right)^T Qx\left(t\right) + u\left(t\right)^T Ru\left(t\right)\right)dt,\ \ Q = C^T C$$

with $\dot{x}(t) = Ax\left(t\right) + Bu\left(t\right),\ \ x\left(t_0\right) = x_0$ and $R \succ 0$:

- if (A, B) is **controllable** (stabilizable) and (A, C) is **observable** (detectable)
- then the optimal control input is given by

$$u(t) = -R^{-1}B^T P_+ x(t)$$

- where $P_+ \left(= P_+^T\right)$ is the positive (semi)definite solution of the **algebraic Riccati equation** (ARE)

$$A^T P + PA - PBR^{-1}B^T P + Q = 0$$

- and the closed-loop system is **asymptotically stable**, with

$$J_{\min} = J^0 = \frac{1}{2}x\left(t_0\right)^T P_+ x\left(t_0\right)$$

Observations

- the control $u(t) = -R^{-1}B^TPx(t)$ is a *constant* state feedback law

Observations

- the control $u(t) = -R^{-1}B^TPx(t)$ is a *constant* state feedback law

- under the optimal control, the closed loop is given by $\dot{x} = Ax - BR^{-1}B^TPx = \underbrace{\left(A - BR^{-1}B^TP\right)}_{A_c} x$ and $J = \frac{1}{2}\int_{t_0}^{\infty}\left(x^TQx + u^TRu\right)dt = \frac{1}{2}\int_{t_0}^{\infty} x^T \underbrace{\left(Q + PBR^{-1}B^TP\right)}_{Q_c} xdt$

Observations

- the control $u(t) = -R^{-1}B^TPx(t)$ is a *constant* state feedback law
- under the optimal control, the closed loop is given by $\dot{x} = Ax - BR^{-1}B^TPx = \underbrace{\left(A - BR^{-1}B^TP\right)}_{A_c} x$ and $J = \frac{1}{2}\int_{t_0}^{\infty}\left(x^TQx + u^TRu\right)dt = \frac{1}{2}\int_{t_0}^{\infty} x^T \underbrace{\left(Q + PBR^{-1}B^TP\right)}_{Q_c} xdt$
- for the above closed-loop system, the Lyapunov Eq. is

$$\begin{aligned} A_c^TP + PA_c &= -Q_c \\ \Leftrightarrow \left(A - BR^{-1}B^TP\right)^T P + P\left(A - BR^{-1}B^TP\right) &= -Q - PBR^{-1}B^TP \\ \Leftrightarrow A^TP + PA - PBR^{-1}B^TP &= -Q \text{ (the ARE!)} \end{aligned}$$

Observations

- the control $u(t) = -R^{-1}B^TPx(t)$ is a *constant* state feedback law
- under the optimal control, the closed loop is given by $\dot{x} = Ax - BR^{-1}B^TPx = \underbrace{\left(A - BR^{-1}B^TP\right)}_{A_c} x$ and $J = \frac{1}{2}\int_{t_0}^{\infty}\left(x^TQx + u^TRu\right)dt = \frac{1}{2}\int_{t_0}^{\infty} x^T \underbrace{\left(Q + PBR^{-1}B^TP\right)}_{Q_c} xdt$
- for the above closed-loop system, the Lyapunov Eq. is

$$\begin{aligned} A_c^TP + PA_c &= -Q_c \\ \Leftrightarrow \left(A - BR^{-1}B^TP\right)^T P + P\left(A - BR^{-1}B^TP\right) &= -Q - PBR^{-1}B^TP \\ \Leftrightarrow A^TP + PA - PBR^{-1}B^TP &= -Q \text{ (the ARE!)} \end{aligned}$$

- when the ARE solution P_+ is positive definite, $\frac{1}{2}x^TP_+x$ is a Lyapunov function for the closed-loop system

Observations

- Lyapunov Eq. and the ARE:

Cost	$\overline{J}=\frac{1}{2}\int_0^\infty x^TQ_cxdt$	$J=\frac{1}{2}\int_{t_0}^\infty\left(x^TQx+u^TRu\right)dt$
Syst. dynamics	$\dot{x}=A_cx$	$\dot{x}=Ax+Bu$ (A,B) controllable/stabilizable (A,C) observable/detectable
Key Eq.	$A_c^TP+PA_c+Q_c=0$	$A^TP+PA-PBR^{-1}B^TP+Q=0$
Optimal control	N/A	$u(t)=-R^{-1}B^TP_+x(t)$
Opt. cost	$\overline{J}^0=\frac{1}{2}x^T(0)P_+x(0)$	$J^0=\frac{1}{2}x(t_0)^TP_+x(t_0)$

Observations

- Lyapunov Eq. and the ARE:

Cost	$\overline{J}=\frac{1}{2}\int_0^\infty x^TQ_cxdt$	$J=\frac{1}{2}\int_{t_0}^\infty\left(x^TQx+u^TRu\right)dt$
Syst. dynamics	$\dot{x}=A_cx$	$\dot{x}=Ax+Bu$ (A,B) controllable/stabilizable (A,C) observable/detectable
Key Eq.	$A_c^TP+PA_c+Q_c=0$	$A^TP+PA-PBR^{-1}B^TP+Q=0$
Optimal control	N/A	$u(t)=-R^{-1}B^TP_+x(t)$
Opt. cost	$\overline{J}^0=\frac{1}{2}x^T(0)P_+x(0)$	$J^0=\frac{1}{2}x(t_0)^TP_+x(t_0)$

- the guaranteed closed-loop stability is an attractive feature
- more nice properties will show up later

Example: Stationary LQR of a pure inertia system

- Consider

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u,\ J = \frac{1}{2}\int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt,\ R > 0$$

Example: Stationary LQR of a pure inertia system

- Consider

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u,\ J = \frac{1}{2}\int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt,\ R > 0$$

- the ARE is

$$0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} P + P \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - P \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{R} \begin{bmatrix} 0 & 1 \end{bmatrix} P \Rightarrow P_+ = \begin{bmatrix} \sqrt{2}R^{1/4} & R^{1/2} \\ R^{1/2} & \sqrt{2}R^{3/4} \end{bmatrix}$$

Example: Stationary LQR of a pure inertia system

- Consider

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \; J = \frac{1}{2}\int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt, \; R > 0$$

- the ARE is

$$0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} P + P \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - P \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{R} \begin{bmatrix} 0 & 1 \end{bmatrix} P \Rightarrow P_+ = \begin{bmatrix} \sqrt{2}R^{1/4} & R^{1/2} \\ R^{1/2} & \sqrt{2}R^{3/4} \end{bmatrix}$$

- the closed-loop A matrix can be computed to be

$$A_c = A - BR^{-1}B^T P_+ = \begin{bmatrix} 0 & 1 \\ -R^{-1/2} & -\sqrt{2}R^{-1/4} \end{bmatrix}$$

Example: Stationary LQR of a pure inertia system

- Consider

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \; J = \frac{1}{2}\int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt, \; R > 0$$

- the ARE is

$$0 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} P + P \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - P \begin{bmatrix} 0 \\ 1 \end{bmatrix} \frac{1}{R} \begin{bmatrix} 0 & 1 \end{bmatrix} P \Rightarrow P_+ = \begin{bmatrix} \sqrt{2}R^{1/4} & R^{1/2} \\ R^{1/2} & \sqrt{2}R^{3/4} \end{bmatrix}$$

- the closed-loop A matrix can be computed to be

$$A_c = A - BR^{-1}B^T P_+ = \begin{bmatrix} 0 & 1 \\ -R^{-1/2} & -\sqrt{2}R^{-1/4} \end{bmatrix}$$

- ⇒ closed-loop eigenvalues:

$$\lambda_{1,2} = -\frac{1}{\sqrt{2}R^{1/4}} \pm \frac{1}{\sqrt{2}R^{1/4}} j$$

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \ J = \frac{1}{2}\int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt$$

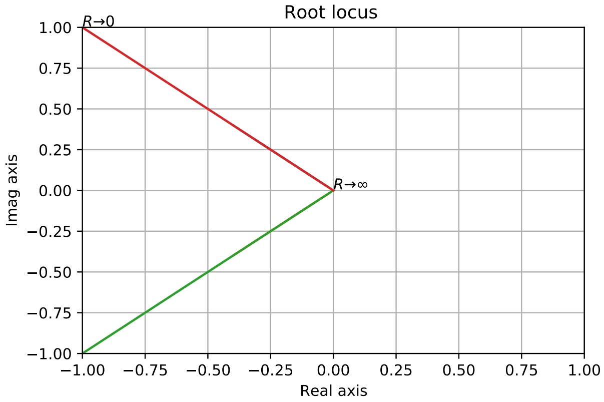


Figure: Eigenvalue $\lambda_{1,2} = -\frac{1}{\sqrt{2}R^{1/4}} \pm \frac{1}{\sqrt{2}R^{1/4}} j$ evolution (root locus)

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \ J = \frac{1}{2}\int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt$$

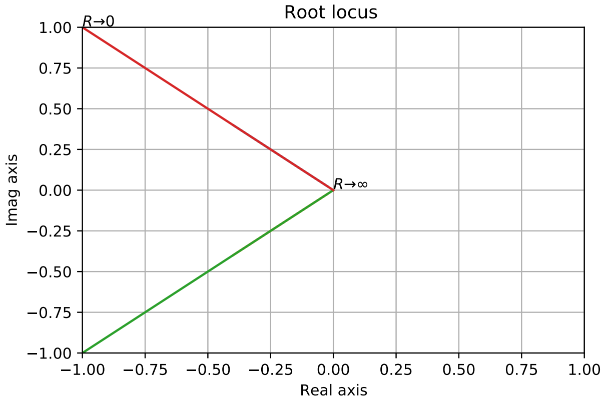


Figure: Eigenvalue $\lambda_{1,2} = -\frac{1}{\sqrt{2}R^{1/4}} \pm \frac{1}{\sqrt{2}R^{1/4}} j$ evolution (root locus)

- $R \uparrow$ (more penalty on the control input) $\Rightarrow$ $\lambda_{1,2}$ move closer to the origin $\Rightarrow$ slower state convergence to zero

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \ J = \frac{1}{2} \int_0^\infty \left(x^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x + Ru^2 \right) dt$$

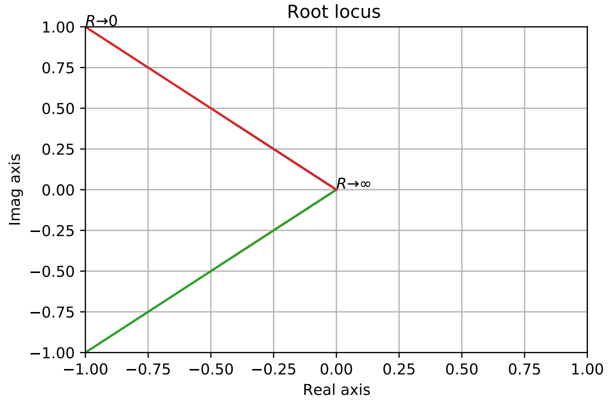


Figure: Eigenvalue $\lambda_{1,2} = -\frac{1}{\sqrt{2}R^{1/4}} \pm \frac{1}{\sqrt{2}R^{1/4}}j$ evolution (root locus)

- $R \uparrow$ (more penalty on the control input) $\Rightarrow \lambda_{1,2}$ move closer to the origin $\Rightarrow$ slower state convergence to zero
- $R \downarrow$ (allow for large control efforts) $\Rightarrow \lambda_{1,2}$ move further to the left of the complex plane $\Rightarrow$ faster speed of closed-loop dynamics

MATLAB commands

- *care*: solves the ARE for a continuous-time system:

$$[P, \Lambda, K] = \text{care}\left(A, B, C^T C, R\right)$$

where $K = R^{-1} B^T P$ and Λ is a diagonal matrix with the closed-loop eigenvalues, i.e., the eigenvalues of $A - BK$, in the diagonal entries.

MATLAB commands

- *care*: solves the ARE for a continuous-time system:

$$[P, \Lambda, K] = \text{care}\left(A, B, C^T C, R\right)$$

  where $K = R^{-1}B^T P$ and Λ is a diagonal matrix with the closed-loop eigenvalues, i.e., the eigenvalues of $A - BK$, in the diagonal entries.

- *lqr* and *lqry*: provide the LQ regulator with

$$[K, P, \Lambda] = \text{lqr}\left(A, B, C^T C, R\right)$$
$$[K, P, \Lambda] = \text{lqry}\left(\text{sys}, Q_y, R\right)$$

  where *sys* is defined by $\dot{x} = Ax + Bu, \; y = Cx + Du$, and

$$J = \frac{1}{2}\int_0^\infty \left(y^T Q_y y + u^T R u\right) dt$$

Additional excellent properties of stationary LQ

- we know stationary LQR yields guaranteed closed-loop stability for controllable (stabilizable) and observable (detectable) systems

Additional excellent properties of stationary LQ

- we know stationary LQR yields guaranteed closed-loop stability for controllable (stabilizable) and observable (detectable) systems

It turns out that LQ regulators with full state feedback has excellent additional properties of:

- at least a 60 degree phase margin
- infinite gain margin
- stability is guaranteed up to a 50% reduction in the gain

Applications and practice

choosing R and Q:

- if there is not a good idea for the structure for Q and R, start with diagonal matrices;

Applications and practice

choosing R and Q:

- if there is not a good idea for the structure for Q and R, start with diagonal matrices;
- gain an idea of the magnitude of each state variable and input variable
- call them $x_{i,\max}$ $(i = 1, \ldots, n)$ and $u_{i,\max}$ $(i = 1, \ldots, r)$
- make the diagonal elements of Q and R inversely proportional to $||x_{i,\max}||^2$ and $||u_{i,\max}||^2$, respectively.

University of Washington
Lecture Notes
Linear Algebra for Controls

Xu Chen
Bryan T. McMinn Endowed Research Professorship
Associate Professor
Department of Mechanical Engineering
University of Washington
chx [AT] uw [DOT] edu

Contents

1 Basic concepts of matrices and vectors

A linear equation set

$$\begin{aligned} 3x_1 + 4x_2 + 10x_3 &= 6 \\ x_1 + 4x_2 - 10x_3 &= 5 \\ 4x_2 + 10x_3 &= -1, \end{aligned} \tag{1}$$

can be simply written as

$$\begin{bmatrix} 3 & 4 & 10 \\ 1 & 4 & -10 \\ 0 & 4 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \\ -1 \end{bmatrix}. \tag{2}$$

Equation (2) wrote x_1, x_2, and x_3 just once rather than two or three times in (1). There are only three unknowns in the above linear equation set. The notational simplicity and many algebraic convenience that will arise, however, are significant when we have thousands of unknowns...

Formally, we write an $m \times n$ matrix A as

$$A = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & \dots & a_{2n} \\ \vdots & \dots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Here,

- $m \times n$ (reads m by n) is the dimension/size of the matrix. It means that A has m rows and n columns.
- Each element a_{jk} is an entry of the matrix. For two matrices A and B to be equal, it must be that $a_{jk} = b_{jk}$ for any j and k.
- If $m = n$, A belongs to the class of square matrices. The entries a_{11}, a_{22}, ..., a_{nn} are then called the diagonal entries of A.
  - Upper triangular matrices : square matrices with nonzero entries only on and above the main diagonal.
  - Lower triangular matrices : nonzero entries only on and below the main diagonal.
  - Diagonal matrices : nonzero entries only on the main diagonal.
  - Identity matrice : diagonal and all diagonal entries are 1.
- Vectors: special matrices whose row or column number is one.
  - A row vector: $a = [a_1, a_2, \dots, a_n]$; its dimension is $1 \times n$.

– A $m \times 1$ column vector:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Example (Matrix and quadratic forms). We can use matrices to express general quadratic functions of vectors. For instance

$$f(x) = x^T Ax + 2bx + c$$

is equivalent to

$$f(x) = \begin{bmatrix} x \\ 1 \end{bmatrix}^T \begin{bmatrix} A & b \\ b^T & c \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}.$$

1.1 Matrix addition and multiplication

The **sum** of two matrices A *and* B *(of the same size)* is

$$A + B = [a_{jk} + b_{jk}].$$

The **product** between a $m \times n$ matrix A and a scalar c is

$$cA = [ca_{jk}],$$

i.e. each entry of A is multiplied by c to generate the corresponding entry of cA.

The **matrix product** $C = AB$ is meaningful only if the column number of A equals the row number of B. The computation is done as shown in the following example:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ \boxed{a_{21}} & \boxed{a_{22}} & \boxed{a_{23}} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \begin{bmatrix} \boxed{b_{11}} & b_{12} \\ \boxed{b_{21}} & b_{22} \\ \boxed{b_{31}} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ \boxed{c_{21}} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{bmatrix},$$

where

$$\begin{aligned} c_{21} &= a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ &= [a_{21}, a_{22}, a_{23}] \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} \\ &= \text{"second row of } A\text{"} \times \text{"first column of } B\text{"}. \end{aligned}$$

More generally:

$$\begin{aligned} c_{jk} &= a_{j1}b_{1k} + a_{j2}b_{2k} + \cdots + a_{jn}b_{nk} \\ &= [a_{j1}, a_{j2}, \ldots, a_{jn}] \begin{bmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{nk} \end{bmatrix}, \end{aligned} \tag{3}$$

namely, the jk entry of C is obtained by multiplying each entry in the jth row of A by the corresponding entry in the kth column of B and then adding these n products. This is called a multiplication of rows into columns.

Matrix multiplication is not commutative: It is a good habit to always check the matrix dimensions when doing matrix products:

$$\begin{array}{cccc} A & B & = & C \\ [m\times n] & [n\times p] & & [m\times p] \end{array}.$$

This way it is clear that AB in general does not equal to BA, e.g.,

$$ABC = (AB)\,C = A\,(BC) \neq BCA.$$

Matrices as combination of vectors: The matrix-vector product

$$Ax = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 \end{bmatrix}$$

is nothing but the weighted sum of the columns of A:

$$Ax = \left[\begin{array}{c|c|c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{41} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ a_{42} \end{bmatrix} + x_3 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ a_{43} \end{bmatrix}.$$

1.2 Matrix transposition

Definition 1 (Transpose)**.** The transpose of an $m \times n$ matrix

$$A = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & \dots & a_{2n} \\ \vdots & \dots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

is the $n \times m$ matrix A^T (reads "A transpose") defined as

$$A^T = [a_{kj}] = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & \dots & \dots & a_{m2} \\ \vdots & \dots & \dots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}.$$

Transposition has the following rules:

- $\left(A^T\right)^T = A$
- $(A+B)^T = A^T + B^T$
- $(cA)^T = cA^T$
- $(AB)^T = B^T A^T$

If $A = A^T$, then A is called symmetric. If $A = -A^T$ then A is called skew-symmetric.

2 Linear systems of equations

A linear system of m equations in n unknowns $x_1, \ldots, x_n$ is a set of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n &= b_2 \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n &= b_m \end{aligned} \tag{4}$$

Here,

- The equation set is linear: each variable x_j appears in the first power only.
- If all the b_j are zero, then the linear equation is called a homogeneous system. Otherwise, it is a nonhomogeneous system.
- Homogeneous systems always have at least the trivial solution $x_1 = x_2 = \cdots = x_n = 0$.

The m equations (4) can be written as a single vector equation

$$Ax = b,$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & \ldots & a_{mn} \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Gauss[1] elimination is a systematic method to solve linear equations. Consider

$$\underbrace{\begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 0 & 10 & 25 \\ 20 & 10 & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 90 \\ 80 \end{bmatrix}}_{b}.$$

The Gauss elimination process is as follows:

[1] Johann Carl Friedrich Gauss, 1777-1855, German mathematician: contributed significantly to many fields, including number theory, algebra, statistics, analysis, differential geometry, geodesy, geophysics, electrostatics, astronomy, Matrix theory, and optics.

Gauss was an ardent perfectionist. He was never a prolific writer, refusing to publish work which he did not consider complete and above criticism. Mathematical historian Eric Temple Bell estimated that, had Gauss published all of his discoveries in a timely manner, he would have advanced mathematics by fifty years.

1. Obtain the augmented matrix of the system

$$\left[\begin{array}{c|c} A & b \end{array}\right] = \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right].$$

2. Perform elementary row operation on the augmented matrix, to obtain the Row Echelon Form. Adding the first row to the second row gives

$$\text{pivot role}: \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ \boxed{-1} & \boxed{1} & \boxed{-1} & \boxed{0} \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right] \xrightarrow[\text{add pivot role}]{\text{row 2}} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right]$$

$$\xrightarrow[\text{add -20}\times\text{pivot role}]{\text{row 4}} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \end{array}\right].$$

What we have done is using the pivot row to eliminate x_1 in the other equations. At this stage, the linear equations look like

$$x_1 - x_2 + x_3 = 0 \tag{5}$$
$$0 = 0 \tag{6}$$
$$10x_2 + 25x_3 = 90 \tag{7}$$
$$30x_2 - 20x_3 = 80. \tag{8}$$

Re-arranging yields

$$x_1 - x_2 + x_3 = 0 \tag{9}$$
$$10x_2 + 25x_3 = 90 \tag{10}$$
$$30x_2 - 20x_3 = 80 \tag{11}$$
$$0 = 0. \tag{12}$$

Moving on, we can get ride of x_2 in the third equation, by adding to it -3 times the second equation. Correspondingly in the augmented matrix, we have

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 0 & -95 & -190 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow[\text{normalizing}]{} \underbrace{\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 1 & 5/2 & 9 \\ 0 & 0 & 1 & 38/19 \\ 0 & 0 & 0 & 0 \end{array}\right]}_{\text{the row echelon form}},$$

namely

$$\begin{aligned} x_3 &= 38/19 \\ x_2 + x_3 &= 9 \\ x_1 - x_2 + x_3 &= 0. \end{aligned}$$

The unknowns can now be readily obtained by back substitution: $x_3 = 38/19$, $x_2 = 9 - x_3$, $x_1 = x_2 - x_3$.

Elementary Row Operations for Matrices What we have done can be summarized by the following elementary matrix row operations:

- Interchange of two rows
- Addition of a constant multiple of one row to another row
- Multiplication of a row by a nonzero constant c

Let the final row echelon form be denoted by

$$\left[\, R \,\middle|\, f \,\right].$$

We have:

1. The two systems $Ax = b$ and $Rx = f$ are equivalent.
2. At the end of the Gauss elimination (before the back substitution), the row echelon form of the augmented matrix will be

$$\left[\begin{array}{cccccc|c} r_{11} & r_{12} & \dots & \dots & \dots & r_{1n} & f_1 \\ & r_{22} & \dots & \dots & \dots & r_{2n} & f_2 \\ & & \ddots & \dots & \dots & \vdots & \vdots \\ & & & r_{rr} & \dots & r_{rn} & f_r \\ & & & & & & f_{r+1} \\ & & & & & & \vdots \\ & & & & & & f_m \end{array}\right],$$

   where all unfilled entries are zero.
3. The number of nonzero rows, r, in the row-reduced coefficient matrix R is called the rank of R and also the rank of A.
4. Solution concepts:

   (a) *No solution* / system is inconsistent: r is less than m and $f_{r+1}, f_{r+2}, \dots, f_m$ are not all zero.

(b) *Unique solution*: if the system is consistent and $r = n$, there is exactly one solution, which can be found by back substitution.

(c) *Infinitely many solutions*: if $f_{r+1} = f_{r+2} = \ldots = f_m = 0$. To obtain any of these solutions, choose values of $x_{r+1}, \ldots, x_n$ arbitrarily. Then solve the r-th equation for x_r (in terms of those arbitrary values), then the $(r-1)$-st equation for x_{r-1}, and so on up the line.

3 Vector space, linear independence, basis, and span

Given a set of m vectors a_1, a_2, ..., a_m with the same size,

$$k_1a_1 + k_2a_2 + \cdots + k_ma_m$$

is called a linear combination of the vectors. If

$$a_1 = k_2a_2 + k_3a_3 + \cdots + k_ma_m,$$

then a_1 is said to be *linearly dependent* on a_2, a_3, ..., a_m. The set

$$\{a_1, a_2, \ldots, a_m\} \tag{13}$$

is then a linearly dependent set. The same idea holds if a_2 or any vector in the set (13) is linearly dependent on others.

Generalizing, if

$$k_1a_1 + k_2a_2 + \cdots + k_ma_m = 0$$

holds if and only if

$$k_1 = k_2 = \cdots = k_m = 0,$$

then the vectors in (13) are linearly dependent. This is saying that at least one of the vectors can be expressed as a linear combination of the other vectors.

Why is linear independence important? If a set of vectors is linearly dependent, then we can get rid of one or perhaps more of the vectors until we get a linearly independent set. This set is then the smallest "truly essential" set with which we can work.

Consider a set of n linearly independent vectors, a_1, a_2, ..., a_n, each with n components. All the possible linear combinations of a_1, a_2, ..., a_n form the vector space $\mathbb{R}^n$. This is the *span* of the n vectors.

Definition 2 (Basis)**.** A *basis* of $\mathbf{V}$ is a set $\mathbf{B}$ of vectors in $\mathbf{V}$, such that any $v \in \mathbf{V}$ can be uniquely expressed as a finite linear combination of vectors in $\mathbf{B}$.

Example 3. In $\mathbb{R}^2$

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

is a linearly independent set and forms a basis.

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ v_3 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

is not a linearly independent set.

4 Matrix properties

4.1 Rank

Definition 4 (Rank)**.** The rank of a matrix A is the maximum number of linearly independent row or column vectors.

Theorem. *Row or column operations do not change the rank of a matrix.*

With the concept of linear dependence, many matrix-matrix operations can be understood from the view point of vector manipulations.

Example (Dyad)**.** $A = uv^T$ is called a dyad, where u and v are vectors of proper dimensions. It is a rank 1 matrix, as can be seen that $A = uv^T$ is formed by linear combinations of the vector u, where the weights of the combinations are coefficients of v.

Fact. *For $A, B \in \mathbb{R}^{n\times n}$, if $rank\,(A) = n$ then $AB = 0$ implies $B = 0$. If $AB = 0$ but $A \neq 0$ and $B \neq 0$, then $rank\,(A) < n$ and $rank\,(B) < n$.*

4.2 Range and null spaces

Definition 5 (Range space)**.** The range space of a matrix A, denoted as $\mathcal{R}\,(A)$, is the span of all the column vectors of A.

Definition 6 (Null space)**.** The null space of a matrix $A \in \mathbb{R}^{n\times n}$, denoted as $\mathcal{N}\,(A)$, is the vector space

$$\{x \in \mathbb{R}^n : \ Ax = 0\}\,.$$

The dimension of the null space is called *nullity* of the matrix.

Fact 7. *The following is true:*

$$\mathcal{N}\left(AA^T\right) = \mathcal{N}\left(A^T\right);\ \mathcal{R}\left(AA^T\right) = \mathcal{R}\,(A)\,.$$

4.3 Determinants

Determinants were originally introduced for solving linear equations in the form of $Ax = y$, with a square A. They are cumbersome to compute for high-order matrices, but their definitions and concepts are partially very important.

We review only the computations of second- and third-order matrices:

- 2×2 matrices:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

- 3×3 matrices:

$$\begin{aligned}\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} &= a\det\begin{bmatrix} e & f \\ h & k \end{bmatrix} - b\det\begin{bmatrix} d & f \\ g & k \end{bmatrix} + c\det\begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ &= aek + bfg + cdh - gec - bdk - ahf,\end{aligned}$$

  where $\det\begin{bmatrix} e & f \\ h & k \end{bmatrix}$, $\det\begin{bmatrix} d & f \\ g & k \end{bmatrix}$, and $\det\begin{bmatrix} d & e \\ g & h \end{bmatrix}$ are called the minors of $\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$.

Caution: $\det(cA) = c^n \det(A)$ (not $c\det(A)$!)

Theorem 8. *The determinant of A is nonzero if and only if A is full rank.*

You should be able to verify the theorem for 2×2 matrices. The proof will be immediate after introducing the concept of eigenvalues.

Definition 9. A linear transformation is called singular if the determinant of the corresponding transformation matrix is zero.

Fact 10. *Determinant facts:*

- *If A and B are square matrices, then*

$$\begin{aligned}\det(AB) &= \det(BA) = \det A \det B \\ \det\left(A^T\right) &= \det(A) \\ \det(A^*) &= \det(A).\end{aligned}$$

- *If X and Z are square, Y with compatible dimensions, then*

$$\det\left(\begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}\right) = \det X \det Z.$$

5 Matrix and linear equations

Consider again, using now concepts in range and null spaces of matrices, the linear equations

$$Ax = y. \tag{14}$$

- *Existence* of solutions requires that $y \in \mathcal{R}(A)$.
- The linear equation is called *overdetermined* if it has more equations than unknowns (i.e. A is a tall skinny matrix), *determined* if A is square, *undetermined* if it has fewer equations than unknowns (A is a wide matrix).

- *Solutions* of the above equation, provided that they exist, is constructed from

$$x = x_o + z: \;\; Az = 0, \tag{15}$$

  where x_0 is any (fixed) solution of (14) and z runs through all the homogeneous solutions of $Az = 0$, namely, z runs through all vectors in the null space of A.

- *Uniqueness* of a solution: if the null space of A is zero, the solution is unique.

You should be familiar with solving 2nd or 3rd-order linear equations by hand.

6 Eigenvector and eigenvalue

6.1 Matrix, mappings, and eigenvectors

Think of Ax this way: A defines a linear operator; Ax is a vector produced by feeding the vector x to this linear operator. In the two-dimensional case, we can look at Fig. 1. Certainly, Ax does not (at all) need to be in the same direction as x. An example is

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

which gives that

$$A_0 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix},$$

namely, Ax is x projected on the first axis in the two-dimensional vector space, which will not be in the same direction as x as long as $x_2 \neq 0$.

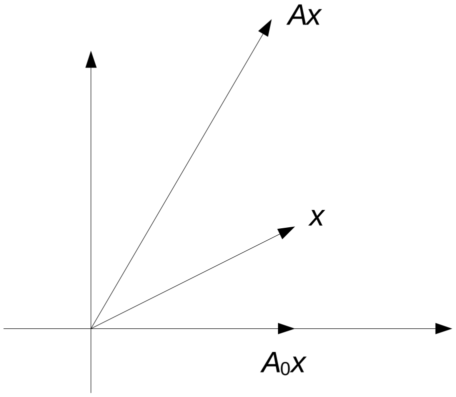


Figure 1: Example relationship between x and Ax.

From here comes the concept of eigenvectors and eigenvalues. It says that there are certain "special directions/vectors" (denoted as v_1 and v_2 in our two-dimensional example) for A such that $Av_i = \lambda_i v_i$. Thus Av_i is on the same line as the original vector v_i, just scaled by the eigenvalue λ_i. It can be shown that if $\lambda_1 \neq \lambda_2$, then v_1 and v_2 are linearly independent (your homework). This is saying that any vector in $\mathbb{R}^2$ can be decomposed as

$$x = a_1 v_1 + a_2 v_2.$$

Therefore

$$Ax = a_1 A v_1 + a_2 A v_2 = a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2.$$

Knowing λ_i and v_i thus can directly tell us how Ax looks like. More important, we have decomposed Ax into small modules that are from time to time more handy for analyzing the system properties. Figs. 2 and 3 demonstrate the above idea graphically.

Remark 11. The above geometric interpretations are for matrices with distinct real eigenvalues.

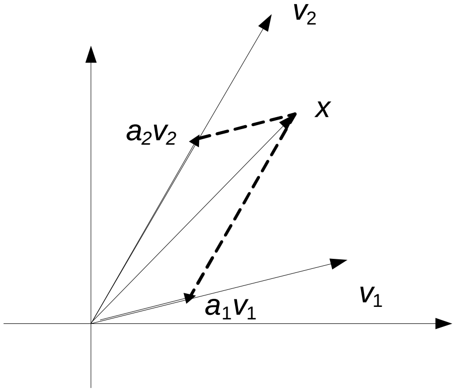


Figure 2: Decomposition of x.

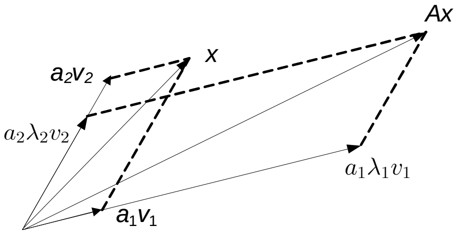


Figure 3: Construction of Ax.

The geometric interpretation above makes eigenvalue a very important concept. *Eigenvalues* are also called *characteristic values* of a matrix. The set of all the eigenvalues of A is called the *spectrum* of A. The largest of the *absolute* values of the eigenvalues of A is called the *spectral radius* of A.

6.2 Computation of eigenvalue and eigenvectors

Formally, eigenvalue and eigenvector are defined as follows. For $A \in \mathbb{R}^{n\times n}$, an eigenvalue λ of A is one for which

$$Ax = \lambda x \tag{16}$$

has a nonzero solution $x \neq 0$. The corresponding solutions are called eigenvectors of A.

Equation (16) is equivalent to

$$(A - \lambda I)\, x = 0. \tag{17}$$

As $x \neq 0$, the matrix $A - \lambda I$ must be singular, so

$$\det(A - \lambda I) = 0. \tag{18}$$

$\det(A - \lambda I)$ is a polynomial of λ, called the characteristic polynomial. Correspondingly, (18) is called the characteristic equation. So eigenvalues are roots of the characteristic equation. If an $n \times n$ matrix A has n eigenvalues $\lambda_1, \ldots, \lambda_n$, it must be that

$$\det(A - \lambda I) = (\lambda_1 - \lambda)\cdots(\lambda_n - \lambda).$$

After obtaining an eigenvalue λ, we can find the associated eigenvector by solving (17). This is nothing but solving a homogeneous system.

Example 12. Consider

$$A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}.$$

Then

$$\begin{aligned} \det(A - \lambda I) = 0 \Rightarrow \det\left(\begin{bmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{bmatrix}\right) = 0 \\ \Rightarrow (5+\lambda)(2+\lambda) - 4 = 0 \\ \Rightarrow \lambda = -1 \text{ or } -6. \end{aligned}$$

So A has two eigenvalues: -1 and -6. The characteristic polynomial of A is $\lambda^2 + 7\lambda + 6$.

To obtain the eigenvector associated to $\lambda = -1$, we solve

$$(A - \lambda I)x = 0 \Leftrightarrow \left(\begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix} + 1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)x = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}x = 0.$$

One solution is

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

As an exercise, show that an eigenvector associated to $\lambda = -6$ is $\begin{bmatrix} 2 & -1 \end{bmatrix}^T$.

Example 13 (Multiple eigenvectors)**.** Obtain the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}.$$

Analogous procedures give that

$$\lambda_1 = 5, \ \lambda_2 = \lambda_3 = -3.$$

So there are repeated eigenvalues. For $\lambda_2 = \lambda_3 = -3$, the characteristic matrix is

$$A + 3I = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ -1 & -2 & 3 \end{bmatrix}.$$

The second row is the first row multiplied by 2. The third row is the negative of the first row. So the characteristic matrix has only rank 1. The characteristic equation

$$(A - \lambda_2 I)x = 0$$

has two linearly independent solutions

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$

Theorem 14 (Eigenvalue and determinant). *Let $A \in \mathbb{R}^{n\times n}$. Then*

$$\det A = \prod_{i=1}^{n} \lambda_i.$$

Proof. Letting $\lambda = 0$ in the characteristic polynomial

$$p\left(\lambda\right) = \det\left(A - \lambda I\right) = \left(\lambda_1 - \lambda\right)\left(\lambda_2 - \lambda\right)\dots$$

gives

$$\det\left(A\right) = p\left(0\right) = \prod_{i=1}^{n} \lambda_i.$$

□

Example 15. For the two-dimensional case

$$A = \left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right] \Rightarrow p\left(\lambda\right) = \det\left(A - \lambda I\right) = \left(a_{11} - \lambda\right)\left(a_{22} - \lambda\right) - a_{12}a_{21}.$$

On the other hand

$$p\left(\lambda\right) = \left(\lambda_1 - \lambda\right)\left(\lambda_2 - \lambda\right).$$

Matching the coefficients we get

$$\begin{aligned} \lambda_1 + \lambda_2 &= a_{11} + a_{22} \\ \lambda_1\lambda_2 &= a_{11}a_{22} - a_{12}a_{21}. \end{aligned}$$

6.3 Eigenbases and diagonalization

Eigenvectors of an $n \times n$ matrix A may (or may not!) form a basis for $\mathbb{R}^n$. If we are interested in a transformation $y = Ax$, such an "eigenbasis" (basis of eigenvectors), if exists, is of great advantage because then we can represent any x in $\mathbb{R}^n$ uniquely as a linear combination of the eigenvectors $x_1, \ldots, x_n$, say, $x = c_1x_1 + c_2x_2 + \ldots + c_nx_n$. And, denoting the corresponding (not necessarily distinct) eigenvalues of the matrix A by $\lambda_1, \ldots, \lambda_n$, we have $Ax_j = \lambda_j x_j$, so that we simply obtain

$$\begin{aligned} y = Ax &= A\left(c_1x_1 + c_2x_2 + \ldots + c_nx_n\right) \\ &= c_1Ax_1 + c_2Ax_2 + \cdots + c_nAx_n \\ &= c_1\lambda_1x_1 + \cdots + c_n\lambda_nx_n. \end{aligned}$$

This shows that we have decomposed the complicated action of A on an arbitrary vector x into a sum of simple actions (multiplication by scalars) on the eigenvectors of A.

Theorem 16 (Basis of Eigenvectors). *If an $n \times n$ matrix A has n distinct eigenvalues, then A has a basis of eigenvectors $x_1, \ldots, x_n$ for $\mathbb{R}^n$.*

Proof. We just need to prove that the n eigenvectors are linearly independent. If not, reorder the eigenvectors and suppose r of them, $\{x_1, x_2, \ldots, x_r\}$, are linearly independent and $x_{r+1}, \ldots, x_n$ are linearly dependent on $\{x_1, x_2, \ldots, x_r\}$. Consider x_{r+1}. There must exist $c_1, \ldots c_{n+1}$, not all zero, such that

$$c_1x_1 + \ldots c_{r+1}x_{r+1} = 0. \tag{19}$$

Multiplying A on both sides yields

$$c_1Ax_1 + \ldots c_{r+1}Ax_{r+1} = 0.$$

Using $Ax_i = \lambda_i x_i$, we have

$$c_1\lambda_1x_1 + \cdots + c_{r+1}\lambda_{r+1}x_{r+1} = 0.$$

But from (19), we know that

$$c_1\lambda_{r+1}x_1 + \ldots c_{r+1}\lambda_{r+1}x_{r+1} = 0.$$

Subtracting the last two equations gives

$$c_1\left(\lambda_1 - \lambda_{r+1}\right)x_1 + \cdots + c_r\left(\lambda_r - \lambda_{r+1}\right)x_r = 0.$$

None of $\lambda_1 - \lambda_{r+1}, \ldots, \lambda_r - \lambda_{r+1}$ are zero, as the eigenvalues are distinct. Hence not all coefficients $c_1\left(\lambda_1 - \lambda_{r+1}\right), \ldots, c_r\left(\lambda_r - \lambda_{r+1}\right)$ are zero. Thus $\{x_1, x_2, \ldots, x_r\}$ is not linearly independent–a contradiction with the assumption at the beginning of the proof. □

Theorem 16 provides an important decomposition–called diagonalization–of matrices. To show that, we briefly review the concept of matrix inverses first.

Definition 17 (Matrix Inverse). The inverse A^{-1} of a square matrix A satisfies

$$AA^{-1} = A^{-1}A = I.$$

If A^{-1} exists, A is called nonsingular; otherwise, A is singular.

Theorem 18 (Diagonalization of a Matrix)**.** *Let an $n \times n$ matrix A have a basis of eigenvectors $\{x_1, x_2, \ldots, x_n\}$, associated to its n distinct eigenvectors $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, respectively. Then*

$$A = XDX^{-1} = [x_1, x_2, \ldots, x_n] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix} [x_1, x_2, \ldots, x_n]^{-1}. \tag{20}$$

Also,

$$A^m = XD^mX^{-1}, \quad (m = 2, 3, \ldots). \tag{21}$$

Remark 19. From (21), you can find some intuition about the benefit of (20): A^m can be tedious to compute while D^m is very simple!

Proof. From Theorem 16, the n linearly independent eigenvectors of A form a basis. Write

$$\begin{aligned} Ax_1 &= \lambda_1 x_1 \\ Ax_2 &= \lambda_2 x_2 \\ &\vdots \\ Ax_n &= \lambda_n x_n \end{aligned}$$

as

$$A[x_1, x_2, \ldots, x_n] = [x_1, x_2, \ldots, x_n] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}.$$

The matrix $[x_1, x_2, \ldots, x_n]$ is square. Linear independence of the eigenvectors implies that $[x_1, x_2, \ldots, x_n]$ is invertible. Multiplying $[x_1, x_2, \ldots, x_n]^{-1}$ on both sides gives (20).

(21) then immediately follows, as

$$A^m = \left(XDX^{-1}\right)^m = XDX^{-1}XDX \ldots XDX^{-1} = XD^mX^{-1}.$$

□

Example 20. Let

$$A = \begin{bmatrix} 2 & -3 \\ 1 & -2 \end{bmatrix}.$$

The matrix has eigenvalues at 1 and -1, with associated eigenvectors

$$\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Then

$$X = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix},\ A = X \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} X^{-1}.$$

Now if we are to compute A^{3000}. We just need to do

$$A^{3000} = X \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}^{3000} X^{-1} = I.$$

7 Similarity transformation

Definition 21 (Similar Matrices. Similarity Transformation)**.** An $n \times n$ matrix $\hat{A}$ is called similar to an $n \times n$ matrix A if

$$\hat{A} = T^{-1}AT$$

for some **nonsingular** $n \times n$ matrix T. This transformation, which gives $\hat{A}$ from A, is called a similarity transformation.

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two vector spaces of the same dimension. Take the *same* point P. Let u be its coordinate in $\mathcal{S}_1$ and $\hat{u}$ be its coordinate in $\mathcal{S}_2$. These coordinates in the two vector spaces are related by some linear transformation T:

$$u = T\hat{u},\ \hat{u} = T^{-1}u$$

Consider Fig. 4. Let the point P go through a linear transformation A in the vector space $\mathcal{S}_1$ to generate an output point P_o. P_o is physically the same point in both $\mathcal{S}_1$ and $\mathcal{S}_2$. However, the coordinates of P_o are different: if we see it from "standing inside" $\mathcal{S}_1$, then

$$y = Au$$

If we see it in $\mathcal{S}_2$, then the coordinate is some other value $\hat{y}$.

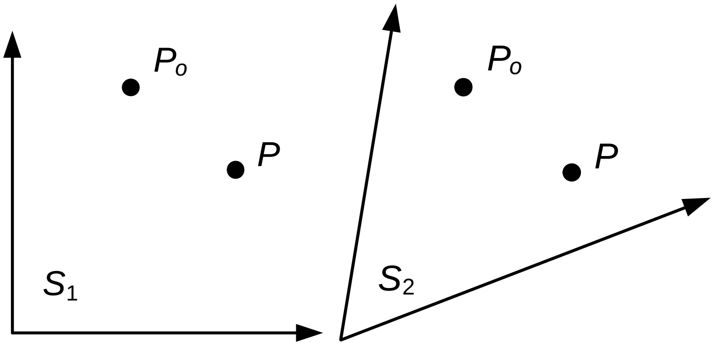


Figure 4: Same points in different vector spaces

How does the linear transformation A mathematically "look like" in $\mathcal{S}_2$?
Result:

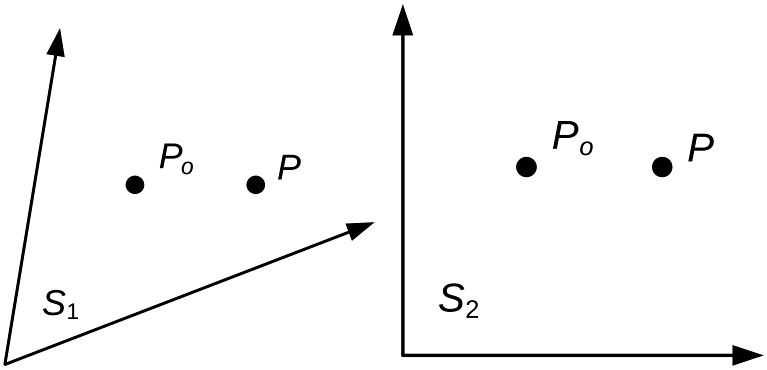
$$\hat{y} = T^{-1}y = T^{-1}Au = \left(T^{-1}AT\right)\hat{u}$$

namely, the linear transformation, viewed from $\mathcal{S}_2$, is

$$\hat{A} = T^{-1}AT$$

It is central to recognize that the physical operation is the same: P goes to another point P_o. Different is our perspective of viewing this transformation. $\hat{A}$ and A are in this sense called similar.

Purpose of doing similarity transformation: $\hat{A}$ can be simpler! Consider, for instance, the following example



In $\mathcal{S}_1$, the transformation changes both coordinates of P while in $\mathcal{S}_2$, only the first coordinate of P is changed.

Theorem 22 (Eigenvalues and Eigenvectors of Similar Matrices). *If $\hat{A}$ is similar to A, then $\hat{A}$ has the same eigenvalues as A. Furthermore, if x is an eigenvector of A, then $y = T^{-1}x$ is an eigenvector of $\hat{A}$ corresponding to the same eigenvalue.*

□

8 Matrix inversion

This section provides a more detailed description of matrix inversion. Recall that the inverse A^{-1} of a square nonsingular matrix A satisfies

$$AA^{-1} = A^{-1}A = I.$$

Theorem 23 (Inverse is unique). *If A has an inverse, the inverse is unique.*

Concepts only. If both B and C are inverses of A, then $BA = AB = I$ and $CA = AC = I$ so that

$$B = IB = (CA)B = CAB = C(AB) = CI = C.$$

Connection with previous topics: The set of all $n \times n$ matrices is not a field. Multiplicative inverse is unique. □

Definition 24 (Existence of a matrix inverse)**.** The inverse A^{-1} of an $n \times n$ matrix A exists if and only if the rank of A is n. Hence A is nonsingular if rank$(A) = n$, and singular if rank$(A) < n$.

Proof. Let $A \in \mathbb{R}^{n\times n}$ and consider the linear equation

$$Ax = b.$$

If A^{-1} exists, then

$$A^{-1}Ax = x = A^{-1}b.$$

Hence $A^{-1}b$ is a solution to the linear equation. It is also unique. If not, then take another solution u; we should have $Au = b$ and $u = A^{-1}b$. Since A^{-1} is unique, it must be that $u = x$.

Conversely, if A has rank n. Then we can solve $Ax = b$ uniquely by Gauss elimination, to get

$$x = Bb,$$

where B is the backward substitution linear transformation in Gauss elimination. Hence

$$Ax = A\left(Bb\right) = \left(AB\right)b = Ib$$

for any b. Hence

$$AB = I.$$

Similarly, substituting $Ax = b$ into $x = Bb$ gives

$$x = B\left(Ax\right) = \left(BA\right)x = Ix,$$

and hence

$$BA = I.$$

Together $B = A^{-1}$ exists. □

There are several ways to compute the inverse of a matrix. One approach for low-order matrices is the method of using adjugate matrix (sometimes also called adjoint matrix):

$$A^{-1} = \frac{1}{\det\left(A\right)}\text{adj}\left(A\right)^T.$$

We explain the computation by two examples. You can find additional details in your undergraduate linear algebra course.

- 2×2 example:

$$\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]^{-1} = \frac{1}{ad - bc}\left[\begin{array}{cc} \left(-1\right)^{1+1}d & \left(-1\right)^{1+2}b \\ \left(-1\right)^{2+1}c & \left(-1\right)^{2+2}a \end{array}\right],$$

  where b in $\left(-1\right)^{1+2}b$ is obtained by:

  - noticing b is at row 1 column 2 of A;
  - looking at the element at row 2 column 1 of A (notice the transpose in adj $\left(A\right)^T$);

  - constructing a submatrix of A by removing row 2 and column 1 from it, i.e., $[b]$ in this 2×2 example;
  - computing the determinant of this submatrix.
  - adding $(-1)^{1+2}$ as a scalar

- 3×3 example:

$$A^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \frac{1}{\det A} \begin{bmatrix} \begin{vmatrix} e & f \\ h & k \end{vmatrix} & -\begin{vmatrix} b & c \\ h & k \end{vmatrix} & \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ -\begin{vmatrix} d & f \\ g & k \end{vmatrix} & \begin{vmatrix} a & c \\ g & k \end{vmatrix} & -\begin{vmatrix} a & c \\ d & f \end{vmatrix} \\ \begin{vmatrix} d & e \\ g & h \end{vmatrix} & -\begin{vmatrix} a & b \\ g & h \end{vmatrix} & \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{bmatrix},$$

where $|\cdot|$ denotes the determinant of a matrix. Similar as before, the row 1 column 2 element $-\begin{vmatrix} b & c \\ h & k \end{vmatrix}$ is obtained via

$$(-1)^{2+1} \det \left(A \text{ with } [d, e, f], \begin{bmatrix} a \\ d \\ g \end{bmatrix} \text{ removed} \right).$$

Example 25. Find the inverse matrices of

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}, \ B = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}, \ C = \begin{bmatrix} -0.5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The answers are:

$$A^{-1} = \begin{bmatrix} 0.4 & -0.1 \\ -0.2 & 0.3 \end{bmatrix}, \ B^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ -1 & 3 & 4 \end{bmatrix}, \ C^{-1} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The related MATLAB command for matrix inversion is *inv()*.

Theorem 26. *Inverse of products of matrices can be obtained from inverses of each factor:*

$$(AB)^{-1} = B^{-1}A^{-1},$$

and more generally

$$(AB \dots YZ)^{-1} = Z^{-1}Y^{-1} \dots B^{-1}A^{-1}. \tag{22}$$

Proof. By definition $(AB)(AB)^{-1} = I$. Multiplying A^{-1} on both sides from the left gives

$$B(AB)^{-1} = A^{-1}.$$

Now multiplying the result by B^{-1} on both sides from the left, we get

$$(AB)^{-1} = B^{-1}A^{-1}.$$

The general case (22) follows by induction. □

Fact 27. **Inverse of upper (lower) triangular matrices are upper (lower) triangular*

Proof. (main idea) We can either use the adjoint matrix method or use the following decomposition of upper(lower) triangular matrices

$$A = D(I + N),$$

where D is diagonal and N is strictly upper (lower) triangular with zeros diagonal elements. Then using matrix Taylor expansion we have

$$\begin{aligned} A^{-1} &= (I + N)^{-1} D^{-1} \\ &= \left(I - N + N^2 - N^3 + N^4 - \dots\right) D^{-1}. \end{aligned}$$

N is nilpotent: N^k are upper (lower) triangular and $N^n = 0$ for n larger than the row dimension of A. D^{-1} is diagonal. Hence A^{-1} is upper (lower) triangular. □

8.1 Block matrix decomposition and inversion

Consider

$$A = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.$$

Recall the key step in performing row operations on matrices in Gauss elimination:

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \to \begin{bmatrix} 3 & 4 \\ 0 & 2/3 \end{bmatrix},$$

where we had substracted one third of the first row in the second row. In matrix representations, the above looks like

$$\begin{bmatrix} 1 & 0 \\ -1/3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 2/3 \end{bmatrix}.$$

For more general two by two matrices, we have

$$\begin{bmatrix} 1 & 0 \\ -ca^{-1} & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & d - ca^{-1}b \end{bmatrix}.$$

If we want to keep the second row unchanged and simplify the first row, we can do

$$\begin{bmatrix} 1 & -bd^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a - bd^{-1}c & 0 \\ c & d \end{bmatrix}.$$

Generalizing the concept to blok matrices (with compatible dimensions), we have

$$\left[\begin{array}{cc} I & 0 \\ -B^TA^{-1} & I \end{array}\right]\left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right]=\left[\begin{array}{cc} A & B \\ 0 & C-B^TAB \end{array}\right],$$

and

$$\left[\begin{array}{cc} A & B \\ 0 & C-B^TAB \end{array}\right]\left[\begin{array}{cc} I & -A^{-1}B \\ 0 & I \end{array}\right]=\left[\begin{array}{cc} A & 0 \\ 0 & C-B^TAB \end{array}\right].$$

Thus

$$\left[\begin{array}{cc} I & 0 \\ -B^TA^{-1} & I \end{array}\right]\left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right]\left[\begin{array}{cc} I & -A^{-1}B \\ 0 & I \end{array}\right]=\left[\begin{array}{cc} A & 0 \\ 0 & C-B^TAB \end{array}\right].$$

Inversion is now very easy:

$$\left\{\left[\begin{array}{cc} I & 0 \\ -B^TA^{-1} & I \end{array}\right]\left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right]\left[\begin{array}{cc} I & -A^{-1}B \\ 0 & I \end{array}\right]\right\}^{-1}=\left[\begin{array}{cc} A & 0 \\ 0 & C-B^TAB \end{array}\right]^{-1}$$

$$\Longrightarrow\left[\begin{array}{cc} I & -A^{-1}B \\ 0 & I \end{array}\right]^{-1}\left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right]^{-1}\left[\begin{array}{cc} I & 0 \\ -B^TA^{-1} & I \end{array}\right]^{-1}=\left[\begin{array}{cc} A & 0 \\ 0 & C-B^TAB \end{array}\right]^{-1},$$

and hence

$$\begin{aligned}\left[\begin{array}{cc} A & B \\ B^T & C \end{array}\right]^{-1}&=\left[\begin{array}{cc} I & -A^{-1}B \\ 0 & I \end{array}\right]\left[\begin{array}{cc} A & 0 \\ 0 & C-B^TAB \end{array}\right]^{-1}\left[\begin{array}{cc} I & 0 \\ -B^TA^{-1} & I \end{array}\right]\\&=\left[\begin{array}{cc} I & -A^{-1}B \\ 0 & I \end{array}\right]\left[\begin{array}{cc} A^{-1} & 0 \\ 0 & \left(C-B^TAB\right)^{-1} \end{array}\right]\left[\begin{array}{cc} I & 0 \\ -B^TA^{-1} & I \end{array}\right].\end{aligned}$$

The above steps work for general partitioned 2 by 2 matrices as well. The result is as follows

$$\left[\begin{array}{cc} I & 0 \\ -CA^{-1} & I \end{array}\right]\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\left[\begin{array}{cc} I & -BA^{-1} \\ 0 & I \end{array}\right]=\left[\begin{array}{cc} A & 0 \\ 0 & D-CA^{-1}B \end{array}\right]$$

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]^{-1}=\left[\begin{array}{cc} I & -BA^{-1} \\ 0 & I \end{array}\right]\left[\begin{array}{cc} A & 0 \\ 0 & D-CA^{-1}B \end{array}\right]^{-1}\left[\begin{array}{cc} I & 0 \\ -CA^{-1} & I \end{array}\right],$$

or

$$\left[\begin{array}{cc} I & -BD^{-1} \\ 0 & I \end{array}\right]\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]\left[\begin{array}{cc} I & 0 \\ -D^{-1}C & I \end{array}\right]=\left[\begin{array}{cc} A-BD^{-1}C & 0 \\ 0 & D \end{array}\right]$$

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]^{-1}=\left[\begin{array}{cc} I & -BD^{-1} \\ 0 & I \end{array}\right]\left[\begin{array}{cc} A-BD^{-1}C & 0 \\ 0 & D \end{array}\right]^{-1}\left[\begin{array}{cc} I & 0 \\ -D^{-1}C & I \end{array}\right].$$

8.2 *LU and Cholesky decomposition

Fact 28. *The following is true for upper and lower triangular matrices:*

$$\left[\begin{array}{cc} I & 0 \\ M & I \end{array}\right]^{-1}=\left[\begin{array}{cc} I & 0 \\ -M & I \end{array}\right]$$

$$\left[\begin{array}{cc} I & M \\ 0 & I \end{array}\right]^{-1}=\left[\begin{array}{cc} I & -M \\ 0 & I \end{array}\right].$$

From the last section

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -BA^{-1} \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix}.$$

Applying Fact 28 to the last equation gives the *block LU decomposition*:

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ 0 & D-CA^{-1}B \end{bmatrix}, \end{aligned}$$

which shows *any square matrix can be decomposed into the product of a lower triangular matrix and an upper triangular matrix.*

There is also *block Cholesky decomposition*

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I \\ CA^{-1} \end{bmatrix} A \begin{bmatrix} I & A^{-1}B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & D-CA^{-1}B \end{bmatrix},$$

or using half matrices

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} A^{\frac{1}{2}} \\ CA^{-\frac{*}{2}} \end{bmatrix} \begin{bmatrix} A^{\frac{*}{2}} & A^{-\frac{1}{2}}B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{*}{2}} \end{bmatrix} \\ Q &= D - CA^{-1}B, \end{aligned}$$

where

$$A^{\frac{1}{2}} A^{\frac{*}{2}} = A, \; Q^{\frac{1}{2}} Q^{\frac{*}{2}} = Q.$$

Hence

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = LU,$$

where

$$LU = \begin{bmatrix} A^{\frac{1}{2}} & 0 \\ CA^{-\frac{*}{2}} & 0 \end{bmatrix} \begin{bmatrix} A^{\frac{*}{2}} & A^{-\frac{1}{2}}B \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & Q^{\frac{*}{2}} \end{bmatrix} = \begin{bmatrix} A^{\frac{1}{2}} & 0 \\ CA^{-\frac{*}{2}} & Q^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} A^{\frac{*}{2}} & A^{-\frac{1}{2}}B \\ 0 & Q^{\frac{*}{2}} \end{bmatrix}.$$

8.3 Determinant and matrix inverse identity

Although $AB \neq BA$ in general, the determinants of products have the following property:

$$\det(AB) = \det(BA) = \det A \det B,$$

where A and B should be square to start with.

Theorem 29 (Sylvester's determinant theorem). *For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$,*

$$\det(I_m + AB) = \det(I_n + BA),$$

where I_m and I_n are the $m \times m$ and $n \times n$ identity matrices, respectively.

Proof. Construct

$$M = \begin{bmatrix} I_m & -A \\ B & I_n \end{bmatrix}.$$

From the decomposition

$$M = \begin{bmatrix} I_m & 0 \\ B & I_n \end{bmatrix} \begin{bmatrix} I_m & -A \\ 0 & I_n + BA \end{bmatrix},$$

we have

$$\det M = \det\left(I_n + BA\right).$$

Alternatively

$$M = \begin{bmatrix} I_m + AB & -A \\ 0 & I_n \end{bmatrix} \begin{bmatrix} I_m & 0 \\ B & I_n \end{bmatrix}.$$

Hence

$$\det M = \det\left(I_m + AB\right).$$

Therefore

$$\det\left(I_m + AB\right) = \det M = \det\left(I_n + BA\right).$$

□

More generally, for any invertible $m \times m$ matrix X

$$\det\left(X + AB\right) = \det\left(X\right)\det\left(I_n + BX^{-1}A\right),$$

which comes from

$$\begin{aligned} X + AB &= X\left(I + X^{-1}AB\right) \\ \Rightarrow \det\left(X + AB\right) &= \det\left[X\left(I + X^{-1}AB\right)\right] = \det X \det\left(I + X^{-1}AB\right). \end{aligned}$$

8.4 Matrix inversion lemma

Fact 30 (Matrix inversion lemma). *Assume A is nonsingular and $(A + BC)^{-1}$ exists. The following is true*

$$(A + BC)^{-1} = A^{-1}\left(I - B\left(CA^{-1}B + I\right)^{-1}CA^{-1}\right). \tag{23}$$

Proof. Consider

$$(A + BC)\,x = y. \tag{24}$$

We aim at getting

$$x = (*)\,y, \text{ where } (*) \text{ will be our } (A + BC)^{-1}. \tag{25}$$

First, let

$$Cx = d. \tag{26}$$

Equation (24) can be written as

$$\begin{aligned} Ax + Bd &= y \\ Cx - d &= 0. \end{aligned}$$

Solving the first equation yields

$$x = A^{-1}(y - Bd). \tag{27}$$

Then (26) becomes

$$CA^{-1}(y - Bd) = d.$$

Combining the terms about d and applying matrix inversion yield

$$d = \left(CA^{-1}B + I\right)^{-1} CA^{-1}y.$$

Putting the result in (27) yields

$$\begin{aligned} x &= A^{-1}\left(y - B\left(CA^{-1}B + I\right)^{-1} CA^{-1}y\right) \\ &= A^{-1}\left(I - B\left(CA^{-1}B + I\right)^{-1} CA^{-1}\right) y. \end{aligned}$$

Comparing the above with (25), we obtain (23). □

Exercise 31. The matrix inversion lemma is a powerful tool useful for many applications. One application in adaptive control and system identification uses

$$\left(A + \phi\phi^T\right)^{-1} = A^{-1}\left(I - \frac{\phi\phi^T A^{-1}}{\phi^T A^{-1}\phi + 1}\right).$$

Prove the above result. Prove also the general case (called rank one update):

$$\left(A + bc^T\right) = A^{-1} - \frac{1}{1 + c^T A^{-1} b}\left(A^{-1}b\right)\left(c^T A^{-1}\right).$$

Fact 32 (More extended matrix inversion lemma). *Assume A, C, and $A + BCB^T$ are nonsingular. The following is true*

$$\begin{aligned} \left(A + BCB^T\right)^{-1} &= A^{-1}\left(I - B\left(CB^T A^{-1}B + I\right)^{-1} CB^T A^{-1}\right) & (28) \\ &= A^{-1} - A^{-1}B\left(CB^T A^{-1}B + I\right)^{-1} CB^T A^{-1} & (29) \\ &= A^{-1} - A^{-1}B\left(B^T A^{-1}B + C^{-1}\right)^{-1} B^T A^{-1}. & (30) \end{aligned}$$

Proof. Consider

$$\left(A + BCB^T\right)x = y.$$

We aim at getting $x = (*)\,y$, where $(*)$ will be our $\left(A + BCB^T\right)^{-1}$. First, let

$$CB^T x = d.$$

We have

$$\begin{aligned} Ax + Bd &= y \\ CB^T x - d &= 0. \end{aligned}$$

Solving the first equation yields

$$x = A^{-1}\left(y - Bd\right).$$

Then

$$CB^TA^{-1}\left(y - Bd\right) = d$$

gives

$$d = \left(CB^TA^{-1}B + I\right)^{-1}CB^TA^{-1}y.$$

Hence

$$\begin{aligned} x &= A^{-1}\left(y - B\left(CB^TA^{-1}B + I\right)^{-1}CB^TA^{-1}y\right) \\ &= A^{-1}\left(I - B\left(CB^TA^{-1}B + I\right)^{-1}CB^TA^{-1}\right)y \end{aligned}$$

and (28) follows. □

The extended matrix inversion lemma is key in transforming the Kalman filter to the information filter when inverting the innovation of covariance matrices.

8.5 Special inverse equalities

Fact 33. *The following matrix equalities are true*

- $(I + GK)^{-1}G = G(I + KG)^{-1}$
  to prove the result, start with $G(I + KG) = (I + GK)G$
- $GK(I + GK)^{-1} = G(I + KG)^{-1}K = (I + GK)^{-1}GK$ (the proof uses the first equality twice)
- generalization 1: $(\sigma^2 I + GK)^{-1}G = G(\sigma^2 I + KG)^{-1}$
- generalization 2: if M is invertible then $(M + GK)^{-1}G = M^{-1}G(I + KM^{-1}G)^{-1}$

Exercise 34. Check validity of the following equality, assuming proper dimensions and invertibility of matrices:

- $Z(I + Z)^{-1} = I - (I + Z)^{-1}$
- $(I + XY)^{-1} = I - XY(I + XY)^{-1} = I - X(I + YX)^{-1}Y$
- extension

$$\begin{aligned} \left(I + XZ^{-1}Y\right)^{-1} &= I - XZ^{-1}Y\left(I + XZ^{-1}Y\right)^{-1} = I - XZ^{-1}\left(I + YXZ^{-1}\right)^{-1}Y \\ &= I - X(Z + YX)^{-1}Y \end{aligned}$$

9 Spectral mapping theorem

Theorem 35 (Spectral Mapping Theorem). *Take any $A \in \mathbb{C}^{n\times n}$ and a polynomial (in s) $f(s)$ (more generally, analytic functions). Then*

$$\text{eig}\left(f\left(A\right)\right) = f\left(\text{eig}\left(A\right)\right).$$

Proof. Let

$$f\left(A\right) = x_0 I + x_1 A + x_2 A^2 + \dots.$$

Let λ be an eigenvalue of A. We first observe that λ^n is an eigenvalue of A^n. This can be seen from $\det\left(\lambda^n I - A^n\right) = \det\left[\left(\lambda I - A\right) p\left(A\right)\right] = \det\left(\lambda I - A\right)\det\left(p\left(A\right)\right)$ where $p\left(A\right)$ is a polynomial of A.

Now consider $f\left(\lambda\right) = x_0 + x_1\lambda + x_2\lambda^2 + \dots$. We have

$$\begin{aligned}\det\left(f\left(\lambda\right)I - f\left(A\right)\right) &= \det\left[x_1\left(\lambda I - A\right) + x_2\left(\lambda^2 I - A^2\right) + x_3\left(\lambda^3 I - A^3\right) + \dots\right]\\ &= \det\left[\left(\lambda I - A\right) q\left(A\right)\right]\\ &= \det\left(\lambda I - A\right)\det\left(q\left(A\right)\right).\end{aligned}$$

Hence $f\left(\lambda\right)$ is an eigenvalue of $f\left(A\right)$.

Conversely, if γ is an eigenvalue of $f\left(A\right)$, we need to prove that γ is in the form of $f\left(\lambda\right)$. Factorize the polynomial

$$f\left(\lambda\right) - \gamma = a_0\left(\lambda - \alpha_1\right)\left(\lambda - \alpha_2\right)\dots\left(\lambda - \alpha_n\right).$$

On the other hand, we note that as a matrix polynomial with the same coefficients:

$$f\left(A\right) - \gamma I = a_0\left(A - \alpha_1 I\right)\left(A - \alpha_2 I\right)\dots\left(A - \alpha_n I\right).$$

But $\det\left(f\left(A\right) - \gamma I\right) = 0$, which means that there is at least one α_i such that

$$\det\left(A - \alpha_i I\right) = 0,$$

which says that α_i is an eigenvalue of A. Hence

$$f\left(\lambda\right) - \gamma = a_0\left(\lambda - \alpha_i\right)\prod_{k\neq i}\left(\lambda - \alpha_k\right) = 0,$$

i.e.

$$\gamma = f\left(\lambda\right),$$

where λ is an eigenvalue of A. □

Example 36. Compute the eigenvalues of

$$A = \begin{bmatrix} 99.8 & 2000 \\ -2000 & 99.8 \end{bmatrix}.$$

Solution:

$$A = 99.8I + 2000\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

So

$$\text{eig}(A) = 99.8 + 2000\,\text{eig}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = 99.8 \pm 2000i.$$

10 Matrix exponentials

Since the Taylor series

$$e^{st} = 1 + st + \frac{s^2t^2}{2!} + \frac{s^3t^3}{3!} + \dots$$

converges everywhere, we can define the exponential of a matrix $A \in \mathcal{C}^{n\times n}$ by

$$e^{At} = I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \dots.$$

Fact 37. *Properties of matrix exponentials*

1. $e^{A0} = I$
2. $e^{A(t+s)} = e^{At}e^{As}$
3. *If* $AB = BA$ *then* $e^{(A+B)t} = e^{At}e^{Bt} = e^{Bt}e^{At}$
4. $\det\left(e^{At}\right) = e^{\textit{trace}(A)t}$
5. e^{At} *is nonsingular for all* $t \in \mathcal{R}$ *and* $\left(e^{At}\right)^{-1} = e^{-At}$
6. e^{At} *is the unique solution* X *of the linear system of ordinary differential equations*

$$\dot{X} = AX, \text{ subject to } X(0) = I$$

11 Inner product

11.1 Inner product spaces

Basics: Inner product, or dot product, brings a metric for vector lengths. It takes two vectors and generates a number. In $\mathbb{R}^n$, we have

$$\langle a, b\rangle \triangleq a^T b = [a_1, a_2, \dots, a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Clearly, $\langle a, b\rangle \triangleq a^T b = \langle b, a\rangle$. Letting $b = a$ above, we get the square of the length of a:

$$||a|| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

Formal definitions:

Definition 38. A real vector space $\mathbf{V}$ is called a real inner product space, if for any vectors a and b in $\mathbf{V}$ there is an associated real number $\langle a, b\rangle$, called the inner product of a and b, such that the following axioms hold:

- (linearity) For all scalars q_1 and q_2 and all vectors $a, b, c \in \mathbf{V}$

$$\langle q_1 a + q_2 b, c\rangle = q_1 \langle a, b\rangle + q_2 \langle b, c\rangle$$

- (symmetry) $\forall a, b \in \mathbf{V}$

$$\langle a, b\rangle = \langle b, a\rangle$$

- (positive definiteness) $\forall a \in \mathbf{V}$

$$\langle a, a\rangle \geq 0$$

  where $\langle a, a\rangle = 0$ if and only if $a = 0$.

Definition 39 (2-norm of vectors). The length of a vector in $\mathbf{V}$ is defined by

$$||a|| = \sqrt{\langle a, a\rangle} \geq 0.$$

For $\mathbb{R}^n$,

$$||a|| = \sqrt{a^T a} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

This is the Euclidean norm or 2-norm of the vector. $\mathbb{R}^n$ equiped with the inner product $\langle a, b\rangle = \sqrt{a^T b}$ is called the n-dimensional Euclidean space.

Example 40 (Inner product for functions, function spaces)**.** The set of all real-valued continuous functions $f(x)$, $g(x)$, $\ldots$ $x \in [\alpha, \beta]$ is a real vector space under the usual addition of functions and multiplication by scalars. An inner product on this function space is

$$\langle f, g \rangle = \int_\alpha^\beta f(x)\, g(x)\, dx$$

and the norm of f is

$$||f(x)|| = \sqrt{\int_\alpha^\beta f(x)^2\, dx}.$$

Inner products is also closely related to the geometric relationships between vectors. For the two-dimensional case, it is readily seen that

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

is a basis of the vector space. The two vectors are additionally orthogonal, by direct observation.

More generally, we have:

Definition 41 (Orthogonal vectors)**.** Vectors whose inner product is zero are called orthogonal.

Definition 42 (Orthonormal vectors)**.** Orthogonal vectors with unity norm is called orthonormal.

Definition 43. The angle between two vectors is defined by

$$\cos\angle(a, b) = \frac{\langle a, b \rangle}{||a|| \cdot ||b||} = \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle} \cdot \sqrt{\langle b, b \rangle}}.$$

11.2 Trace (standard matrix inner product)

The trace of an $n \times n$ matrix $A = [a_{jk}]$ is given by

$$\mathrm{Tr}(A) = \sum_{i=1}^n a_{ii}. \tag{31}$$

Trace defines the so-called **matrix inner product**:

$$\langle A, B \rangle = \mathrm{Tr}\left(A^T B\right) = \mathrm{Tr}\left(B^T A\right) = \langle B, A \rangle, \tag{32}$$

which is closely related to vector inner products. Take an example in $\mathbb{R}^{3\times 3}$: write the matrices in the column-vector form $B = [\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3]$, $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$, then

$$A^T B = \begin{bmatrix} \mathbf{a}_1^T \mathbf{b}_1 & * & * \\ * & \mathbf{a}_2^T \mathbf{b}_2 & * \\ * & * & \mathbf{a}_3^T \mathbf{b}_3 \end{bmatrix}. \tag{33}$$

So

$$\operatorname{Tr}\left(A^T B\right) = \mathbf{a}_1^T \mathbf{b}_1 + \mathbf{a}_2^T \mathbf{b}_2 + \mathbf{a}_3^T \mathbf{b}_3,$$

which is nothing but the inner product of the two "stacked" vectors $\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix}$. Hence

$$\langle A, B \rangle = \operatorname{Tr}\left(A^T B\right) = \left\langle \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}, \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \right\rangle.$$

Exercise 44. If x is a vector, show that

$$\operatorname{Tr}(xx^T) = x^T x.$$

12 Norms

Previously we have used $||\cdot||$ to denote the Euclidean length function. At different times, it is useful to have more general notions of size and distance in vector spaces. This section is devoted to such generalizations.

12.1 Vector norm

Definition 45. A *norm* is a function that assigns a real-valued length to each vector in a vector space $\mathbb{C}^m$. To develop a reasonable notion of length, a norm must satisfy the following properties: for any vectors a, b and scalars $\alpha \in \mathbb{C}$,

- the norm of a nonzero vector is positive: $||a|| \geq 0$, and $||a|| = 0$ if and only if $a = 0$
- scaling a vector scales its norm by the same amount: $||\alpha a|| = |\alpha| \, ||a||$
- triangle inequality: $||a + b|| \leq ||a|| + ||b||$

Let w_1 be a $n \times 1$ vector. The most important class of vector norms, the p norms, of w are defined by

$$||w||_p = \left(\sum_{i=1}^{n} |w_i|^p \right)^{1/p}, \ 1 \leq p < \infty.$$

Specifically, we have

$\|w\|_1 = \sum_{i=1}^n |w_j|$ (absolute column sum)

$\|w\|_\infty = \max_i |w_i|$

$\|w\|_2 = \sqrt{w^H w}$ (Euclidean norm)

Remark 46. When unspecified, $||\cdot||$ refers to 2 norm in this set of notes.

Intuitions for the infinity norm By definition

$$||w||_\infty = \lim_{p\to\infty}\left(\sum_{i=1}^{n}|w_i|^p\right)^{1/p}.$$

Intuitively, as p increases, $\max_i |w_i|$ takes more and more weighting in $\sum_{i=1}^n |w_i|^p$. More rigorously, we have

$$\lim_{p\to\infty}\left((\max|w_i|)^p\right)^{1/p} \le \lim_{p\to\infty}\left(\sum_{i=1}^{n}|w_i|^p\right)^{1/p} \le \lim_{p\to\infty}\left(\sum_{i=1}^{n}(\max|w_i|)^p\right)^{1/p}.$$

Both $\lim_{p\to\infty}\left((\max|w_i|)^p\right)^{1/p}$ and $\lim_{p\to\infty}\left(\sum_{i=1}^n (\max|w_i|)^p\right)^{1/p}$ equals $\max_i |w_i|$. Hence $||w||_\infty = \max|w_i|$.

12.2 Induced matrix norm

As matrices define linear transformations between vector spaces, it makes sense to have a measure of the "size" of the transformation. Induced matrix norms[2] are defined by

$$||M||_{p\leftarrow q} = \max_{x\neq 0}\frac{||Mx||_p}{||x||_q}. \tag{34}$$

In other words, $||M||_{q\leftarrow q}$ is the maximum factor by which M can "stretch" a vector x.

In particular, the following matrix norms are common:

$\|M\|_{1\leftarrow 1} = \max_j \sum_{i=1}^n |M_{ij}|$ maximum absolute column sum

$\|M\|_{\infty\leftarrow\infty} = \max_i \sum_{j=1}^m |M_{ij}|$ maximum absolute row sum

$\|M\|_{2\leftarrow 2} = \sqrt{\lambda_{\max}(M^*M)}$ maximum singular value

The induced 2 norm can be understood as follows:

$$||M||_{2\leftarrow 2} = \max_{x\neq 0}\frac{||Mx||_2}{||x||_2} = \max_{x\neq 0}\sqrt{\frac{x^*M^*Mx}{\langle x,x\rangle^2}} = \sqrt{\lambda_{\max}(M^*M)}.$$

Remark 47. When $p = q$ in (34), often the induced matrix norm is simply written as $||M||_p$.

12.3 Frobenius norm and general matrix norms

Matrix norms do not have to be induced by vector norms.

[2] It is 'induced' from other vector norms as shown in the definition.

Formal definition: Let $\mathcal{M}_n$ be the set of all $n \times n$ real- or complex-valued matrices. We call a function $||\cdot|| : \mathcal{M}_n \to \mathbb{R}$ a matrix norm if for all $A, B \in \mathcal{M}_n$ it satisfies the following axioms:

1. $||A|| \geq 0$
2. $\|A\| = 0$ if and only if $A = 0$
3. $||cA|| = |c|||A||$ for all complex scalars c
4. $\|A + B\| \leq \|A\| + \|B\|$
5. $\|AB\| \leq \|A\|\|B\|$

The formal definition of matrix norms is slightly amended from vector norms. This is because although $\mathcal{M}_n$ is itself a vector space of dimension n^2, it has a natural multiplication operation that is absent in regular vector spaces. A vector norm on matrices that satisfies the first four axioms and not necessarily axiom 5 is often called a generalized matrix norm.

Frobenius norm: The most important matrix norm which is not induced by a vector norm is the Frobenius norm, defined by

$$||A||_F \triangleq \sqrt{\mathrm{Tr}\,(A^*A)} = \sqrt{< A, A >} = \sqrt{\sum_{i,j} |a_{i,j}|^2}.$$

Frobenius norm is just the Euclidean norm of the matrix, written out as a long column vector:

$$||A||_F = (\mathrm{Tr}\,(A^*A))^{\frac{1}{2}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{m} |a_{i,j}|^2\right)^{\frac{1}{2}}.$$

We also have bounds for Frobenius norms:

$$||AB||_F^2 \leq ||A||_F^2 ||B||_F^2.$$

Transforming from one matrix norm to another:

Theorem 48. *If $||\cdot||$ is a matrix norm on $\mathcal{M}_n$ and if $S \in \mathcal{M}_n$ is nonsingular, then*

$$||A||_S = ||S^{-1}AS|| \ \forall A \in \mathcal{M}_n$$

is a matrix norm.

Exercise 49. Prove Theorem 48.

12.4 Norm inequalities

1. Cauchy-Schwartz Inequality:

$$|\langle x,y\rangle| \le ||x||_2||y||_2,$$

which is the special case of the Holder inequality

$$|\langle x,y\rangle| \le ||x||_p||y||_q,\ \frac{1}{p}+\frac{1}{q}=1,\ 1\le p,q\le\infty. \tag{35}$$

Both bounds are tight: for certain choices of x and y, the inequalities become equalities.

2. Bounding induced matrix norms:

$$||AB||_{l\leftarrow n} \le ||A||_{l\leftarrow m}||B||_{m\leftarrow n}, \tag{36}$$

which comes from

$$||ABx||_l \le ||A||_{l\leftarrow m}||Bx||_m \le ||A||_{l\leftarrow m}||B||_{m\leftarrow n}||x||_n.$$

In general, the bound is not tight. For instance, $||A^n||=||A||^n$ does not hold for $n\ge 2$ unless A has special structures.

3. (35) and (36) are useful for computing bounds of difficult-to-compute norms. For instance, $||A||_2^2$ is expensive to compute but $||A||_1$ and $||A||_\infty$ are not. As a special case of (36), we have

$$||A||_2^2 \le ||A||_1||A||_\infty.$$

We can obtain an upper bound of $||A||_2^2$ by computing $||A||_1||A||_\infty$.

4. Any matrix induced norms of A are larger than the absolute eigenvalues of A:

$$|\lambda(A)| \le ||A||_p.$$

Hence as a special case, the spectral radius is upper bounded by any matrix norms:

$$\rho(A) \le \|A\|.$$

5. Let $A\in\mathcal{M}_n$ and $\epsilon>0$ be given. There is a matrix norm such that

$$\rho(A) \le \|A\| \le \rho(A)+\epsilon.$$

Hint: A can be decomposed as $A=U^*\Delta U$ where U is unitary and Δ is upper triangular [Schur triangulariztion theorem]. Let $D_t=\text{diag}(t,t^2,\dots,t^n)$ and compute

$$D_t\Delta D_t^{-1}=\begin{bmatrix} \lambda_1 & t^{-1}d_{12} & \dots & & \dots & t^{-n+1}d_{1n}\\ 0 & \lambda_2 & t^{-1}d_{23} & & \dots & t^{-n+2}d_{1n}\\ \vdots & \ddots & \lambda_3 & \ddots & & \vdots\\ & & \ddots & \ddots & \ddots & \vdots\\ \vdots & & & \ddots & \ddots & t^{-1}d_{n-1,n}\\ 0 & \dots & & \dots & 0 & \lambda_n \end{bmatrix}.$$

For t large enough, the sum of the absolute values of the off-diagonal entries is less than ϵ and in particular

$$\|D_t\Delta D_t^{-1}\|_1 \le \rho(A)+\epsilon.$$

12.5 Exercises

1. Let x be an m vector and A be an $m \times n$ matrix. Verify each of the following inequalities, and give an example when the equality is achieved.

   (a) $||x||_\infty \leq ||x||_2$

   (b) $||x||_2 \leq \sqrt{m}||x||_\infty$

   (c) $||A||_\infty \leq \sqrt{n}||A||_2$

   (d) $||A||_2 \leq \sqrt{m}||A||_\infty$

2. Let x be a random vector with mean $\mathsf{E}\left[x\right] = 0$ and covariance $\mathsf{E}\left(xx^T\right) = I$, then

$$\|A\|_F^2 = \mathrm{E}\left[\|Ax\|_2^2\right].$$

   Hint: use Exercise 44.

13 Symmetric, skew-symmetric, and orthogonal matrices

13.1 Definitions and basic properties

A real square matrix A is called **symmetric** if $A = A^T$, **skew-symmetric** if $A = -A^T$.

Fact 50. *Any real square matrix A may be written as the sum of a symmetric matrix R and a skew-symmetric matrix S, where*

$$R = \frac{1}{2}\left(A + A^T\right),\ S = \frac{1}{2}\left(A - A^T\right).$$

If $A = [a_{jk}]$, then the **complex conjugate** of A is denoted as$\overline{A} = [\overline{a}_{jk}]$, i.e., each element $a_{jk} = \alpha + i\beta$ is replaced with its complex conjugate $\overline{a}_{jk} = \alpha - i\beta$.

A square matrix A is called **Hermitian** if $A^T = \overline{A}$; **skew-Hermitian** if $A^T = -\overline{A}$.

Example 51. Find the symmetric, skew-symmetric, Hermitian, and skew-Hermitian matrices in the set:

$$\left\{ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2i \\ 2i & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2i \\ -2i & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2+2i \\ 2-2i & 0 \end{bmatrix} \right\}.$$

We introduce one more class of important matrices: a real square matrix A is called **orthogonal**[3] if

$$A^TA = AA^T = I. \tag{37}$$

Writing A in the column-vector notation

$$A = [a_1, a_2, \ldots, a_n],$$

[3] Some people also call use the notion of orthonormal matrix. But orthogonal matrix is more often used (we can say orthonormal basis).

we get the equivalent form of (37):

$$A^TA = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} a_1, & a_2, & \ldots, & a_n \end{bmatrix} = \begin{bmatrix} a_1^Ta_1 & a_1^Ta_2 & \ldots & a_1^Ta_n \\ a_2^Ta_1 & a_2^Ta_2 & \ldots & a_2^Ta_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n^Ta_1 & a_n^Ta_2 & \ldots & a_n^Ta_n \end{bmatrix} = I.$$

Hence it must be that

$$\begin{aligned} a_j^Ta_j &= 1 \\ a_j^Ta_m &= 0 \ \forall j \neq m, \end{aligned}$$

namely, a_j and a_m are orthonormal for any $j \neq m$.

The complex version of an orthogonal matrix is the **unitary matrix**. A square matrix A is called unitary if $A\overline{A}^T = \overline{A}^TA = I$, namely $A^{-1} = \overline{A}^T$.

Remark 52. Often the complex conjugate transpose $\overline{A}^T$ is written as A^*.

Theorem 53. *The eigenvalues of symmetric matrices are all real.*

Proof. $\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = A$. $Au = \lambda u \Rightarrow \overline{u}^TAu = \lambda\overline{u}^Tu$, where $\overline{u}$ is the complex conjugate of u. $\overline{u}^TAu$ is a real number, as

$$\begin{aligned} \overline{\overline{u}^TAu} &= u^T\overline{A}\overline{u} \\ &= u^TA\overline{u} \quad \because A \in \mathbb{R}^{n\times n} \\ &= u^TA^T\overline{u} \quad \because A = A^T \\ &= \lambda u^T\overline{u} \quad \because (Au)^T = (\lambda u)^T \\ &= \lambda\overline{u}^Tu \quad \because u^T\overline{u} \in \mathbb{R} \\ &= \overline{u}^TAu \quad \because Au = \lambda u. \end{aligned}$$

By definition of complex conjugate numbers, $\overline{u}^Tu \in \mathbb{R}$. So $\lambda = \frac{\overline{u}^TAu}{\overline{u}^Tu}$ is also a real number. □

Theorem 54. *The eigenvalues of skew-symmetric matrices are all imaginary or zero.*

The proof is left as an exercise.

Fact 55. *An orthogonal transformation preserves the value of the inner product of vectors a and b in $\mathbb{R}^n$. That is, for any a and b in $\mathbb{R}^n$, orthogonal $n \times n$ matrix A, and $u = Aa$, $v = Ab$ we have $\langle u, v\rangle = \langle a, b\rangle$, as*

$$u^Tv = a^TA^TAb = a^Tb.$$

Hence the transformation also preserves the length or 2-norm of any vector a in $\mathbb{R}^n$ given by $||a||_2 = \sqrt{\langle a, a\rangle}$.

Theorem 56. *The determinant of an orthogonal matrix is either 1 or -1.*

Proof. $UU^T = I \Rightarrow \det U \det U^T = (\det U)^2 = 1.$ □

Theorem 57. *The eigenvalues of an orthogonal matrix A are real or complex conjugates in pairs and have absolute value 1.*

Proof. $Au = \lambda u \Rightarrow A^T Au = \lambda A^T u \Rightarrow u = \lambda A^T u \Rightarrow \overline{u}^T u = \lambda \overline{u}^T A^T u \Rightarrow \overline{u}^T u = \lambda \overline{u}^T \overline{A}^T u = \lambda\overline{\lambda}\overline{u}^T u \Rightarrow \left(|\lambda|^2 - 1\right) \overline{u}^T u = 0.$ □

Properties of the special matrices From the above results, we have the following table:

real matrix	complex matrix	properties
symmetric ($A = A^T$)	Hermitian ($A^* = A$)	eigenvalues are all real
orthogonal ($A^T A = AA^T = I$)	unitary ($A^* A = AA^* = I$)	eigenvalues have unity magnitude; Ax maintains the 2-norm of x
skew-symmetric ($A^T = -A$)	skew-Hermitian ($A^* = -A$)	eigenvalues are all imaginary or zero

Based on the eigenvalue characteristics, we have:

- symmetric and Hermitian matrices are like the real line in the complex domain
- skew-symmetric/Hermitian matrices are like the imaginary line
- orthogonal/unitary matrices are like the unit circle

Exercise 58 (Representation of matrices using special matrices)**.** Many unitary matrices can be mapped as functions of skew-Hermitian matrices as follows

$$U = (I - S)^{-1} (I + S),$$

where $S \neq I$. Show that if S is skew-Hermitian, then U is unitary.

13.2 Symmetric eigenvalue decomposition (SED)

When $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, we have seen the useful result of matrix diagonalization:

$$A = U\Lambda U^{-1} = [u_1, \dots, u_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} [u_1, \dots, u_n]^{-1}, \tag{38}$$

where λ_i's are the distinct eigenvalues associated to the eigenvector u_i's.

The inverse matrix in (38) can be cumbersome to compute though.

The spectral theorem, aka symmetric eigenvalue decomposition theorem,[4] significantly simplifies the result when A is symmetric.

[4] Recall that the set of all the eigenvalues of A is called the spectrum of A. The largest of the absolute values of the eigenvalues of A is called the spectral radius of A.

Theorem 59. *$\forall : A \in \mathbb{R}^{n\times n},\ A^T = A$, there always exist λ_i and u_i, such that*

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U\Lambda U^T, \tag{39}$$

where:[5]

- *λ_i's: eigenvalues of A*
- *u_i: eigenvector associated to λ_i, normalized to have unity norms*
- *$U = [u_1, u_2, \cdots, u_n]^T$ is an orthogonal matrix, i.e., $U^TU = UU^T = I$*
- *$\{u_1, u_2, \cdots, u_n\}$ forms an orthonormal basis*
- *$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.*

To understand the result, we show an important theorem first.

Theorem 60. *$\forall : A \in \mathbb{R}^{n\times n}$* with *$A^T = A$, then eigenvectors of A, associated with different eigenvalues, are orthogonal.*

Proof. Let $Au_i = \lambda_i u_i$ and $Au_j = \lambda_j u_j$. Then $u_i^T A u_j = u_i^T \lambda_j u_j = \lambda_j u_i^T u_j$. In the meantime, $u_i^T A u_j = u_i^T A^T u_j = (Au_i)^T u_j = \lambda_i u_i^T u_j$. So $\lambda_i u_i^T u_j = \lambda_j u_i^T u_j$. But $\lambda_i \neq \lambda_j$. It must be that $u_i^T u_j = 0$. □

Theorem 59 now follows. If A has distinct eigenvalues, then $U = [u_1, u_2, \cdots, u_n]^T$ is orthogonal if we normalize all the eigenvectors to unity norm. If A has $r(< n)$ distinct eigenvalues, we can *choose* multiple orthogonal eigenvectors for the eigenvalues with none-unity multiplicities.

Observations:

- If we "walk along" u_j, then

$$Au_j = \left(\sum_i \lambda_i u_i u_i^T\right) u_j = \lambda_j u_j u_j^T u_j = \lambda_j u_j, \tag{40}$$

  where we used the orthonormal condition of $u_i^T u_j = 0$ if $i \neq j$. This confirms that u_j is an eigenvector.

[5] $u_i u_i^T \in \mathbb{R}^{n\times n}$ is a symmetric dyad, the so-called outerproduct of u_i and u_i. It has the following properties:

- $\forall\ v \in \mathbb{R}^{n\times 1}$, $\left(vv^T\right)_{ij} = v_i v_j$. (Proof: $\left(vv^T\right)_{ij} = e_i^T\left(vv^T\right)e_j = v_i v_j$, where e_i is the unit vector with all but the i_{th} elements being zero.)
- link with quadratic functions: $q(x) = x^T\left(vv^T\right)x = \left(v^Tx\right)^2$

- $\{u_i\}_{i=1}^n$ is a orthonormal basis $\Rightarrow \forall x \in \mathbb{R}^n$, $\exists\ x = \sum_i \alpha_i u_i$, where $\alpha_i =< x, u_i >$. And we have

$$Ax = A\sum_i \alpha_i u_i = \sum_i \alpha_i A u_i = \sum_i \alpha_i \lambda_i u_i = \sum_i (\alpha_i \lambda_i)\, u_i, \tag{41}$$

  which gives the (intuitive) picture of the geometric meaning of Ax: decompose first x to the space spanned by the eigenvectors of A, scale each components by the corresponding eigenvalue, sum the results up, then we will get the vector Ax.

With the spectral theorem, next time we see a symmetric matrix A, we immediately know that

- λ_i is real for all i
- associated with λ_i, we can always find one or more real eigenvectors
- $\exists$ an orthonormal basis $\{u_i\}_{i=1}^n$, which consists of the eigenvectors
- if $A \in \mathbb{R}^{2\times 2}$, then if you compute first λ_1, λ_2 and u_1, you won't need to go through the regular math to get u_2, but can simply solve for a u_2 that is orthogonal to u_1 with $\|u_2\| = 1$.

Example 61. Consider the matrix $A = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}$. Computing the eigenvalues gives

$$\det \begin{bmatrix} 5-\lambda & \sqrt{3} \\ \sqrt{3} & 7-\lambda \end{bmatrix} = 35 - 12\lambda + \lambda^2 - 3 = (\lambda - 4)(\lambda - 8) = 0$$
$$\Rightarrow \lambda_1 = 4,\ \lambda_2 = 8.$$

We can know one of the eigenvectors from

$$(A - \lambda_1 I)\, t_1 = 0 \Rightarrow \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Note here we normalized t_1 such that $||t_1||_2 = 1$. With the above computation, we no more need to do $(A - \lambda_2 I)\, t_2 = 0$ for getting t_2. Keep in mind that A here is symmetric, so has eigenvectors orthogonal to each other. By direct observation, we can see that

$$x = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

is orthogonal to t_1 and $||x||_2 = 1$. So $t_2 = x$.

Theorem 62 (Eigenvalues of symmetric matrices). *If $A = A^T \in \mathbb{R}^{n\times n}$, then the eigenvalues of A satisfy*

$$\lambda_{\max} = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{42}$$

$$\lambda_{\min} = \min_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^T A x}{\|x\|_2^2}. \tag{43}$$

Proof. Perform SED to get

$$A = \sum_{i=1}^{n} \lambda_i u_i^T u_i,$$

where $\{u_i\}_{i=1}^n$ form a basis of $\mathbb{R}^n$. Then any vector $x \in \mathbb{R}^n$ can be decomposed as

$$x = \sum_{i=1}^{n} \alpha_i u_i.$$

Thus

$$\max_{x\neq 0} \frac{x^T A x}{\|x\|_2^2} = \max_{\alpha_i} \frac{\left(\sum_i \alpha_i u_i\right)^T \sum_i \lambda_i \alpha_i u_i}{\sum_i \alpha_i^2} = \max_{\alpha_i} \frac{\sum_i \lambda_i \alpha_i^2}{\sum_i \alpha_i^2} = \lambda_{\max}.$$

The proof for (43) is analogous and omitted. □

13.3 Symmetric positive-definite matrices

Definition 63. A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **positive-definite**, written $P \succ 0$, if $x^T P x > 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$. P is called **positive-semidefinite,** written $P \succeq 0$, if $x^T P x \geq 0$ for all $x \in \mathbb{R}^n$

Definition 64. A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **negative-definite**, written $P \prec 0$, if $-P \succ 0$, i.e., $x^T P x < 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$. P is called **negative-semidefinite,** written $P \preceq 0$, if $x^T P x \leq 0$ for all $x \in \mathbb{R}^n$

When A and B have compatible dimensions, $A \succ B$ means $A - B \succ 0$.
Positive-definite matrices can have negative entries, as shown in the next example.

Example 65. The following matrix is positive-definite

$$P = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

as $P = P^T$ and take any $v = [x, y]^T$, we have

$$v^T P v = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 2y^2 - 2xy = x^2 + y^2 + (x+y)^2 \geq 0,$$

and the equality sign holds only when $x = y = 0$.

Conversely, matrices whose entries are all positive are not necessarily positive-definite.

Example 66. The following matrix is not positive-definite

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

as

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 < 0.$$

Theorem 67. *For a symmetric matrix P, $P \succ 0$ if and only if all the eigenvalues of P are positive.*

Proof. Since P is symmetric, we have

$$\lambda_{\max}(P) = \max_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{44}$$

$$\lambda_{\min}(P) = \min_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2}, \tag{45}$$

which gives

$$x^T A x \in \left[\lambda_{\min}\|x\|_2^2,\ \lambda_{\max}\|x\|_2^2\right].$$

For $x \neq 0$, $\|x\|_2^2$ is always positive. It can thus be seen that $x^T A x > 0,\ x \neq 0 \Leftrightarrow \lambda_{\min} > 0$. □

Lemma. *For a symmetric matrix P, $P \succeq 0$ if and only if all the eigenvalues of P are none-negative.*

Theorem. *If A is symmetric positive definite, X is full column rank, then $X^T A X$ is positive definite.*

Proof. Consider $y\left(X^T A X\right) y = x^T A x$, which is always positive unless $x = 0$. But X is full rank so $Xy = x = 0$ if and only if $y = 0$. This proves $X^T A X$ is positive definite. □

Fact. *All principle submatrices of A are positive definite.*

Proof. Use the last theorem. Take $X = e_1$, $X = [e_1, e_2]$, etc. Here e_i is a column vector whose ith-entry is 1 and all other entries are zero. □

Example 68. The following matrices are all not positive-definite:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Positive-definite matrices are like positive real numbers. We can have the concept of *square root* of positive-definite matrices.

Definition 69. Let $P \succeq 0$. We can perform symmetric eigenvalue decomposition to obtain $P = UDU^T$ where U is orthogonal with $UU^T = I$ and D is diagonal with all diagonal elements being none negative

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} \succeq 0$$

. Then the square root of P, written $P^{\frac{1}{2}}$, is defined as

$$P^{\frac{1}{2}} = UD^{\frac{1}{2}}U^T$$

,where

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_n} \end{bmatrix}$$

.

13.4 General positive-definite matrices

Definition 70. A general square matrix $Q \in \mathbb{R}^{n\times n}$ is called positive-definite, written as $Q \succ 0$, if $x^TQx > 0 \ \forall x \neq 0$.

We have discussed the case when Q is symmetric. If not, recall that any real square matrix can be decomposed as the sum of a symmetric matrix and a skew symmetric matrix:

$$Q = \frac{Q+Q^T}{2} + \frac{Q-Q^T}{2},$$

where $\frac{Q+Q^T}{2}$ is symmetric.

Notice that $x^T\frac{Q-Q^T}{2}x = x^TQx - \left(x^TQx\right)^T = 0$. So for a general square real matrix:

$$Q \succ 0 \Leftrightarrow Q + Q^T \succ 0.$$

Example 71. The following matrices are positive definite but not symmetric

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

For complex matrices with $Q = Q^* = Q_R + jQ_I$, we have

$$\begin{aligned} Q \succ 0 &\Leftrightarrow x^*Qx > 0, \ \forall x \neq 0 \\ &\Leftrightarrow \left(x_R^T - jx_I^T\right)\left(Q_R + jQ_I\right)\left(x_R + jx_I\right) > 0 \\ &\Leftrightarrow \begin{pmatrix} x_R \\ x_I \end{pmatrix}^T \begin{pmatrix} 1 \\ j \end{pmatrix} \begin{pmatrix} Q_R & Q_I \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix}^T \begin{pmatrix} x_R \\ x_I \end{pmatrix} \\ &\Leftrightarrow \begin{pmatrix} x_R \\ x_I \end{pmatrix}^T \begin{pmatrix} Q_R & Q_I \\ -Q_I & Q_R \end{pmatrix} \begin{pmatrix} x_R \\ x_I \end{pmatrix} > 0 \\ &\Leftrightarrow x_R^TQ_Rx_R - x_I^TQ_Ix_R + x_R^TQ_Ix_I + x_I^TQ_Rx_I > 0. \end{aligned}$$

13.5 *Positive-definite functions and non-constant matrices

We can further extend the concept of positive definiteness to general and even time-varying functions, by placing upper and/or lower bounds that are "positive-definite like".

Define first two special functions:

1. class-K function: $\psi \in C^0 : [0, a] \to [0, \infty)$ with $\psi(0) = 0$ and ψ strictly increasing,
2. class-K_∞ function: if the domain $a = \infty$ and $\psi(r) \to \infty$ as $r \to \infty$.

Note: ψ is continuous but does not need to be continuously differentiable, e.g.

$$\psi = \min\left\{x, x^2\right\}$$

is a class-K function.

Lemma 72. *Let $V : D \to \mathbb{R}$ be a continuous, positive definite function. Let $B_r \subset D$ for some $r > 0$. Then there exist class-K functions ψ and ϕ defined on $[0, r]$ such that*

$$\phi(\|x\|) \leq V(x) \leq \psi(\|x\|)$$

for all $x \in B_r$.

- if the domain $D = \mathbb{R}^n$ then $r = \infty$,
- if $V(x)$ is radially unbounded, then ψ and ϕ can be class-K_∞.

Definition 73. A time-dependent function $V(t, x)$ is positive-semidefinite if

$$V(t, x) \geq \phi(\|x\|),$$

where ϕ is class-K.

Definition 74. A time-varying matrix $P(t)$ is positive definite if there exists a lower-bounding positive definite matrix such that

$$P(t) \succeq c_3 I \succ 0, \ \forall t \geq 0.$$

14 Singular value and singular value decomposition (SVD)

14.1 Motivation

Symmetric eigenvalue decomposition is great but many matrices are not symmetric. A general matrix A may actually not even be square. Singular value decomposition is an important matrix decomposition technique that works for arbitrary matrices.[6]

For a general none-square matrix $A \in \mathbb{C}^{m\times n}$, eigenvalues and eigenvectors are generalized to

$$Av_j = \sigma_j u_j \tag{46}$$

Be careful about the dimensions: if $m > n$, we have

$$\begin{bmatrix} & \vdots & \\ \cdots & A & \cdots \\ & \vdots & \end{bmatrix} \underbrace{\left[\begin{array}{c|c|c|c} v_1 & v_2 & \dots & v_n \end{array}\right]}_{V} = \underbrace{\left[\begin{array}{c|c|c|c} u_1 & u_2 & \dots & u_n \end{array}\right]}_{\hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}}_{\hat{\Sigma}}.$$

It turns out that, if A has full column rank n, then we can find a V that is unitary ($VV^* = V^*V = I$) and a $\hat{U}$ that has orthonormal columns. Hence

$$A = \hat{U}\hat{\Sigma}V^*. \tag{47}$$

14.2 SVD

(47) forms the so-called reduced singular value decomposition (SVD). The idea of a "full" SVD is as follows. The columns of $\hat{U}$ are n orthonormal vectors in the m-dimensional space $\mathbb{C}^m$. They do not form a basis for $\mathbb{C}^m$ unless $m = n$. We can add additional $m - n$ orthonormal columns to $\hat{U}$ and augment it to a unitary matrix U. Now the matrix dimension has changed, $\hat{\Sigma}$ needs to be augmented to compatible dimensions as well. To maintain the equality (47), the newly added elements to $\hat{\Sigma}$ are set to zero.

Theorem 75. *Let $A \in \mathbb{C}^{m\times n}$ with rank r. Then we can find orthogonal matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ such that*

$$A = U\Sigma V^*,$$

[6] History of SVD: discovered between 1873 and 1889, independently by several pioneers; did not became widely known in applied mathematics until the late 1960s, when it was shown that SVD can be computed effectively and used as the basis for solving many problems.

where

$$\begin{aligned}\Sigma &\in \mathbb{R}^{m\times n} \text{ is diagonal}\\ U &\in \mathbb{C}^{m\times m} \text{ is unitary}\\ V &\in \mathbb{C}^{n\times n} \text{ is unitary.}\end{aligned}$$

In addition, the diagonal entries σ_j of Σ are nonnegative and in nonincreasing order; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Proof. Notice that A^*A is positive semi-definite. Hence, A^*A has a full set of orthonormal eigenvectors; its eigenvalues are real and nonnegative. Order these eigenvalues as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0.$$

[7]Let $\{v_1, \dots, v_n\}$ be an orthonormal choice of eigenvectors of A^*A corresponding to these eigenvalues:

$$A^*Av_i = \lambda_i v_i.$$

Then,

$$||Av_i||^2 = v_i^*A^*Av_i = \lambda_i v_i^* v_i = \lambda_i.$$

For $i > r$, it follows that $Av_i = 0$.

For $1 \leq i \leq r$, we have

$$A^*Av_i = \lambda_i v_i.$$

Recall (46), we define $\sigma_i = \sqrt{\lambda_i}$ and get

$$\begin{aligned}Av_i &= \sigma_i u_i\\ A^*u_i &= \sigma_i v_i.\end{aligned}$$

For $1 \leq i, j \leq r$, we have

$$\langle u_i, u_j \rangle = u_i^* u_j = \frac{1}{\sigma_i\sigma_j} v_i^* A^* A v_j = \frac{1}{\sigma_i\sigma_j}\lambda_j v_i^* v_j = \frac{\sigma_j}{\sigma_i} v_i^* v_j = \begin{cases}1 & i = j,\\ 0 & i \neq j.\end{cases}$$

Hence $\{u_{1,} \dots, u_r\}$ is an orthonormal set of eigenvectors. Extending this set to form an orthonormal basis for $\mathbb{C}^m$ gives

$$U = \left[\begin{array}{ccc|ccc} u_1, & \dots, & u_r & u_{r+1}, & \dots, & u_m\end{array}\right].$$

For $i \leq r$, we already have

$$Av_i = \sigma_i u_i,$$

[7] Fact: rank (A) = rank (A^*A). To see this, notice first, that rank $(A) \geq$ rank (A^*A) by definition of rank. Second, $A^*Ax = 0 \Rightarrow x^*A^*Ax = 0 \Rightarrow ||Ax|| = 0 \Rightarrow Ax = 0$, hence rank $(A) \leq$ rank (A^*A).

namely

$$A\left[v_1,\dots v_r\right]=\left[u_1,\dots,u_r\right]\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}$$

$$=\left[\begin{array}{ccc|ccc} u_1, & \dots, & u_r & u_{r+1}, & \dots, & u_m\end{array}\right]\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \\ & & & 0 \\ & & & \vdots \\ & & & 0\end{bmatrix}.$$

For $v_{r+1},\dots$, we have already seen that $Av_{r+1}=Av_{r+2}=\dots=0$, hence

$$A\underbrace{\left[v_1,\dots v_r|v_{r+1},\dots,v_n\right]}_{n\times n}=\underbrace{\left[\begin{array}{ccc|ccc} u_1, & \dots, & u_r & u_{r+1}, & \dots, & u_m\end{array}\right]}_{m\times m}\underbrace{\begin{bmatrix}\sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & & & \vdots \\ & & & & & 0\end{bmatrix}}_{m\times n}$$

$$\Rightarrow A=U\Sigma V^*.$$

□

Theorem 76. *The range space of A is spanned by $\{u_1,\dots,u_r\}$. The null space of A is spanned by $\{v_{r+1},\dots,v_n\}$.*

□

Theorem 77. *The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A or AA^*.*

□

Theorem 78. *$||A||_2=\sigma_1$, i.e., the induced two norm of A is the maximum singular value of A.*

The next important theorem can be easily proved via SVD.

Theorem (Fundermental theory of linear algebra). *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$\mathcal{R}\left(A\right)+\mathcal{N}\left(A^T\right)=\mathbb{R}^m,$$

and

$$\mathcal{R}\left(A\right)\perp\mathcal{N}\left(A^T\right).$$

Proof. By singular value decomposition, we have

$$\begin{aligned}A&=U\Sigma V^T\\ A^T&=V\Sigma U^T.\end{aligned}$$

The range space of A is the first r columns of U, from the first equation. The null space of A^T is the last $m-r$ columns of U, from the second equation. □

New intuition of matrix vector operation With $A=U\Sigma V^*$, a new intuition for $Ax=U\Sigma V^*x$ is formed. Since V is unitary, it is norm-preserving, in the sense that V^*x does not change the 2-norm of the vector x. In other words, V^*x only rotates x in $\mathbb{C}^n$. The diagonal matrix Σ then functions to scale (by its diagonal values) the rotated vector. Finally, U is another rotation in $\mathbb{C}^m$.

14.3 Properties of singular values

Fact. *Let A and B be matrices with compatible dimensions. The following are true*
$\overline{\sigma}\left(A+B\right)\leq\overline{\sigma}\left(A\right)+\overline{\sigma}\left(B\right),$
$\overline{\sigma}\left(AB\right)\leq\overline{\sigma}\left(A\right)\overline{\sigma}\left(B\right).$

Proof. The first inequality comes from

$$\overline{\sigma}\left(A+B\right)=\max_{v\neq0}\frac{||Av+Bv||_2}{||v||_2}\leq\max_{v\neq0}\frac{||Av||_2+||Bv||_2}{||v||_2}.$$

The second inequality uses

$$\overline{\sigma}\left(AB\right)=\max_{v\neq0}\frac{||ABv||_2}{||v||_2}\leq\max_{v\neq0}\frac{||A||_2||Bv||_2}{||v||_2}.$$

□

14.4 Exercises

1. Compute the singular values of the following matrices

$$(a)\begin{bmatrix}3 & \\ & -2\end{bmatrix},\ (b)\begin{bmatrix}2 & \\ & 3\end{bmatrix},\ (c)\begin{bmatrix}0 & 2\\ 0 & 0\\ 0 & 0\end{bmatrix},\ (d)\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix},\ (e)\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}.$$

2. Show that if A is real, then it has a real SVD (i.e., U and V are both real).

3. For any matrix $A \in \mathbb{R}^{n\times m}$, construct

$$M = \begin{bmatrix} \overbrace{0}^{n\times n} & \overbrace{A}^{n\times m} \\ \underbrace{A^T}_{m\times n} & \underbrace{0}_{m\times m} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)},$$

which satisfies

$$M^T = M.$$

M is Hermitian, and hence has real eigenvalues and eigenvectors:

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_j \\ v_j \end{bmatrix} = \sigma_j \begin{bmatrix} u_j \\ v_j \end{bmatrix}. \tag{48}$$

(a) Show that

i. v_j is in the co-kernal (perpendicular to kernal/null space) of A and u_j is in the range of A.

ii. if σ_j and $\begin{bmatrix} u_j \\ v_j \end{bmatrix}$ form a eigen pair for M, then$-\sigma_j$ and $\left[u_j^T, -v_j^T\right]^T$ also form an eigen pair for M

iii. eigenvalues of M always appear in pairs that are symmetric to the imaginary axis.

(b) Use the results to show that, if

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 4 & 32 \end{bmatrix},$$

then M must have eigenvalues that are equal to 0.

4. Suppose $A \in \mathbb{C}^{m\times m}$ and has an SVD $A = U\Sigma V^*$. Find an eigenvalue decomposition of

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}.$$

5. **Worst input direction** in matrix vector multiplications. Recall that any matrix defines a linear transformation:

$$Mw = z$$

What is the worst input direction for the vector w? Here *worst* means: if we fix the input norm, say $\|w\| = 1$, $\|z\|$ will reach a maximum value (the worst case) for a specific input direction in w.

(a) Show that the worst $||z||$ is $||M||$ when $||w|| = 1$.

(b) Provide procedures to obtain the w that gives the maximum $||z||$, for the cases of 1 norm, ∞ norm, and 2 norm.

Index