# Lecture 1, Sep 3, 2025

## Theory of Linear Control Systems

- *Dynamical systems* (aka *systems* or *dynamical processes*) are mathematical models that describe how quantities of interest evolve over time
  - At a high level: input $u(t)$, output $y(t)$
- We are interested in the theory of *analysis* (how is the system behaving?) and *control* (how can we make the system behave well/better?) for *linear time-invariant* (LTI) systems in continuous time

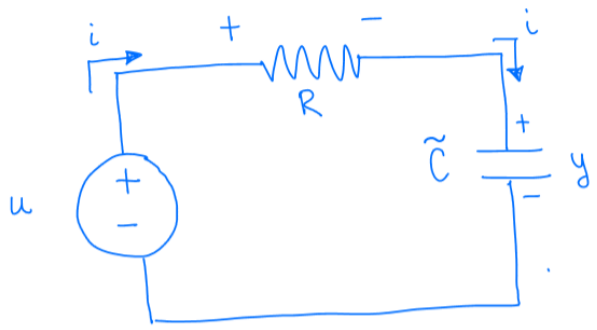


Figure 1: Circuit example of a dynamical system.

- Consider the circuit above; let $y$ be the voltage across the capacitor and $u$ be the input voltage
  - $i = \tilde{C}\dfrac{\mathrm{d}y}{\mathrm{d}t}$
  - By KVL: $u - iR - y = 0 \implies u - R\tilde{C}\dfrac{\mathrm{d}y}{\mathrm{d}t} - y = 0$
  - Rearrange: $\dfrac{\mathrm{d}y}{\mathrm{d}t} = -\dfrac{1}{R\tilde{C}}y + \dfrac{1}{R\tilde{C}}u$
  - We now have an ODE representing the LTI system

**Definition**

A *linear time-invariant* system in *state-space* form is represented by the following:

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t), t \geq 0$$
$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{x}(t) + \boldsymbol{D}\boldsymbol{u}(t)$$

where

$$\boldsymbol{x} : [0, \infty) \in \mathbb{R}^n (\text{state})$$
$$\boldsymbol{u} : [0, \infty) \in \mathbb{R}^m (\text{input})$$
$$\boldsymbol{y} : [0, \infty) \in \mathbb{R}^p (\text{output})$$

The first is known as the *state equation* while the second is the *measurement equation*.

- $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the *system* matrix.
- $\boldsymbol{B} \in \mathbb{R}^{n \times m}$ is the *input* matrix.
- $\boldsymbol{C} \in \mathbb{R}^{p \times n}$ is the *output* matrix.
- $\boldsymbol{D} \in \mathbb{R}^{p \times m}$ is the *feedforward* matrix.

- For our example system: $x = y, \dot{x} = -\dfrac{1}{R\tilde{C}}x + \dfrac{1}{R\tilde{C}}u, y = x + u$ (note in this example we can fully measure the state)
  - $A = -\dfrac{1}{R\tilde{C}}$
  - $B = \dfrac{1}{R\tilde{C}}$

  - $C = 1$
  - $D = 0$
- Within dynamical systems we have several forms of representation:
  - Model-based representation (for systems represented as ODEs)
    * We are interested in continuous time, deterministic (no noise), LTI systems
  - Data-based representation (for streams of data)
  - Computer-based representation (for complicated systems that we cannot write down)

# Lecture 2, Sep 5, 2025

## Converting Between State Space and Transfer Functions

- Recall the transfer function representation: $Y(s) = G(s)U(s)$ where $U(s) = \mathcal{L}\{u(t)\}$ (input), $Y(s) = \mathcal{L}\{y(t)\}$ (output), the transfer function is $G(s) = \mathcal{L}\{g(t)\}$ (impulse response)
  - Also known as the *input-output representation*
  - Note this assumes zero initial conditions
- Using the circuit example: $\dfrac{\mathrm{d}y}{\mathrm{d}t} + \dfrac{1}{R\tilde{C}}y = \dfrac{1}{R\tilde{C}}u$
  - Assuming zero initial conditions, $\mathcal{L}\{y\}$ and $\mathcal{L}\{u\}$ exist in the right-half complex plane
  - Apply Laplace: $sY(s) + \dfrac{1}{R\tilde{C}}Y(s) = \dfrac{1}{R\tilde{C}}U(s) \implies \left(s + \dfrac{1}{R\tilde{C}}\right)Y(s) = \dfrac{1}{R\tilde{C}}U(s)$
  - Therefore $G(s) = \dfrac{\frac{1}{R\tilde{C}}}{s + \frac{1}{R\tilde{C}}}$
  - To go from state space to transfer function representation, we can take the Laplace transform and rearrange into the $Y(s) = G(s)U(s)$ form
- To transform transfer function to state space: Let $G(s) = \dfrac{b_m s^m + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} = \dfrac{N(s)}{D(s)}$ and assume $a_i, b_i \in \mathbb{R}$ (*rational*) and $m < n$ (*strictly proper*)
  - Break into 2 blocks, $\dfrac{1}{D(s)}$ and then $N(s)$, and let the intermediate output be $V(s)$; the first block will give us our state equation, the second will give the measurement equation
  - Block 1: $\dfrac{V(s)}{U(s)} = \dfrac{1}{D(s)} = \dfrac{1}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}$
    * $\left(s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0\right)V(s) = U(s)$
    * Inverse Laplace assuming zero initial conditions: $\dfrac{\mathrm{d}^n v}{\mathrm{d}t^n} + a_{n-1}\dfrac{\mathrm{d}^{n-1}v}{\mathrm{d}t^{n-1}} + \cdots + a_1\dfrac{\mathrm{d}v}{\mathrm{d}t} + a_0 v = u$
    * Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} v \\ \dfrac{\mathrm{d}v}{\mathrm{d}t} \\ \vdots \\ \dfrac{\mathrm{d}^{n-1}v}{\mathrm{d}t^{n-1}} \end{bmatrix} \implies \dot{\boldsymbol{x}} = \begin{bmatrix} \dfrac{\mathrm{d}v}{\mathrm{d}t} \\ \dfrac{\mathrm{d}^2 v}{\mathrm{d}t^2} \\ \vdots \\ \dfrac{\mathrm{d}^n v}{\mathrm{d}t^n} \end{bmatrix} = \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ -a_{n-1}x_n - \cdots - a_1 x_2 - a_0 x_1 + u \end{bmatrix}$
    * $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$
  - Block 2: $Y(s) = V(s)N(s) = \left(b_m s^m + \cdots + b_1 s + b_0\right)V(s)$
    * Again inverse Laplace assuming zero IC ($v(0) = \dot{v}(0) = \cdots = \dfrac{\mathrm{d}^{m-1}v}{\mathrm{d}t^{m-1}}(0) = 0$)
    * Using the definition of $\boldsymbol{x}$: $y(t) = b_m\dfrac{\mathrm{d}^m v}{\mathrm{d}t^m} + \cdots + b_1\dfrac{\mathrm{d}v}{\mathrm{d}t} + b_0 v = b_m x_{m+1} + \cdots + b_1 x_2 + b_0 x_1$
      - Here is where we use the $m < n$ assumption

    * Therefore: $\boldsymbol{C} = \begin{bmatrix} b_0 & b_1 & \cdots & b_m \end{bmatrix}, \boldsymbol{D} = 0$
  - Note there are many other sets of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ that satisfy this

**Note**

Given the state-space representation with $\boldsymbol{x}(0) = \boldsymbol{0}$, we can show that the corresponding transfer function is

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

Note $\boldsymbol{G} \in \mathbb{R}^{p \times m}$ is a matrix. This can be derived by taking the Laplace transform, then isolating and substituting $\boldsymbol{X}$.

# Lecture 3, Sep 5, 2025

## Linearization

- Consider a general nonlinear function $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}), \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u})$ where $\boldsymbol{f}, \boldsymbol{h}$ are differentiable; we want to approximate this system by an LTI model by linearization around an equilibrium point

**Definition**

A pair $(\boldsymbol{x}^*, \boldsymbol{u}^*)$ is an *equilibrium condition* if $\boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*) = \boldsymbol{0}$. In this case $\boldsymbol{x}^*$ is an *equilibrium point* with control $\boldsymbol{u}^*$.

- Consider the example of an actuated pendulum affected by gravity $\boldsymbol{x} = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix}, y = x_1$
  - $\dot{x}_1 = x_2$
  - $\dot{x}_2 = -\dfrac{mgl}{J}\sin x_1 + \dfrac{u}{J}$ (torque balance) where $J$ is the moment of inertia
  - $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) = \begin{bmatrix} x_2 \\ -\dfrac{mgl}{J}\sin x_1 + \dfrac{u}{J} \end{bmatrix}$
  - With a control of $u^* = 0$, $\boldsymbol{f}(\boldsymbol{x}^*, 0) = \dfrac{x_2^*}{-\frac{mgl}{J}\sin x_1^*} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \boldsymbol{x}^* = \begin{bmatrix} k\pi \\ 0 \end{bmatrix}, k \in \mathbb{Z}$
    * Physically this corresponds to the pendulum being perfectly up or down with zero velocity
  - With a control of $u^* = mgl$, $\dfrac{x_2}{-\frac{mgl}{J}\sin x_1^* + \frac{mgl}{J}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \boldsymbol{x}^* = \begin{bmatrix} \frac{\pi}{2} + 2k\pi \\ 0 \end{bmatrix}$
    * This corresponds to the applied torque being balanced by gravity
- Consider $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^*) + \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) + \boldsymbol{R}(\boldsymbol{x})$ where $\boldsymbol{R}$ is a remainder term
  - $\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_{n_1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{n_2}}{\partial x_1} & \cdots & \frac{\partial f_{n_2}}{\partial n_1} \end{bmatrix} \in \mathbb{R}^{n_2 \times n_1}$ is the *Jacobian* of $\boldsymbol{f}$
  - For differentiable $\boldsymbol{f}$, $\lim\limits_{\boldsymbol{x} \to \boldsymbol{x}^*} \dfrac{\boldsymbol{R}(\boldsymbol{x})}{\|\boldsymbol{x} - \boldsymbol{x}^*\|} = \boldsymbol{0}$
- Let $\boldsymbol{z}^* = (\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{u})$, then $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{z}) = \boldsymbol{f}(\boldsymbol{z}^*) + \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}}(\boldsymbol{z}^*)(\boldsymbol{z} - \boldsymbol{z}^*) + \boldsymbol{R}(\boldsymbol{z})$ where $\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}} = \begin{bmatrix} \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} & \dfrac{\partial \boldsymbol{x}}{\partial \boldsymbol{u}} \end{bmatrix}$
  - Therefore $\dot{\boldsymbol{x}} \approx \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{z}}(\boldsymbol{z}^*)(\boldsymbol{z} - \boldsymbol{z}^*) = \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*)(\boldsymbol{x} - \boldsymbol{x}^*) + \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*)(\boldsymbol{u} - \boldsymbol{u}^*)$
    * Note we used the fact that $\boldsymbol{z}^*$ is an equilibrium condition
  - Let $\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^*, \delta\boldsymbol{u} = \boldsymbol{u} - \boldsymbol{u}^*$ then $\dot{\delta\boldsymbol{x}} = \dot{\boldsymbol{x}} \approx \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*)\delta\boldsymbol{x} + \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*)\delta\boldsymbol{u}$

– Similarly let $\delta \boldsymbol{y} = \boldsymbol{y} - \boldsymbol{h}(\boldsymbol{x}^*, \boldsymbol{u}^*)$ then $\delta \boldsymbol{y} \approx \dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*)\delta \boldsymbol{x} + \dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*)\delta \boldsymbol{u}$

– Therefore: $\boldsymbol{A} = \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{B} = \dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{C} = \dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{D} = \dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*)$

**Summary**

To linearize a general nonlinear system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}), \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u})$ where $\boldsymbol{f}, \boldsymbol{h}$ are differentiable, let $(\boldsymbol{x}^*, \boldsymbol{u}^*)$ be an equilibrium condition, then a linear approximation is

$$\delta \dot{\boldsymbol{x}} = \boldsymbol{A}\delta \boldsymbol{x} + \boldsymbol{B}\delta \boldsymbol{u}$$
$$\boldsymbol{y} = \boldsymbol{C}\delta \boldsymbol{x} + \boldsymbol{D}\delta \boldsymbol{u}$$

where $\delta \boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^*, \delta \boldsymbol{u} = \boldsymbol{u} - \boldsymbol{u}^*$ and

$$\boldsymbol{A} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{B} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{C} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \boldsymbol{u}^*), \boldsymbol{D} = \frac{\partial \boldsymbol{h}}{\partial \boldsymbol{u}}(\boldsymbol{x}^*, \boldsymbol{u}^*)$$

# Lecture 4, Sep 10, 2025

## The Matrix Exponential

**Definition**

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, then the *matrix exponential* is defined as:

$$e^{\boldsymbol{A}} = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{A}^k$$

Note $\boldsymbol{A}^0 = \boldsymbol{I}_n$.

- To define the matrix exponential based on a series, we have to first define convergence for matrices, and then show that this series definition of the matrix exponential converges

**Definition**

A series of matrices $\sum_{k=0}^{\infty} \boldsymbol{M}_k$ *converges* if every element $(\boldsymbol{S}_n)_{ij}$ of the partial sum $\boldsymbol{S}_n = \sum_{k=0}^{n} \boldsymbol{M}_k$ converges to a number as $n \to \infty$, i.e. $\lim_{n \to \infty} (\boldsymbol{S}_n)_{ij} = a_{ij}$ for all $i, j$.

Formally, we require

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } n > N \implies |(\boldsymbol{S}_n)_{ij} - a_{ij}| < \epsilon$$

**Definition**

A *norm* on $\mathrm{R}^n$ is a function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ with the following properties:

1. $\|\boldsymbol{x}\| \geq 0 \ \forall x \in \mathbb{R}^n$
2. $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \boldsymbol{0} \in \mathbb{R}^n$
3. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\| \ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$
4. $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\| \ \forall \boldsymbol{x} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}$

**Definition**

The *induced norm* on $\mathbb{R}^{n\times n}$ is a function $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}$ defined as

$$\|\boldsymbol{A}\| = \max_{\{\,\boldsymbol{x}\in\mathbb{R}^n \,|\, \|\boldsymbol{x}\|=1\,\}} \|\boldsymbol{A}\boldsymbol{x}\|$$

i.e. max norm of $\boldsymbol{A}\boldsymbol{x}$ over all $\boldsymbol{x}$ in the unit sphere. Note that this matrix norm is defined in terms of (*induced by*) the vector norm.

- We can show that the induced norm is a valid norm, and it has property $\|\boldsymbol{A}^k\| \leq \|\boldsymbol{A}\|^k$

**Theorem**

If the scalar series $\sum_{k=0}^{\infty} \|\boldsymbol{M}_k\|$ converges, then the matrix series $\sum_{k=0}^{\infty} \boldsymbol{M}_k$ converges. Such a series is called *absolutely convergent*.

- We will now prove that $e^{\boldsymbol{A}} = \sum_{k=0}^{\infty} \frac{1}{k!}\boldsymbol{A}^k$ is absolutely convergent:
  - We need to show that $\lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{k=0}^{n} \left\| \frac{\boldsymbol{A}^k}{k!} \right\|$ converges
  - We will rely on the fact that an increasing sequence that is bounded above always converges
  - $S_n$ is an increasing sequence, since $S_{n+1} - S_n = \left\| \frac{\boldsymbol{A}^{n+1}}{(n+1)!} \right\| = \frac{1}{(n+1)!}\|\boldsymbol{A}^{n+1}\| \geq 0$
  - To bound $S_n$ from above, we will show that $S_n \leq e^{\|\boldsymbol{A}\|}$
    - $S_n = \sum_{k=0}^{n} \left\| \frac{\boldsymbol{A}^k}{k!} \right\| \leq \sum_{k=0}^{\infty} \frac{1}{k!}\|\boldsymbol{A}^k\| \leq \sum_{k=0}^{\infty} \frac{1}{k!}\|\boldsymbol{A}\|^k = e^{\|\boldsymbol{A}\|}$

**Theorem**

The matrix exponential satisfies the following properties:

1. For any invertible $\boldsymbol{P} \in \mathbb{R}^{n\times n}, e^{\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}} = \boldsymbol{P}e^{\boldsymbol{A}}\boldsymbol{P}^{-1}$
2. For any $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ such that $AB = BA$ (commutativity) $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}}e^{\boldsymbol{B}} = e^{\boldsymbol{B}}e^{\boldsymbol{A}}$
3. $(e^{\boldsymbol{A}})^{-1} = e^{-\boldsymbol{A}}$
4. For $t \in \mathbb{R}$, $\frac{\mathrm{d}}{\mathrm{d}t}e^{\boldsymbol{A}t} = \boldsymbol{A}e^{\boldsymbol{A}t} = e^{\boldsymbol{A}t}\boldsymbol{A}$

# Lecture 5, Sep 12, 2025

## Solving $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$

- Consider the *autonomous* (i.e. no control input) LTI system, $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbb{R}^n$; we will show that this is solved by $e^{\boldsymbol{A}t}\boldsymbol{x}_0$
- We will show that $\frac{\mathrm{d}}{\mathrm{d}t}e^{\boldsymbol{A}t} = e^{\boldsymbol{A}t}\boldsymbol{A}$

- $\dfrac{\mathrm{d}}{\mathrm{d}t}e^{\boldsymbol{A}t} = \lim_{h\to 0}\dfrac{e^{\boldsymbol{A}(t+h)} - e^{\boldsymbol{A}t}}{h}$
$$= \lim_{h\to 0}\frac{e^{\boldsymbol{A}t}e^{\boldsymbol{A}h} - e^{\boldsymbol{A}t}}{h} \qquad \text{Since } \boldsymbol{A}h \text{ and } \boldsymbol{A}t \text{ commute}$$
$$= \lim_{h\to 0}\frac{e^{\boldsymbol{A}t}(e^{\boldsymbol{A}h} - \boldsymbol{I})}{h}$$
$$= e^{\boldsymbol{A}t}\lim_{h\to 0}\frac{1}{h}\left(-\boldsymbol{I} + \sum_{k=0}^{\infty}\frac{(\boldsymbol{A}h)^k}{k!}\right) \qquad \text{Matrix exponential definition}$$
$$= e^{\boldsymbol{A}t}\lim_{h\to 0}\frac{1}{h}\left(-\boldsymbol{I} + \boldsymbol{I} + h\sum_{k=1}^{\infty}\frac{\boldsymbol{A}^k h^{k-1}}{k!}\right) \qquad \text{Take out first term and factor } h$$
$$= e^{\boldsymbol{A}t}\lim_{h\to 0}\sum_{k=1}^{\infty}\frac{\boldsymbol{A}^k \boldsymbol{h}^{k-1}}{k!}$$
$$= e^{\boldsymbol{A}t}\lim_{h\to 0}\left(\frac{\boldsymbol{A}h^0}{1!} + \sum_{k=2}^{\infty}\frac{\boldsymbol{A}^k h^{k-1}}{k!}\right) \qquad \text{Take out first term}$$
$$= e^{\boldsymbol{A}t}\boldsymbol{A}$$
  - Note due to commutativity, we could've also taken out $e^{\boldsymbol{A}t}$ on the right and get $\dfrac{\mathrm{d}}{\mathrm{d}t}e^{\boldsymbol{A}t} = \boldsymbol{A}e^{\boldsymbol{A}t}$

**Theorem**

The differential equation $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbb{R}^n$ has the **unique solution**

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0, t \geq 0$$

- To show existence:
  - $\dot{\boldsymbol{x}}(t) = \dfrac{\mathrm{d}}{\mathrm{d}t}(e^{\boldsymbol{A}t}\boldsymbol{x}_0) = \dfrac{\mathrm{d}}{\mathrm{d}t}(e^{\boldsymbol{A}t})\boldsymbol{x}_0 = \boldsymbol{A}e^{\boldsymbol{A}t}\boldsymbol{x}_0 = \boldsymbol{A}\boldsymbol{x}(t)$
  - $\boldsymbol{x}(0) = e^{\boldsymbol{0}}\boldsymbol{x}_0 = \boldsymbol{I}\boldsymbol{x}_0 = \boldsymbol{x}_0$
- To show uniqueness, let $\boldsymbol{y}(t)$ be any other solution to the differential equation; we want to show that $\boldsymbol{y} = \boldsymbol{x}$
  - $\dfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{y}(t) = \boldsymbol{A}\boldsymbol{y}(t)$ and $\boldsymbol{y}(0) = \boldsymbol{x}_0$
  - We want to show $\boldsymbol{y}(t) = e^{\boldsymbol{A}t}x_0$, equivalently $e^{-\boldsymbol{A}t}\boldsymbol{y}(t) = e^{-\boldsymbol{A}t}e^{\boldsymbol{A}t}\boldsymbol{x}_0 = \boldsymbol{x}_0$
  - Notice that both sides are constants, so we can try taking the derivative and seeing what we get
  - $\dfrac{\mathrm{d}}{\mathrm{d}t}e^{-\boldsymbol{A}t}\boldsymbol{y}(t) = \left(\dfrac{\mathrm{d}}{\mathrm{d}t}e^{-\boldsymbol{A}t}\right)\boldsymbol{y}(t) + e^{-\boldsymbol{A}t}\dfrac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{y}(t)$
$$= -e^{-\boldsymbol{A}t}\boldsymbol{A}\boldsymbol{y}(t) + e^{-\boldsymbol{A}t}\boldsymbol{A}\boldsymbol{y}(t)$$
$$= -e^{\boldsymbol{A}t}\left(-\boldsymbol{A}\boldsymbol{y}(t) + \boldsymbol{A}\boldsymbol{y}(t)\right)$$
$$= \boldsymbol{0}$$
  - Since $e^{-\boldsymbol{A}t}\boldsymbol{y}(t)$ has a zero derivative, it must be constant, so $e^{-\boldsymbol{A}t}\boldsymbol{y}(t) = e^{-0\boldsymbol{A}}\boldsymbol{y}(0) = \boldsymbol{y}(0)$, but $\boldsymbol{y}(0) = \boldsymbol{x}_0$ since $\boldsymbol{y}$ is a solution to the differential equation
  - Therefore we've shown $e^{-\boldsymbol{A}t}\boldsymbol{y}(t) = \boldsymbol{x}_0$ and so $\boldsymbol{y}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0$, and thus $\boldsymbol{x}(t) = \boldsymbol{y}(t)$

## Computing the Matrix Exponential

### Matrix Exponential by Laplace Transform

- Consider $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0 \in \mathbb{R}^n$, taking the Laplace transform:

- $\mathcal{L}\{\dot{\boldsymbol{x}}\} = \mathcal{L}\{\boldsymbol{A}\boldsymbol{x}\}$
  $\implies s\boldsymbol{X}(s) - \boldsymbol{x}(0) = \boldsymbol{A}\boldsymbol{X}(s)$
  $\implies s\boldsymbol{X}(s) - \boldsymbol{A}\boldsymbol{X}(s) = \boldsymbol{x}(0)$
  $\implies (s\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X}(s) = \boldsymbol{x}_0$
  $\implies \boldsymbol{X}(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{x}_0$
  $\implies \boldsymbol{x}(t) = \mathcal{L}^{-1}\left\{(s\boldsymbol{I} - \boldsymbol{A})^{-1}\right\}\boldsymbol{x}_0$
- Because we know that the unique solution is $\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0$, $e^{\boldsymbol{A}t}\boldsymbol{x}_0 = \mathcal{L}^{-1}\left\{(s\boldsymbol{I} - \boldsymbol{A})^{-1}\right\}\boldsymbol{x}_0$
  - Since this holds for all $\boldsymbol{x}_0$, it must be that $e^{\boldsymbol{A}t} = \mathcal{L}^{-1}\left\{(s\boldsymbol{I} - \boldsymbol{A})^{-1}\right\}$
  - Formally, to justify this, consider the case where $\boldsymbol{x}_0 = \boldsymbol{e}_i$, i.e. all zeros except 1 in the $i$th row; substituting this into the equation we get that the $i$th column of the LHS must be equal to the $i$th column of the RHS, so do this for all $n$ columns

**Matrix Exponential by Modal Decomposition (Eigenvectors & Eigenvalues)**

- Recall that $\lambda \in \mathbb{C}$ is an eigenvalue of $\boldsymbol{A}$ if and only if $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$; i.e. $\lambda$ are the roots of the characteristic polynomial of $\boldsymbol{A}$
  - To find eigenvectors corresponding to each $\lambda$, we find a basis for $\mathcal{N}(\lambda\boldsymbol{I} - \boldsymbol{A})$ where $\mathcal{N}$ denotes the null space

**Definition**

If there exists a nonsingular matrix $\boldsymbol{P} \in \mathbb{C}^{n\times n}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is diagonal, then $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is *diagonalizable*.

**Theorem**

$\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if and only if it has $n$ linearly independent eigenvectors.

- Suppose $\boldsymbol{A}$ has $n$ linearly independent eigenvectors, and let $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n\end{bmatrix}$
  - $\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix}\boldsymbol{A}\boldsymbol{v}_1 & \cdots & \boldsymbol{A}\boldsymbol{v}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\boldsymbol{v}_1 & \cdots & \lambda_n\boldsymbol{v}_n\end{bmatrix} = \boldsymbol{P}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix} = \boldsymbol{P}\boldsymbol{\Lambda}$
  - Since $\boldsymbol{P}$ has all linearly independent columns it is invertible, therefore $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$
- Note $\boldsymbol{A}$ is diagonalizable if it has $n$ distinct eigenvalues (but diagonalizability does not always imply distinct eigenvalues); $\boldsymbol{A}$ is also diagonalizable if it is symmetric (the *spectral theorem*)

**Theorem**

If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable, then $e^{\boldsymbol{A}} = \boldsymbol{P}e^{\boldsymbol{\Lambda}}\boldsymbol{P}^{-1}$, where

$$e^{\boldsymbol{\Lambda}} = \begin{bmatrix}e^{\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_n}\end{bmatrix}$$

and $\lambda_i$ are eigenvalues of $\boldsymbol{A}$.

- We can show by induction that $\boldsymbol{A}^n = (\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1})^n = \boldsymbol{P}\boldsymbol{\Lambda}^n\boldsymbol{P}^{-1}$, then we can prove the above by substituting this into the definition of the matrix exponential, and noting that taking a diagonal matrix to a power is equivalent to taking each of the components to that power

# Lecture 6, Sep 17, 2025

## Computing the Matrix Exponential – Continued

### Complex Eigenvalues

- Consider $\boldsymbol{A} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ -1 \end{bmatrix}$; the eigenvectors and eigenvalues are $\left( \begin{bmatrix} 2 \\ i-1 \end{bmatrix}, i \right)$ and $\left( \begin{bmatrix} 2 \\ -i-1 \end{bmatrix}, -i \right)$
  - Recall that eigenvector/eigenvalues always come in complex conjugate pairs, so for a 2x2 we can take the conjugate to find the other one
  - $\boldsymbol{A}$ has two distinct eigenvalues, so it is diagonalizable
- Direct calculation: $e^{\boldsymbol{A}t} = \boldsymbol{P} e^{\boldsymbol{\Lambda} t} \boldsymbol{P}^{-1}$

$$= \cdots$$

$$= \begin{bmatrix} \sin t + \cos t & 2\sin t \\ -\sin t & \cos t - \sin t \end{bmatrix}$$

  - Note we applied Euler's formula
- Alternatively, we can define $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$ differently to avoid dealing with complex numbers
  - Denote $\boldsymbol{v} = \begin{bmatrix} 2 \\ i-1 \end{bmatrix}$ (the other eigenvector is $\bar{\boldsymbol{v}}$
  - Let $\tilde{\boldsymbol{P}} = \begin{bmatrix} \mathrm{Re}(v) & \mathrm{Im}(v) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}$
  - Let $\tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ (the imaginary parts of $\lambda_1, \lambda_2$ are in the off-diagonal entries)
  - We can show that $e^{\begin{bmatrix} a & b \\ -b & a \end{bmatrix} t} = e^{at} \begin{bmatrix} \cos(bt) & \sin(bt) \\ -\sin(bt) & \cos(bt) \end{bmatrix}$
    - When we have complex eigenvalues, we see that this results in a rotation with a rate of decay
  - Now we can compute $\tilde{\boldsymbol{P}} e^{\tilde{\boldsymbol{\Lambda}} t} \tilde{\boldsymbol{P}}^{-1}$ and this leads to the same answer

### Non-Diagonalizable Case – Jordan Forms

- Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ have less than $n$ linearly independent eigenvectors, i.e. it is non-diagonalizable; in this case we cannot form an invertible $\boldsymbol{P}$ with the eigenvectors
  - We will introduce the notion of *generalized eigenvectors* which allow us to form $\boldsymbol{P}$ in this case
- Recall that the characteristic polynomial of $\boldsymbol{A}$ is $x_A(s) = \det(s\boldsymbol{I} - \boldsymbol{A}) = \prod_{i=1}^{\sigma} (s - \lambda_i)^{m_i}$, where $\sum_{i=1}^{\sigma} m_i = n$
  - $m_i$ is the *algebraic multiplicity* of the eigenvalue $\lambda_i$
- The *minimal polynomial* $\psi_A(s)$ is the polynomial of least degree such that $\psi_A(\boldsymbol{A}) = \boldsymbol{0}$
  - We can show that such a polynomial always exists
  - The minimal polynomial has the form $\psi_A(s) = \prod_{i=1}^{\sigma} (s - \lambda_i)^{l_i}$ where $l_i \leq m_i$
    - Note that normally $s \in \mathbb{C}$; when we substitute $\boldsymbol{A}$ into the polynomial we replace $\lambda_i$ with $\lambda_i \boldsymbol{I}$
    - This is the same form as the characteristic polynomial but we may not have to repeat each term as many times, i.e. some of the information in the characteristic polynomial is redundant
  - The $l_i$ are known as the *geometric multiplicity* of $\lambda_i$
  - Each $l_i$ is also the number of linearly independent eigenvectors corresponding to $\lambda_i$ (i.e. $\dim(\mathcal{N}(\lambda_i \boldsymbol{I} - \boldsymbol{A}))$ where $\mathcal{N}$ denotes null space)
- Using this, we can decompose $\mathbb{C}^n$ into $\sigma$ subspaces: $\mathbb{C}^n = \mathcal{N}(\lambda_1 \boldsymbol{I} - \boldsymbol{A})^{l_1} \oplus \cdots \oplus \mathcal{N}(\lambda_\sigma \boldsymbol{I} - \boldsymbol{A})^{l_\sigma}$
  - $\sigma$ denotes a direct sum, $\mathcal{V} \oplus \mathcal{W} = \{ \boldsymbol{v} + \boldsymbol{w} \mid \boldsymbol{v} \in \mathcal{V}, \boldsymbol{w} \in \mathcal{W} \}$, where it is required that $\mathcal{V} \cap \mathcal{W} = \{ 0 \}$
- Example: Let $\boldsymbol{A} \in \mathbb{R}^{6 \times 6}$ have the characteristic polynomial $x_A(s) = (s - \lambda)^6$ (i.e. $\lambda$ repeated 6 times) and $\psi_A(s) = (s - \lambda)^3$ (i.e. only 3 independent eigenvectors)
  - $\mathcal{N}(\lambda \boldsymbol{I} - \boldsymbol{A}) = \mathrm{span} \{ \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3 \}$ from the 3 independent eigenvectors

- Let $\boldsymbol{v}_1, \boldsymbol{v}_2$ be linearly independent solutions to $\begin{cases}(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{v}_1 = -\boldsymbol{e}_1 \\ (\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{v}_2 = -\boldsymbol{e}_2\end{cases}$ and $\boldsymbol{w}_1$ be a solution to $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{w}_1 = -\boldsymbol{v}_1$ and all of $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{w}_1$ are linearly independent
  * Multiplying by $(\lambda \boldsymbol{I} - \boldsymbol{A})$ on both sides, the right hand side goes to zero, so we see that $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{N}((\lambda \boldsymbol{I} - \boldsymbol{A})^2)$ and $\boldsymbol{w}_1 \in \mathcal{N}((\lambda \boldsymbol{I} - \boldsymbol{A})^3)$
  * In general, a *generalized eigenvector* is a vector such that $(\lambda \boldsymbol{I} - \boldsymbol{A})^n \boldsymbol{v} = \boldsymbol{0}$ where $n \in \mathbb{N}$
  * For each of the eigenvectors, we can create an entire chain of these generalized eigenvectors
- Now notice that if we rearrange the expressions we get $\boldsymbol{A}\boldsymbol{v}_1 = \lambda \boldsymbol{v}_1 + \boldsymbol{e}_1$, and $\boldsymbol{A}\boldsymbol{w}_1 = \lambda \boldsymbol{w}_1 + \boldsymbol{v}_1$ (and so on for longer chains)
- Let $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{e}_1 & \boldsymbol{v}_1 & \boldsymbol{w}_1 & \boldsymbol{e}_2 & \boldsymbol{v}_2 & \boldsymbol{e}_3\end{bmatrix}$
  * We group together the chain related to $\boldsymbol{e}_1$, then the chain of $\boldsymbol{e}_2$ and so on
- $\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix}\lambda \boldsymbol{e}_1 & \lambda \boldsymbol{v}_1 + \boldsymbol{e}_1 & \lambda \boldsymbol{w}_1 + \boldsymbol{v}_1 & \lambda \boldsymbol{e}_2 & \lambda \boldsymbol{v}_2 + \boldsymbol{e}_2 & \lambda \boldsymbol{e}_3\end{bmatrix} = \boldsymbol{P}\begin{bmatrix}\lambda & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda\end{bmatrix} = \boldsymbol{P}\boldsymbol{J}$
- The resulting matrix $\boldsymbol{J}$ is the *Jordan form*, which is block diagonal; each of the blocks is known as a *Jordan block*
  * The number of Jordan blocks is always equal to the number of linearly independent eigenvectors
- Note there are multiple other forms that are possible, depending on how we choose to do the generalized eigenvectors: we can have 3 Jordan blocks of size 2, 1 block of size 4 and 2 blocks of size 1; as long as we only have 3 blocks, any combination is possible
  * Also, in general we have multiple distinct eigenvalues, so we have a block for each distinct eigenvalue, and then within each block we have a block for each linearly independent eigenvector for that eigenvalue (and its chain)
- In this way we can generalize the notion of diagonalizability to matrices without a full set of independent eigenvectors

**Definition**

The *Jordan form* of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ has the form

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{bmatrix}\boldsymbol{J}_{\lambda_1} & & \\ & \ddots & \\ & & \boldsymbol{J}_{\lambda_k}\end{bmatrix}$$

where $k$ is the number of distinct eigenvalues, and each $\boldsymbol{J}_{\lambda_i}$ has form

$$\boldsymbol{J}_{\lambda_i} = \begin{bmatrix}\boldsymbol{J}^1_{\lambda_i} & & \\ & \ddots & \\ & & \boldsymbol{J}^{l_i}_{\lambda_i}\end{bmatrix} \in \mathbb{C}^{m_i \times m_i}$$

where $m_i$ is the *algebraic multiplicity* of $\lambda_i$ (number of times it appears as a root in the characteristic equation) and $l_i$ is its *geometric multiplicity* (number of linearly independent eigenvectors for $\lambda_i$); each $\boldsymbol{J}^j_{\lambda_i}$ takes the form

$$\boldsymbol{J}^j_{\lambda_i} = \begin{bmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i\end{bmatrix}$$

with $\lambda_i$ repeated along the diagonal, and 1s above each $\lambda$.

- With this we can now write $e^{\boldsymbol{A}t} = \boldsymbol{P}e^{\boldsymbol{J}t}\boldsymbol{P}^{-1}$ (following the same proof as the case of normal diagonalization for $\boldsymbol{A}$)
  - The block diagonal form means $e^{\boldsymbol{J}t} = \begin{bmatrix} e^{\boldsymbol{J}_{\lambda_1}t} & & \\ & \ddots & \\ & & e^{\boldsymbol{J}_{\lambda_k}t} \end{bmatrix}$
  - Similarly $e^{\boldsymbol{J}_{\lambda_i}t} = \begin{bmatrix} e^{\boldsymbol{J}^1_{\lambda_i}} & & \\ & \ddots & \\ & & e^{\boldsymbol{J}^{l_i}_{\lambda_i}} \end{bmatrix}$
  - $\boldsymbol{J}^j_{\lambda_i}t = (\lambda_i\boldsymbol{I} + \boldsymbol{N})\,t$ where $\boldsymbol{N}$ is a matrix with 1s above the diagonal
  - Therefore $e^{\boldsymbol{J}^j_{\lambda_i}t} = e^{\lambda_i t}e^{\boldsymbol{N}t}$
    - $\boldsymbol{N}$ is a nilpotent matrix, so we can show that eventually the higher order terms in the infinite series expansion for $e^{\boldsymbol{N}t}$ all go to zero
  - We can show that $e^{\boldsymbol{N}t}$ has form $\begin{bmatrix} 1 & t & \cdots & \frac{t^{p-1}}{(p-1)!} \\ & \ddots & \ddots & \vdots \\ & & 1 & t \\ & & & 1 \end{bmatrix} \in \mathbb{C}^{p\times p}$
    - Here $p$ is the size of the block $\boldsymbol{J}^j_{\lambda_i}$ (i.e. the size of the chain of generalized eigenvectors corresponding to $\lambda_i$), which is not unique

# Lecture 7, Sep 19, 2025

## Reasoning About System Behaviour With Eigenvalues and Eigenvectors

- Consider a system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0$
  - Assume that $\boldsymbol{A}$ is diagonalizable, so the solution is $\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0 = \boldsymbol{P}e^{\boldsymbol{\Lambda}t}\boldsymbol{P}^{-1}\boldsymbol{x}_0$
- Consider the transformed coordinate space $\boldsymbol{z}(t) = \boldsymbol{P}^{-1}\boldsymbol{x}(t)$; how does the system look in this coordinate system?
  - $\dot{\boldsymbol{z}} = \boldsymbol{P}^{-1}\dot{\boldsymbol{x}}(t) = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{x}(t) = \boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1}\boldsymbol{x}(t) = \boldsymbol{\Lambda}\boldsymbol{z}(t)$
  - Since $\boldsymbol{\Lambda}$ is diagonal, we get $\dot{z}_i(t) = \lambda_i z_i(t)$, in other words, a set of $n$ decoupled linear differential equations
  - Each one is solved by $z_i(t) = e^{\lambda_i t}z_i(0)$, resulting in much easier to analyze system behaviour
  - $\boldsymbol{x}(t) = \boldsymbol{P}\boldsymbol{z}(t) = \sum_{i=1}^{n}\boldsymbol{v}_i\boldsymbol{z}_i(t) = \sum_{i=1}^{n}\boldsymbol{v}_i e^{\lambda_i t}z_i(0)$
    - Each term of the sum is called the $i$-th *mode* of $x(t)$; the entire operation is known as a *modal decomposition*
    - We denote $\boldsymbol{h}_i(t) = \sum_{i=1}^{n}\boldsymbol{v}_i e^{\lambda_i t}z_i(0)$
- Geometrically, we can imagine drawing each of the $\boldsymbol{v}_i$ as a line; if the associated $\lambda_i$ is negative, solutions shrink and go towards 0 along this line; conversely if $\lambda_i$ is positive, solutions expand and go to infinity along the line
  - For any initial condition $\boldsymbol{x}_0$ we can decompose it into components along each $\boldsymbol{v}_i$, and each of those components will evolve according to $\lambda_i$ (towards or away from the origin at a speed determined by the magnitude)
  - In the $z$ coordinate system this is easier to see since the $\boldsymbol{v}_i$ are now along the coordinate axes

### System Behaviour According to Eigenvalues

- With the above knowledge we can now categorize systems according to their eigenvalues
- Case 1: Real and nonzero eigenvalues

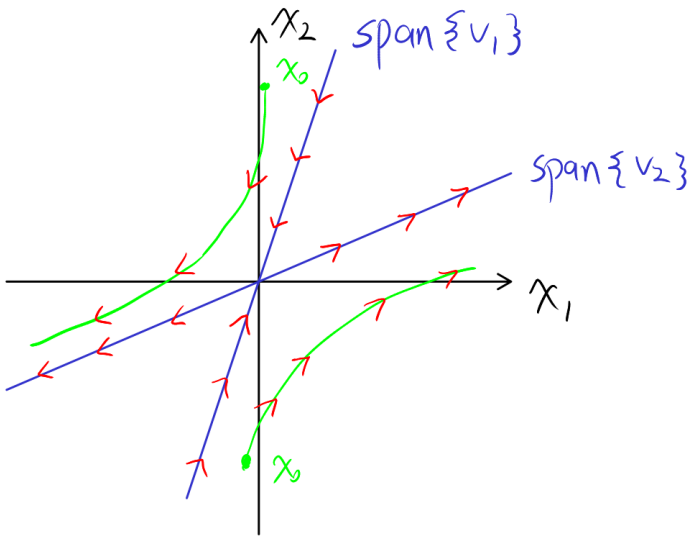


Figure 2: Illustrations of how solutions evolve in the $x$ coordinate system, for an example where $\lambda_1 < 0, \lambda_2 > 0$.

  - If all eigenvalues are less than 0, we have a *stable node* since all initial conditions converge towards zero
  - If all eigenvalues are greater than 0, we have an *unstable node* since all initial conditions explode to infinity (except for 0, which stays at 0)
  - If eigenvalues have mixed signs, we get a *saddle point* as initial conditions will move towards zero along one axis but diverge away from it on another axis; again, zero is the only initial condition that does not diverge
- Case 2: Complex conjugate eigenvalues $\lambda_1 = a + ib, \lambda_2 = a - ib$ (recall that the solution in this case is $e^{at}\begin{bmatrix}\cos(bt) & \sin(bt)\\ -\sin(bt) & \cos(bt)\end{bmatrix}$)
  - If $a < 0$, we get a *stable focus* as solutions spiral in towards zero
  - If $a > 0$, we get an *unstable focus* as solutions spiral outwards from zero towards infinity
  - If $a = 0$, we get a *centre* since all solutions stay orbiting the origin in a circle, not converging or diverging
  - In all cases, the magnitude determines the rate of spiral
- Case 3: One nonzero eigenvalue
  - The eigenvector with zero eigenvalue forms a line, where every point on the line is an equilibrium
  - If the other eigenvalue is less than zero, all solutions converge towards that line; if the other eigenvalue is greater than zero then all solutions diverge from the line
  - All initial conditions follow a straight path towards the equilibrium line, defined by the other eigenvector (nonzero eigenvalue)

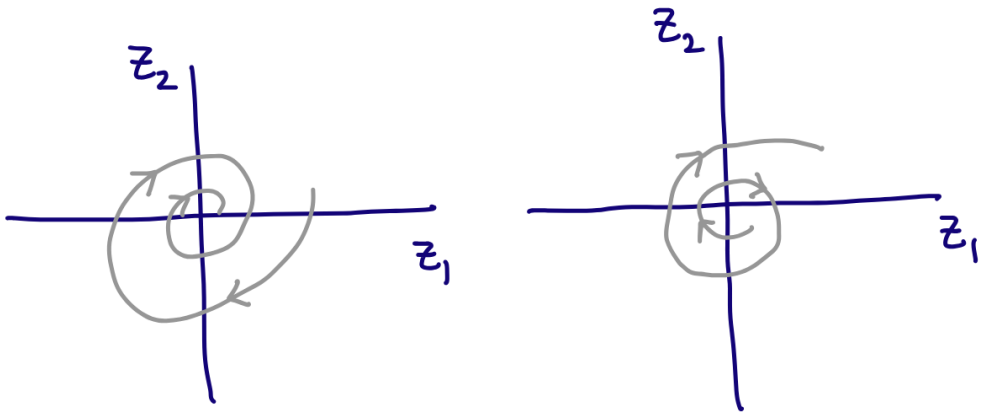


Figure 3: Behaviour for complex eigenvalues.

# Lecture 8, Sep 24, 2025

## Solution to a Non-Autonomous LTI System

**Theorem**

The solution to the LTI system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \boldsymbol{x}(0) = \boldsymbol{x}_0$$
$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}$$

is given by

$$\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0 + \int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau$$

- We can show the initial condition is satisfied trivially
- $\dot{\boldsymbol{x}} = \dfrac{\mathrm{d}}{\mathrm{d}t}\left(e^{\boldsymbol{A}t}\boldsymbol{x}_0 + \displaystyle\int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau\right)$

  $= \boldsymbol{A}e^{\boldsymbol{A}t}\boldsymbol{x}_0 + \boldsymbol{A}\displaystyle\int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau + \boldsymbol{B}\boldsymbol{u}(t)$

  $= \boldsymbol{A}\left(e^{\boldsymbol{A}t}\boldsymbol{x}_0 + \displaystyle\int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau\right) + \boldsymbol{B}\boldsymbol{u}(t)$

  $= \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t)$

  - Note $\dfrac{\mathrm{d}}{\mathrm{d}t}\displaystyle\int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau = \boldsymbol{A}\int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau + \boldsymbol{B}\boldsymbol{u}(t)$ by Leibniz rule
    - $\dfrac{\mathrm{d}}{\mathrm{d}x}\displaystyle\int_{a(x)}^{b(x)} f(x,t)\,\mathrm{d}t = f(x,b(x))\frac{\mathrm{d}}{\mathrm{d}x}b(x) - f(x,a(x))\frac{\mathrm{d}}{\mathrm{d}x}a(x) + \int_{a(x)}^{b(x)}\frac{\partial}{\partial x}f(x,t)\,\mathrm{d}t$
- Note due to the fundamental theorem of differential equations (the existence and uniqueness theorem), as our system is linear (and therefore continuous), we know that the solution above is the unique solution

# Lecture 9, Sep 24, 2025

## Stability

**Definition**

A system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0$ is said to be *stable* if for every $\boldsymbol{x}_0 \in \mathbb{R}^n$, the solution $\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0$ is *bounded*, i.e.

$$\exists M < \infty \text{ s.t. } \|\boldsymbol{x}(t)\| \leq M, \forall t \geq 0$$

The system is *asymptotically stable* if for every $\boldsymbol{x}_0 \in \mathbb{R}^n$,

$$\lim_{t\to\infty} x(t) = \mathbf{0} \in \mathbb{R}^n$$

**Theorem**

The system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}(0) = \boldsymbol{x}_0$ is asymptotically stable if and only if $\mathrm{Re}(\lambda_i) < 0$ for all eigenvalues $\lambda_i$ of $\boldsymbol{A}$.

- The intuition here is that we can decouple the system using the eigenvectors like we showed before, and if all eigenvalues have negative real parts, then all components must decay to 0
- Asymptotic stability is equivalent to $e^{\boldsymbol{A}t}$ converging to 0 as $t \to \infty$ (since this is the unique solution)
  - Recall that using the Jordan form, this is equivalent to $e^{\boldsymbol{J}_{\lambda_i}t}$ converging to 0 for each $i$, which can be expanded to $e^{\lambda_i t}\boldsymbol{N}$ where $\boldsymbol{N}$ is a matrix of polynomials of $t$
  - Suppose all $\mathrm{Re}(\lambda_i) < 0$; then $e^{\lambda_i t}$ times any polynomial of $t$ will decay to 0 as $t \to \infty$ for all $i$, since the exponential grows faster than any polynomial
  - Therefore every term in every Jordan block will converge to 0, and so $e^{\boldsymbol{A}t}$ converges to 0 and the system is asymptotically stable
- We can also define some notions of stability when the input $\boldsymbol{u}$ is involved:

**Definition**

The system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \boldsymbol{x}(0) = \boldsymbol{x}_0$$
$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}$$

is *bounded-input-bounded-output stable* (*BIBO stable*) if, when $\boldsymbol{x}_0 = \mathbf{0}$, as long as the input $\boldsymbol{u}(t)$ is bounded, the output $\boldsymbol{y}(t)$ is bounded.

The system

$$\boldsymbol{y}(t) = \int_0^t \boldsymbol{h}(t-\tau)\boldsymbol{u}(\tau)\,\mathrm{d}\tau$$

where $\boldsymbol{h} : [0, \infty) \mapsto \mathbb{R}^{m\times p}$ is BIBO stable if for all bounded $\boldsymbol{u}(t)$, $\boldsymbol{y}(t)$ is also bounded. Note this model implicitly assumes zero inital conditions.

**Definition**

The system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \boldsymbol{x}(0) = \boldsymbol{x}_0$$
$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}$$

is *input-output stable* if **for all inital conditions** $\boldsymbol{x}_0 \in \mathbb{R}^n$, a bounded $\boldsymbol{u}(t)$ implies a bounded $\boldsymbol{y}(t)$. Note input-output stability implies BIBO stability.

**Theorem**

If $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ is asymptotically stable, then for any $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$, the system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \boldsymbol{x}(0) = \boldsymbol{x}_0$$
$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{u}$$

is both BIBO and input-output stable.

- Recall that PSD for a real symmetric matrix means $\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v} \geq 0$ for all $\boldsymbol{v} \in \mathbb{R}^n$, and positive definite means $\boldsymbol{v}^T\boldsymbol{P}\boldsymbol{v} > 0$ for all nonzero $\boldsymbol{v}$
  - PSD is sometimes denoted $\boldsymbol{P} \in \mathbb{S}_n^+$; positive definite is denoted $\boldsymbol{P} \in \mathbb{S}_n$
- For complex $\boldsymbol{v}$, $\boldsymbol{v}^*\boldsymbol{P}\boldsymbol{v} > 0$ for positive definite $\boldsymbol{P}$, and greater than or equal to zero for PSD $\boldsymbol{P}$ ($\boldsymbol{v}^*$ denotes a conjugate-transpose or *Hermitian transpose*)
  - $\boldsymbol{v}^*\boldsymbol{P}\boldsymbol{v} = (\boldsymbol{x} + i\boldsymbol{y})^*\boldsymbol{P}(\boldsymbol{x} + i\boldsymbol{y})$
    $= (\boldsymbol{x}^T - i\boldsymbol{y}^T)\boldsymbol{P}(\boldsymbol{x} + i\boldsymbol{y})$
    $= \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} - i\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{x} + i\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{y} - i^2\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}$
    $= \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} + \boldsymbol{y}^T\boldsymbol{P}\boldsymbol{y}$
    * Since $\boldsymbol{P}$ is positive definite (or PSD) the last two remaining terms are both positive (or nonnegative for PSD)
    * Note that $\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{x}$ is a scalar, so we can take its transpose, and since $\boldsymbol{P}$ is symmetric we can show the expression is equal to $\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{y}$ and so the middle terms cancel

**Theorem**

Let $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ be symmetric; then $\boldsymbol{P}$ is positive definite if and only if all its eigenvalues are positive; $\boldsymbol{P}$ is positive semidefinite if and only if all its eigenvalues are non-negative. This result is sometimes known as the *spectral theorem*.
Note since $\boldsymbol{P}$ is real and symmetric, all eigenvalues are real.

This theorem also applies for negative (semi-)definite matrices and negative (nonpositive) eigenvalues.

**Theorem**

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and suppose there exists a symmetric positive definite matrix $\boldsymbol{P}$ such that

$$\boldsymbol{Q} = -\boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{P}\boldsymbol{A}$$

is also positive definite, then $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ is asymptotically stable. $\boldsymbol{Q}$ is known as the *continuous-time Lyapunov operator*.

- Let $\boldsymbol{e} \in \mathbb{C}^n$ be an eigenvector of $\boldsymbol{A}$; we know $\boldsymbol{e}^*\boldsymbol{Q}\boldsymbol{e} = \boldsymbol{e}^*(-\boldsymbol{A}^T\boldsymbol{P} - \boldsymbol{P}\boldsymbol{A})\boldsymbol{e} > 0$ and we can expand the right hand side, using $\boldsymbol{A}\boldsymbol{e} = \lambda\boldsymbol{e}$, to show that all the eigenvalues of $\boldsymbol{A}$ have negative real parts
- In discrete time, the analogous equation is $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k$, which turns out to be asymptotically stable if and only if $|\lambda_i| < 1$ for all eigenvalues of $\boldsymbol{A}$
  - The analogous definition of $\boldsymbol{Q}$ is $\mathcal{L}_d(\boldsymbol{P}) = \boldsymbol{P} - \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A}$ and it also holds that if there exists a $\boldsymbol{P}$ that makes this positive definite, then $\boldsymbol{A}$ has all eigenvalues with magnitude less than 1 and therefore the system is stable
- Now consider $\boldsymbol{A}$ with eigenvalues less than or equal to zero; if we take its Jordan form and expand $e^{\boldsymbol{J}t}$, we find that in some Jordan blocks we only have $e^{\lambda t}$, but in other blocks we have terms with $e^{\lambda t}$ times a polynomial of $t$
  - We allow the blocks that only have $e^{\lambda t}$ to have a zero eigenvalue, since this becomes a constant
  - However the blocks containing $e^{\lambda t}$ times a polynomial must have a negative eigenvalue, because

otherwise that term will grow to infinity as $t \to \infty$
  - This means for all Jordan (sub-)blocks that are bigger than $1 \times 1$, its $\lambda$ value must be strictly negative
  - Recall that we get bigger Jordan blocks when the algebraic multiplicity is greater than the geometric multiplicity for some eigenvalue
  - This is the intuition for the next theorem

**Theorem**

$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ is stable if and only if $\text{Re}(\lambda_i) \leq 0$ for all $i$, and if for all eigenvalues that have $\text{Re}(\lambda_i) = 0$, the algebraic multiplicity equals the geometric multiplicity for that eigenvalue.

- To prove this, we equate stability to $e^{\boldsymbol{A}t}$ being bounded, which is equal to $e^{\boldsymbol{J}t}$ being bounded, which we can consider separately for negative and zero eigenvalues:
  - $\text{Re}(\lambda_j) < 0$: then all $\lim_{t\to\infty} e^{\boldsymbol{J}_{\lambda_j} t} = 0$ since the exponential grows faster than all polynomials
  - $\text{Re}(\lambda_j) = 0$: $e^{\boldsymbol{J}_{\lambda_j} t}$ is bounded if and only if it has no polynomials in $t$; this only happens if we have Jordan blocks of size 1, which happens if and only if the geometric and algebraic multiplicities are equal
- Example: $\boldsymbol{A} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$; under what conditions is $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ stable? Asymptotically stable?
  - For both stability and asymptotic stability, we require $\lambda < 0$, since for this matrix $\lambda$ has algebraic multiplicity of 2 but geometric multiplicity of 1
  - We can see this from $e^{\boldsymbol{A}t} = \begin{bmatrix} e^{\lambda t} & te^{\lambda t} \\ 0 & e^{\lambda t} \end{bmatrix}$ since the matrix is already in Jordan form

**Theorem**

If $\text{Re}(\lambda_i) > 0$ for any eigenvalue $\lambda_i$ of $\boldsymbol{A}$, then $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ is unstable.

# Lecture 10, Oct 1, 2025

## Linear Algebra Concepts

**Definition**

A *vector space* $\mathcal{X}$ over a field $\mathbb{F}$ (can be e.g. $\mathbb{R}$ or $\mathbb{C}$) is a set of elements (vectors) with 2 operations: vector addition (between two elements of $\mathcal{X}$) and scalar multiplication (between an element of $\mathcal{X}$ and an element of the field $\mathbb{F}$), with the properties:

- Closure: $v + w \in \mathcal{X}$ and $\lambda v \in \mathcal{X}$
- Commutativity: $v + w = w + v$
- Associativity: $(v + w) + z = v + (w + z)$ and $(\lambda\mu)v = \lambda(\mu v)$
- Additive identity: $\exists \bar{0} \in \mathcal{X}$ such that $v + \bar{0} = v$
- Multiplicative identity: $\exists 1 \in \mathbb{F}$ and $0 \in \mathbb{F}$ such that $1v = v$ and $0v = \bar{0}$
- Additive inverse: $\exists(-v) \in \mathcal{X}$ such that $v + (-v) = \bar{0}$
- Distributivity: $\lambda(v + w) = \lambda v + \lambda w$ and $(\lambda + \mu)v = \lambda v + \mu v$

**Definition**

Let $\mathcal{X}$ be a vector space over $\mathbb{F}$ and let $x_1, \ldots, x_m \in \mathcal{X}$, then the *span* of these vectors is

$$\operatorname{span}\{x_1, \ldots, x_m\} = \left\{ \sum_{i=1}^{m} c_i x_i \;\middle|\; c_i \in \mathbb{F} \right\}$$

i.e. it is the set formed by all possible linear combinations of vectors in the set.

**Definition**

A vector space $\mathcal{X}$ is *finite-dimensional* if it can be expressed as

$$\mathcal{X} = \operatorname{span}\{x_1, \ldots, x_m\}$$

where $m \in \mathbb{N}$ is a finite integer.

The smallest $m$ that satisfies this relation is the *dimension* of $\mathcal{X}$.

- Some examples: $\mathbb{R}^n, \mathbb{R}^{m\times n}$, and the vector space of $n$-th degree polynomials $P^n$ are all finite dimensional; but the vector space of square-integrable functions in $[a, b]$, $L^2([a, b])$ is not finite dimensional

**Definition**

A set of vectors $\{x_1, \ldots, x_n\}$ is *linearly independent* if

$$\forall c_1, \ldots, c_m \in \mathbb{F}, \sum_{i=1}^{m} c_i x_i = \bar{0} \iff c_1 = \cdots = c_m = 0$$

i.e. the only linear combination of the vectors to get the zero vector is all zeros.

**Definition**

A set of vectors $\{x_1, \ldots, x_m\} \subseteq \mathcal{X}$ is a *basis* for $\mathcal{X}$ if $\mathcal{X} = \operatorname{span}\{x_1, \ldots, x_m\}$ and all vectors in the set are linearly independent.

**Definition**

Let $\mathcal{X}$ be a vector space over $\mathbb{F}$ and let $\{x_1, \ldots, x_m\}$ be a basis for $\mathcal{X}$; then any vector $v \in \mathcal{X}$ can be written as $v = c_1 x_1 + \cdots + c_m x_m$, where

$$\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \in \mathbb{F}^m$$

is the *coordinate representation* of $v$ under this basis. It can be shown that the values of $c_1, \ldots, c_m$ are uniquely determined by $v$ and the basis $\{x_1, \ldots, x_m\}$.

**Definition**

A subset $\mathcal{V} \subseteq \mathcal{X}$ is a *subspace* if it is closed, i.e. $x, y \in \mathcal{V}, \lambda \in \mathbb{F} \implies x + \lambda y \in \mathcal{V}$ and if it contains the zero vector, $\bar{0} \in \mathcal{V}$.

---

The *direct sum* of two subspaces is

$$\mathcal{V} \oplus \mathcal{W} = \{\, v + w \mid v \in \mathcal{V}, w \in \mathcal{W} \,\}$$

which can be shown to be another subspace.

**Definition**

Two subspaces $\mathcal{V}, \mathcal{W}$ of $\mathcal{X}$ are *independent* if $\mathcal{V} \cap \mathcal{W} = \{\, \bar{0}_{\mathcal{X}} \,\}$, i.e. their intersection contains only zero.

**Theorem**

Let $\mathcal{V}$ be a subspace of the vector space $\mathcal{X}$, then $\mathcal{V}$ has an *independent complement*, which is another subspace which is independent from $\mathcal{V}$ and sums with $\mathcal{V}$ to get the entirety of $\mathcal{X}$, i.e.

$$\exists \mathcal{W} \subseteq \mathcal{X} \text{ s.t. } \mathcal{V} \cap \mathcal{W} = \{\, \bar{0}_{\mathcal{X}} \,\}, \mathcal{V} \oplus \mathcal{W} = \mathcal{X}$$

Not to be confused with an orthogonal complement.

**Definition**

Let $\mathcal{X}$ be a vector space over $\mathbb{R}$. An *inner product* on $\mathcal{X}$ is an operation $\langle \cdot, \cdot \rangle : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ with the following properties:

1. $\langle x, y \rangle = \langle y, x \rangle$ (this property requires a conjugate for $\mathbb{C}$)
2. $\langle \lambda x_1 + x_2, y \rangle = \lambda \langle x_1, y \rangle + \langle x_2, y \rangle$
3. $\langle x, x \rangle \geq 0$
4. $\langle x, x \rangle = 0 \iff x = \bar{0}$

If $\langle x, y \rangle = 0$, then $x$ and $y$ are *orthogonal* under the inner product $\langle \cdot, \cdot \rangle$.

---

Combining a vector space with an inner product, $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ is called an *inner product space*.

- For example, $\mathbb{R}^n$ and $\langle x, y \rangle = x^T y$ is an inner product space
  - On $L^2([a, b])$ we can define an inner product $\langle f, g \rangle = \int_a^b f(\tau) g(\tau) \, \mathrm{d}\tau$

**Definition**

Let $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ be an inner product space and let $\mathcal{V} \subseteq \mathcal{X}$ be a subspace. Then the *orthogonal complement* of $\mathcal{V}$ in $\mathcal{X}$ is the subspace

$$\mathcal{V}^{\perp} = \{\, w \in \mathcal{X} \mid \langle w, v \rangle = 0, \forall v \in \mathcal{V} \,\}$$

Note $\mathcal{V} \oplus \mathcal{V}^{\perp} = \mathcal{X}$.

# Lecture 11, Oct 3, 2025

## Linear Maps and Matrix Representations

**Definition**

A function $f : \mathcal{X} \mapsto \mathcal{Y}$ is *injective* (one-to-one) if

$$\forall x_1, x_2 \in \mathcal{X},\ f(x_1) = f(x_2) \implies x_1 = x_2$$

or contrapositively $x_1 \neq x_2 \implies f(x_1) \neq f(x_2)$, i.e. different inputs always map to different outputs.
$f$ is *surjective* (onto) if

$$\forall y \in \mathcal{Y},\ \exists x \in \mathcal{X} \text{ s.t. } f(x) = y$$

i.e. the output reaches the entirety of $\mathcal{Y}$.

A function that is both injective and surjective is called *bijective*.

**Definition**

Let $\mathcal{X}, \mathcal{Y}$ be vector spaces, then a function $L : \mathcal{X} \mapsto \mathcal{Y}$ is a *linear transformation* (or *linear map*) if

$$\forall x_1, x_2 \in \mathcal{X}, \lambda \in \mathbb{F},\ L(x + \lambda y) = L(x) + \lambda L(y)$$

- Consider finite dimensional vector spaces $\mathcal{X}, \mathcal{Y}$ where $\{ x^1, \dots, x^n \}$ is a basis for $\mathcal{X}$ and $\{ y^1, \dots, y^m \}$ is a basis for $\mathcal{Y}$
  - For each $x_i$, $L(x_i) \in \mathcal{Y}$ so it can be expressed as coordinates $L(x^i) = \sum_{j=1}^{m} a_{ji} y^j$
  - From this, we can form $\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ where column $i$ contains the coordinates of $L(x^i)$
  - Now consider $x \in \mathcal{X} \implies x = \sum_{i=1}^{n} c_i x^i$ and $y \in \mathcal{Y} \implies \sum_{j=1}^{m} d_j y^j$ such that $L(x) = y$, then:
    * $$\begin{aligned} & L(x) = y \\ \implies & L\left( \sum_{i=1}^{n} c_i x^i \right) = \sum_{j=1}^{m} d_j y^j \\ \implies & \sum_{i=1}^{n} c_i L(x^i) = \sum_{j=1}^{m} d_j y^j \\ \implies & \sum_{i=1}^{n} c_i \sum_{j=1}^{m} a_{ji} y^j = \sum_{j=1}^{m} d_j y^j \\ \implies & \sum_{j=1}^{m} \left( \sum_{i=1}^{n} a_{ji} c_i \right) y^j = \sum_{j=1}^{m} d_j y^j \\ \implies & \sum_{i=1}^{n} a_{ji} c_i = d_i \\ \implies & \boldsymbol{A} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} \end{aligned}$$
    - Note the last step uses the uniqueness of coordinate representations

- The key idea is that we can perform a linear transformation between the abstract vector spaces $\mathcal{X}$ and $\mathcal{Y}$ by first going from $\mathcal{X}$ to $\mathbb{R}^n$ using a coordinate representation, then performing the transformation $\mathbb{R}^n \mapsto \mathbb{R}^m$ through a matrix multiplication by $\boldsymbol{A}$ to obtain coordinates for a vector in $\mathcal{Y}$, then mapping back to $\mathcal{Y}$ through the basis
- Note that a transformation has a matrix representation if and only if it is linear and maps between finite dimensional vector spaces

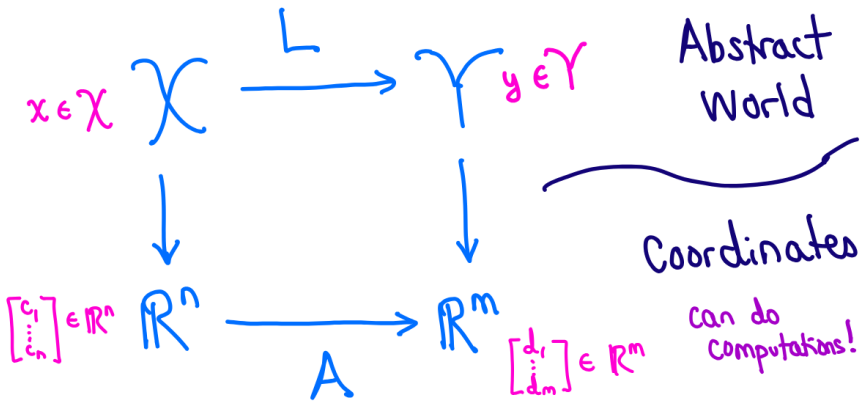


Figure 4: A matrix $\boldsymbol{A}$ as the representation of a linear transformation $L$ between two abstract vector spaces represented with coordinates.

- Example: The matrix representation of a counterclockwise rotation by $\theta$ in $\mathbb{R}^2$, using the standard basis, is $\boldsymbol{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$; what is the equivalent transformation, using the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$?
  - Denote the standard basis $\mathcal{E} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$
  - We want to find a matrix $\bar{\boldsymbol{A}}$ that takes us from $\mathbb{R}^2$ represented with $\mathcal{B}$ to another $\mathbb{R}^2$ represented with $\mathcal{B}$; we know that $\boldsymbol{A}$ takes us from $\mathbb{R}^2$ represented with $\mathcal{E}$ to another $\mathbb{R}^2$ represented with $\mathcal{E}$
  - Suppose we can get from basis $\mathcal{B}$ to basis $\mathcal{E}$ through $\boldsymbol{M}$, then we can get back to basis $\mathcal{B}$ by $\boldsymbol{M}^{-1}$
  - Therefore $\bar{\boldsymbol{A}}\boldsymbol{z} = \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M}\boldsymbol{z}$ – first applying $\boldsymbol{M}$ to get to $\mathcal{E}$, then applying $\boldsymbol{A}$ in basis $\mathcal{E}$, and then applying $\boldsymbol{M}^{-1}$ to get back to $\mathcal{B}$
    * Therefore $\bar{\boldsymbol{A}} = \boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M}$ – a similarity transform
    * Note the order that we write this is kind of reversed
  - Let $\boldsymbol{z}$ have coordinates $\begin{bmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{bmatrix}$ in $\mathcal{B}$ and $\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$ in $\mathcal{E}$, i.e. $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 \end{bmatrix} \begin{bmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$
  - We want to find $\boldsymbol{M}$ such that $\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \boldsymbol{M} \begin{bmatrix} \bar{\xi}_1 \\ \bar{\xi}_2 \end{bmatrix} \implies \boldsymbol{M} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 \end{bmatrix}$
  - Therefore $\boldsymbol{M} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and we can use this to find $\bar{\boldsymbol{A}}$

**Definition**

Let $L : \mathcal{X} \mapsto \mathcal{Y}$ be a linear transformation. The *null space* or *kernel* of $L$ is

$$\mathcal{N}(L) = \{\, \boldsymbol{x} \in \mathcal{X} \mid L(\boldsymbol{x}) = \bar{0} \,\}$$

i.e. all the vectors that map to zero. This is a subspace.

---

The *range* or *image* of $L$ is

$$\mathcal{R}(L) = \{\, \boldsymbol{y} \in \mathcal{Y} \mid \exists \boldsymbol{x} \in \mathcal{X},\, \boldsymbol{y} = L(\boldsymbol{x}) \,\}$$

i.e. all the vectors that can be reached via $L$. This is another subspace.

- Note for a subspace $\mathcal{V}$ of $\mathcal{X}$, then we denote, in general, the range of $\mathcal{V}$ under a linear transformation $L$ as $L(\mathcal{V}) = \{\, \boldsymbol{y} \in \mathcal{Y} \mid \exists \boldsymbol{x} \in \mathcal{V},\, \boldsymbol{y} = L(x) \,\}$

**Definition**

Let $L : \mathcal{X} \mapsto \mathcal{Y}$ be a linear transformation between finite dimensional vector spaces $\mathcal{X}, \mathcal{Y}$, then the *rank* of $L$ is defined as

$$\operatorname{rank}(L) = \dim(\mathcal{R}(L))$$

**Theorem**

$L : \mathcal{X} \mapsto \mathcal{Y}$ for finite dimensional $\mathcal{X}, \mathcal{Y}$ satisfies the following properties:

1. $L$ is injective if and only if $\mathcal{N}(L) = \{\, \bar{0} \,\}$
2. $\dim(\mathcal{R}(L)) + \dim(\mathcal{N}(L)) = \dim(\mathcal{X})$

- The second property (rank-nullity) can be proven as follows:
  - Let $k = \dim(\mathcal{N}(L))$ and $n = \dim(\mathcal{X})$; we want to show $n - k = \dim(\mathcal{R}(L))$
  - Let $\operatorname{span} \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k$ be a basis for $\mathcal{N}(L)$, and so $L(\boldsymbol{x}^i) = \bar{0}$ for $i \in [1, k]$
  - Complete the basis such that $\operatorname{span} \boldsymbol{x}^1, \ldots, \boldsymbol{x}^k, \boldsymbol{x}^{k+1}, \ldots, \boldsymbol{x}^n$ be a basis for $\mathcal{X}$
  - Let $\boldsymbol{x} \in \mathcal{X}$, which has a unique coordinate representation $\boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{x}^i$ with respect to this basis
  - $$\begin{aligned} L(\boldsymbol{x}) &= L\left( \sum_{i=1}^{n} c_i \boldsymbol{x}^i \right) \\ &= \sum_{i=1}^{k} c_i L(\boldsymbol{x}^i) + \sum_{i=k+1}^{n} c_i L(\boldsymbol{x}^i) \\ &= \sum_{i=k+1}^{n} c_i L(\boldsymbol{x}^i) \end{aligned}$$
  - This suggests $\{\, L(\boldsymbol{x}^i) \,\}$ for $i = k+1, \ldots, n$ forms a basis for $\mathcal{R}(L)$
    - To do this, we need to prove that they span $\mathcal{R}(L)$ and that they are linearly independent (in the notes)

**Theorem**

Let $L : \mathcal{X} \mapsto \mathcal{Y}$, then for any matrix representation $\boldsymbol{A}$ of the linear map $L$,

$$\dim(\mathcal{R}(L)) = \dim(\mathcal{R}(\boldsymbol{A})) \tag{1}$$
$$\dim(\mathcal{N}(L)) = \dim(\mathcal{N}(\boldsymbol{A})) \tag{2}$$

1. $L$ is surjective if and only if $\text{rank}(\boldsymbol{A}) = \dim(\mathcal{R}(\boldsymbol{A})) = \dim(\mathcal{Y})$, i.e. all rows of $\boldsymbol{A}$ are linearly independent (full row rank)
2. $L$ is injective if and only if $\dim(\mathcal{N}(\boldsymbol{A})) = 0$, i.e. all columns of $\boldsymbol{A}$ are linearly independent (full column rank)
3. $L$ is bijective if and only if $\boldsymbol{A}$ is square and invertible

# Lecture 12, Oct 10, 2025

## Invariant Subspaces and the Representation Theorem

**Definition**

A subspace $\mathcal{V} \in \mathbb{R}^n$ is $\boldsymbol{A}$-*invariant* for $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ if and only if

$$\forall \boldsymbol{x} \in \mathcal{V},\ \boldsymbol{A}\boldsymbol{x} \in \mathcal{V}$$

i.e. any vector in the subspace stays within the subspace under a linear transformation $\boldsymbol{A}$. We denote this by $\boldsymbol{A}\mathcal{V} \subseteq \mathcal{V}$.

- Note this is equivalent to $\forall \boldsymbol{x}_0 \in \mathcal{V}, e^{\boldsymbol{A}t}\boldsymbol{x}_0 \in \mathcal{V}$
- Some examples:
  - $\mathcal{N}(\boldsymbol{A}), \mathcal{R}(\boldsymbol{A})$ are both $\boldsymbol{A}$-invariant
  - If $\boldsymbol{w}_1, \dots, \boldsymbol{w}_n \in \mathbb{R}^n$ are eigenvectors of $\mathcal{N}$, then span $\boldsymbol{w}_1, \dots, \boldsymbol{w}_n$ is $\boldsymbol{A}$-invariant

**Theorem**

*Representation theorem*: Let $\mathcal{X}$ be a finite dimensional vector space over $\mathbb{F}$ ($\dim(\mathcal{X}) = n$) and let $L : \mathcal{X} \mapsto \mathcal{X}$ be a linear map, and let $\mathcal{V}$ be an $L$-invariant subspace of $\mathcal{X}$ ($\dim(\mathcal{V}) = k$). Then there exists a basis $\{ \boldsymbol{x}^1, \dots, \boldsymbol{x}^n \}$ for $\mathcal{X}$ such that the matrix representation of $L$ in this basis has the form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0}_{(n-k)\times k} & \boldsymbol{A}_{22} \end{bmatrix} \qquad \boldsymbol{A}_{11} \in \mathbb{F}^{k\times k}, \boldsymbol{A}_{12} \in \mathbb{F}^{k\times(n-k)}, \boldsymbol{A}_{22} \in \mathbb{F}^{(n-k)\times(n-k)}$$

- Note that if $L$ has a matrix representation $\boldsymbol{B}$ in the standard basis, then $\boldsymbol{A} = \boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$, where $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}^1, \dots, \boldsymbol{x}^n \end{bmatrix}$
- Proof:
  - $\mathcal{V}$ is a subspace so it has an independent complement $\mathcal{W}$
  - Let $\{ \boldsymbol{v}^1, \dots, \boldsymbol{v}^k \}$ be a basis for $\mathcal{V}$ and $\{ \boldsymbol{v}^{k+1}, \dots, \boldsymbol{v}^n \}$ be a basis for $\mathcal{W}$, then $\{ \boldsymbol{v}^1, \dots, \boldsymbol{v}^n \}$ is a basis for $\mathcal{X}$
  - $\mathcal{V}$ is $L$-invariant, so $L(\boldsymbol{v}^i) \in \mathcal{V}$ for $i = 1, \dots, k$ so we can express each $L(\boldsymbol{v}^i) = \sum_{i=1}^{k} a_{ji}\boldsymbol{v}^j + \sum_{l=k+1}^{n} 0\boldsymbol{v}^l$
    * For $i = k+1, \dots, n$ we no longer have $L(\boldsymbol{v}^i)$ since $\mathcal{W}$ is not $L$-invariant, so for these terms the second sum does not have all zeros
  - Recall column $i$ of the matrix representation of $L$ in this basis are the coordinates of $L(\boldsymbol{v}^i)$, so

columns $i = 1, \dots, k$ have the form $\begin{bmatrix} a_{1i} \\ \vdots \\ a_{ki} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, and the rest of the columns are nonzero in general

  - Putting it all together, we get the form of $\boldsymbol{A}$ stated in the theorem
- The representation theorem allows us to split up a linear map into parts that are invariant and parts that are not
- Consider $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ and $\mathcal{V}$ as an $\boldsymbol{A}$-invariant subspace of $\mathbb{R}^n$, then the representation theorem tells us that there exists a basis $\{ \boldsymbol{v}^1, \dots, \boldsymbol{v}^n \}$ of $\mathbb{R}^n$ such that the matrix representation of $\boldsymbol{A}$ has the form $\hat{\boldsymbol{A}} = \begin{bmatrix} \hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ \boldsymbol{0} & \hat{\boldsymbol{A}}_{22} \end{bmatrix}$
  - Now let $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{v}^1 & \cdots & \boldsymbol{v}^n \end{bmatrix}$, then $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\hat{\boldsymbol{A}}$, i.e. $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \hat{\boldsymbol{A}}$
    * $\boldsymbol{A}\boldsymbol{v}^j = \sum_{i=1}^{n} \hat{a}_{ij}\boldsymbol{v}^i = \begin{bmatrix} \boldsymbol{v}^1 & \cdots & \boldsymbol{v}^n \end{bmatrix} \begin{bmatrix} \hat{a}_{ij} \\ \vdots \\ \hat{a}_{nj} \end{bmatrix} = \boldsymbol{P} \begin{bmatrix} \hat{a}_{ij} \\ \vdots \\ \hat{a}_{nj} \end{bmatrix}$
    * Recall that for a matrix representation of a linear map, column $i$ contains the coordinates of the $i$-th basis vector after transformation by the linear map
      - This means $\hat{a}_{ij}$ are the coordinates of $\boldsymbol{A}\boldsymbol{v}^j$ with respect to our basis
      - The last column vector here is the $j$th column of $\hat{\boldsymbol{A}}$
    * Repeat this for every column
  - Let $\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{x}$ so $\dot{\boldsymbol{z}} = \boldsymbol{P}^{-1}\dot{\boldsymbol{x}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{z} = \hat{\boldsymbol{A}}\boldsymbol{z}$
  - Then $\begin{bmatrix} \dot{\boldsymbol{z}}^1 \\ \dot{\boldsymbol{z}}^2 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ \boldsymbol{0} & \hat{\boldsymbol{A}}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix}$
  - Now $\dot{\boldsymbol{z}}^1 = \hat{\boldsymbol{A}}_{11}\boldsymbol{z}^1 + \hat{\boldsymbol{A}}_{12}\boldsymbol{z}^2$ and $\dot{\boldsymbol{z}}^2 = \hat{\boldsymbol{A}}_{22}\boldsymbol{z}^2$
  - Notice now that the $\boldsymbol{z}^2$ subsystem is decoupled
  - We will later make use of this to define the Kalman decomposition and the notion of stabilizability

# Lecture 13, Oct 15, 2025

## Controllability

**Definition**

An LTI system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$ is *completely controllable* (or just controllable) if, for some positive time $T$, for all possible initial and final states $\boldsymbol{x}_0, \boldsymbol{x}_f \in \mathbb{R}^n$, there exists some piecewise continuous input $\boldsymbol{u}(t), t \in [0, T]$ that brings the system from the initial state to the final state, i.e.

$$\boldsymbol{x}_f = \boldsymbol{x}(T) = e^{\boldsymbol{A}T}\boldsymbol{x}_0 + \int_0^T e^{\boldsymbol{A}(T-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau = e^{\boldsymbol{A}\boldsymbol{T}} + L_c(\boldsymbol{u}(\cdot))$$

- $L_c(\boldsymbol{u}(\cdot))$ is a map from real piecewise continuous functions to $\mathbb{R}^n$, the impact of the input on the final state (compared to just an autonomous system)
- Let $\bar{\mathcal{R}}_T(\boldsymbol{x}_0) = \{ \boldsymbol{x}_f \in \mathbb{R}^n \mid \exists u : [0, T] \mapsto \mathbb{R}^m, \boldsymbol{x}_f = e^{\boldsymbol{A}T}\boldsymbol{x}_0 + L_c(\boldsymbol{u}(\cdot)) \}$
  - This denotes the set of all possible final states $\boldsymbol{x}_f$ that we can reach from an initial state $\boldsymbol{x}_0$ with piecewise continuous inputs $\boldsymbol{u}$
- Lemma: The LTI system $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable if and only if $\bar{\mathcal{R}}_T(0) = \mathbb{R}^n$, or equivalently $\mathcal{R}(L_c) = \mathbb{R}^n$
  - Assume $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable, then we can let $\boldsymbol{x}_0 = 0$, and there exists $\boldsymbol{u}(t)$ such that we can reach any $\boldsymbol{x}_f \in \mathbb{R}^n$; therefore by definition, $\bar{\mathcal{R}}_T(0) = \mathbb{R}^n$

### Coordinate and Feedback Transformations

- Consider the coordinate transformation $\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{x} \implies \dot{\boldsymbol{z}} = \boldsymbol{P}^{-1}\dot{\boldsymbol{x}}$
$$= \boldsymbol{P}^{-1}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u})$$
$$= \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{u}$$
  - Therefore the coordinate transform does $(\boldsymbol{A}, \boldsymbol{B}) \to (\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}, \boldsymbol{P}^{-1}\boldsymbol{B})$
- A *feedback transformation* is $\boldsymbol{u} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{v}$ where $\boldsymbol{v}$ is the new input, for some feedback matrix $\boldsymbol{K}$ (now the system's input contains feedback based on its state)
  - $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{K}\boldsymbol{x} + \boldsymbol{v}) = (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{v}$
  - The feedback transform does $(\boldsymbol{A}, \boldsymbol{B}) \to (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K}, \boldsymbol{B})$
- We will see that coordinate and feedback transformations do not affect the controllability of a system
  - This is useful because we can see the system under a different transformation, which may lead to more insights, and obtain information applicable to the original system

**Theorem**

For any nonsingular $\boldsymbol{P}$, the system $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable if and only if $(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}, \boldsymbol{P}^{-1}\boldsymbol{B})$ is completely controllable.
For any $\boldsymbol{K}$, the system $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable if and only if $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K}, \boldsymbol{B})$ is completely controllable.

In other words, controllability is invariant under coordinate and feedback transformations.

- Proof for coordinate transform invariance:
  - From the lemma, completely controllable $(\boldsymbol{A}, \boldsymbol{B})$ means $\mathcal{R}(L_c) = \mathbb{R}^n$
  - Consider the transformed $\tilde{L}_c(\boldsymbol{u}(\cdot)) = \displaystyle\int_0^T e^{\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}(T-\tau)}\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau$ so $\mathcal{R}(\tilde{L}_c) = \mathbb{R}^n$ if and only if the transformed system is controllable
  - The idea is that $L_c$ and $\tilde{L}_c$ are related by a nonsingular matrix, so they should have the same range space (similar to how $\boldsymbol{B}$ and $\boldsymbol{P}\boldsymbol{B}$ have the same range space for nonsingular $\boldsymbol{P}$)

### Controllability Matrix

**Theorem**

*Cayley-Hamilton Theorem*: Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ have characteristic polynomial

$$\det(s\boldsymbol{I} - \boldsymbol{A}) = s^n + a_{n-1}s^{n-1} + \cdots + a_0$$

then $\boldsymbol{A}$ satisfies

$$\boldsymbol{A}^n + a_{n-1}\boldsymbol{A}^{n-1} + \cdots + a_0\boldsymbol{I} = \boldsymbol{0}_{n\times n}$$

i.e. every square matrix satisfies its own characteristic equation. This allows us to express $\boldsymbol{A}^n$ and any higher powers of $\boldsymbol{A}$ as a linear combination of $\boldsymbol{I}, \boldsymbol{A}, \ldots, \boldsymbol{A}^{n-1}$.

**Definition**

The *controllability matrix* for a system $(\boldsymbol{A}, \boldsymbol{B})$ is defined as

$$\boldsymbol{Q}_c = \begin{bmatrix}\boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B}\end{bmatrix} \in \mathbb{R}^{n\times nm}$$

The system $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable if and only if $\mathcal{R}(\boldsymbol{Q}_c) = \mathbb{R}^n$, or $\operatorname{rank}(\boldsymbol{Q}_c) = n$.

- Proof ($\operatorname{rank}(\boldsymbol{Q}_c) = n \implies (\boldsymbol{A}, \boldsymbol{B})$ is controllable):

- Proof by contradiction: first assume the negation of the statement, i.e. let $\text{rank}(\boldsymbol{Q}_c) = n$ but $(\boldsymbol{A}, \boldsymbol{B})$ not controllable; we will show that this shows $\text{rank}(\boldsymbol{Q}_c) \neq n$, leading to a contradiction
- Let $\hat{L}_c$ such that $L_c(\boldsymbol{u}(\cdot)) = \displaystyle\int_0^T e^{\boldsymbol{A}(T-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau = e^{\boldsymbol{A}T}\int_0^T e^{-\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau = e^{\boldsymbol{A}T}\hat{L}_c(\boldsymbol{u}(\cdot))$
  - $(\boldsymbol{A}, \boldsymbol{B})$ not controllable means $\dim(\mathcal{R}(L_c)) < n$
  - Since $e^{\boldsymbol{A}T}$ is always invertible, $\text{rank}(L_c) = \text{rank}(\hat{L}_c) \implies \dim(\mathcal{R}(\hat{L}_c)) < n$
- $\mathcal{R}(\hat{L}_c)$ has an orthogonal component $\mathcal{R}(\hat{L}_c)^\perp$ where $\mathcal{R}(\hat{L}_c)^\perp \oplus \mathcal{R}(\hat{L}_c) = \mathbb{R}^n$, and $\dim(\mathcal{R}(\hat{L}_c)) < n$ means there exists a nonzero $\boldsymbol{v} \in \mathcal{R}(\hat{L}_c)^\perp$ orthogonal to all elements in $\hat{L}_c$
- Then for any any piecewise continuous $\boldsymbol{u}(\cdot) : [0, T] \mapsto \mathbb{R}^m$, we have $\boldsymbol{v}^T \displaystyle\int_0^T e^{-\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau = 0$
  - Consider the control input $\boldsymbol{u}^{i,s}(t) = \begin{cases} \boldsymbol{e}_i & t \in [0, s] \\ 0 & t \in (s, T]) \end{cases}$ for $s \in [0, T]$ and $i = 1, \ldots, m$
    - Note $\boldsymbol{e}_i$ denotes a vector with a 1 in the $i$th element and all 0s everywhere else
    - This input picks out the $i$th column of $\boldsymbol{B}$ for $t \in [0, s]$
  - Therefore $0 = \boldsymbol{v}^T \displaystyle\int_0^T e^{-\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{u}^{i,s}(\tau)\,\mathrm{d}\tau = \boldsymbol{v}^T \int_0^s e^{-\boldsymbol{A}\tau}\boldsymbol{b}_i \mathrm{d}\tau$ holds $\forall s \in [0, T]$
    - $\dfrac{\mathrm{d}}{\mathrm{d}s}\left(\boldsymbol{v}^T \displaystyle\int_0^s e^{-\boldsymbol{A}\tau}\boldsymbol{b}_i \mathrm{d}\tau\right) = \boldsymbol{v}^T e^{-\boldsymbol{A}s}\boldsymbol{b}_i = 0$, evaluate at $s = 0$ gives us $\boldsymbol{v}^T\boldsymbol{b}_i = 0$
    - $\dfrac{\mathrm{d}^2}{\mathrm{d}s^2}\left(\boldsymbol{v}^T \displaystyle\int_0^s e^{-\boldsymbol{A}\tau}\boldsymbol{b}_i \mathrm{d}\tau\right) = -\boldsymbol{v}^T\boldsymbol{A}e^{-\boldsymbol{A}s}\boldsymbol{b}_i = 0$, again at $s = 0$ gives $\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{b}_i = 0$
    - Do this for up to the $n$th derivative, then $\boldsymbol{v}^T\boldsymbol{A}^k\boldsymbol{b}_i = 0$ for all $k = 0, \ldots, n-1$
  - Altogether $\boldsymbol{v}^T \begin{bmatrix} \boldsymbol{b}_i & \boldsymbol{A}\boldsymbol{b}_i & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{b}_i \end{bmatrix} = \boldsymbol{0} \in \mathbb{R}^{1\times n}$
  - Repeat for all $i$, then $\boldsymbol{v}^T \begin{bmatrix} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B} \end{bmatrix} = \boldsymbol{0} \in \mathbb{R}^{1\times nm}$
- Therefore we've shown there exists a nonzero $\boldsymbol{v}$ where $\boldsymbol{v}^T\boldsymbol{Q}_c = \boldsymbol{0}$, meaning the rows of $\boldsymbol{Q}_c$ are linearly dependent; since row rank equals column rank, $\text{rank}(\boldsymbol{Q}_c) < n$

- Example: RLC circuit with $\boldsymbol{A} = \begin{bmatrix} 0 & \frac{1}{C} \\ -\frac{1}{L} & -\frac{R}{L} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 \\ \frac{1}{L} \end{bmatrix}$
  - $\boldsymbol{Q}_c = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{LC} \\ \frac{1}{L} & -\frac{R}{L^2} \end{bmatrix}$
  - For all nonzero $L, C$ this matrix has rank 2, therefore this system is always completely controllable

# Lecture 14, Oct 17, 2025

## Kalman Decomposition and Controllable Canonical Form

**Theorem**

*PBH Controllability Test*: The system $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable if and only if

$$\text{rank}(\begin{bmatrix} s\boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}) = n, \forall s \in \mathbb{C}$$

Note if $s$ is not an eigenvalue of $\boldsymbol{A}$, this matrix always has rank $n$, so the condition only needs to be tested for eigenvalues of $\boldsymbol{A}$.

- Proof:
  - Forward direction: $\text{rank}(\boldsymbol{Q}_c) = n \implies \text{rank}(\begin{bmatrix} s\boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}) = n, \forall s \in \mathbb{C}$
    - Take the contrapositive, $\exists s \in \mathbb{C}$ s.t. $\text{rank}(\begin{bmatrix} s\boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}) < n \implies \text{rank}(\boldsymbol{Q}_c) < n$
    - Since the matrix is not full rank, there exists $\boldsymbol{v}$ such that $\boldsymbol{v}^T \begin{bmatrix} s\boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix} = \boldsymbol{0}$
    - Therefore $s\boldsymbol{v}^T = \boldsymbol{v}^T\boldsymbol{A}$ and $\boldsymbol{v}^T\boldsymbol{B} = \boldsymbol{0} \in \mathbb{R}^{1\times m}$

    * Multiply by $\boldsymbol{B}$, $s\boldsymbol{v}^T\boldsymbol{B} = \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{B}$, but $\boldsymbol{v}^T\boldsymbol{B} = 0$ so $\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{B} = 0$
        - We can repeat this for all powers of $\boldsymbol{A}$, e.g. $s\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{B} = \boldsymbol{v}^T\boldsymbol{A}^2\boldsymbol{B} = 0$
    * Therefore $\boldsymbol{v}^T \begin{bmatrix}\boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{B}\end{bmatrix} = \boldsymbol{v}^T\boldsymbol{Q}_c = \boldsymbol{0}$, and so $\boldsymbol{Q}_c$ is not full rank
- Note that since multiplying by a non-singular matrix does not change rank, we can show that the PBH test is coordinate invariant
- Suppose $\text{rank}(\boldsymbol{Q}_c) < n$, i.e. $\mathcal{R}(\boldsymbol{Q}_c) \subsetneq \mathbb{R}^n$; $\mathcal{R}(\boldsymbol{Q}_c)$ has the following properties:
    - $\mathcal{R}(\boldsymbol{Q}_c)$ is $\boldsymbol{A}$-invariant
    - $\mathcal{R}(\boldsymbol{B}) \subseteq \mathcal{R}(\boldsymbol{Q}_c)$
- As a consequence of the above and the representation theorem, there exists a nonsingular matrix $\boldsymbol{P}$ such that $\begin{bmatrix}\hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ \boldsymbol{0} & \hat{\boldsymbol{A}}_{12}\end{bmatrix} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ and $\begin{bmatrix}\hat{\boldsymbol{B}}_1 \\ \boldsymbol{0}\end{bmatrix} = \boldsymbol{P}^{-1}\boldsymbol{B}$, where $\hat{\boldsymbol{A}}_{11}, \hat{\boldsymbol{B}}_1$ have dimension $\dim(\mathcal{R}(\boldsymbol{Q}_c))$
- Let $\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{x} = \begin{bmatrix}\boldsymbol{z}^1 \\ \boldsymbol{z}^2\end{bmatrix}$ where $\boldsymbol{z}^1 \in \mathbb{R}^{\dim(\mathcal{R}(\boldsymbol{Q}_c))}$ and $\boldsymbol{z}^2$ has the dimensions of its independent complement
    - $\dot{\boldsymbol{z}}^1 = \hat{\boldsymbol{A}}_{11}\boldsymbol{z}^1 + \hat{\boldsymbol{A}}_{12}\boldsymbol{z}^2 + \hat{\boldsymbol{B}}_1\boldsymbol{u}$
    - $\dot{\boldsymbol{z}}^2 = \hat{\boldsymbol{A}}_{22}\boldsymbol{z}^2$
    - This is the *Kalman decomposition for controllability*
- The Kalman decomposition separates the system into a part that we can control and a part we cannot, so if the eigenvalues of $\hat{\boldsymbol{A}}_{22}$ don't have negative real parts, our system cannot be controlled

**Definition**

For a system $(\boldsymbol{A}, \boldsymbol{b})$ where $\text{rank}(\boldsymbol{Q}_c) = k < n$, let

$$\begin{bmatrix}\hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ \boldsymbol{0} & \hat{\boldsymbol{A}}_{12}\end{bmatrix} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}, \begin{bmatrix}\hat{\boldsymbol{B}}_1 \\ \boldsymbol{0}\end{bmatrix} = \boldsymbol{P}^{-1}\boldsymbol{B}$$

for some nonsingular $\boldsymbol{P}$, where $\hat{\boldsymbol{A}}_{11}, \hat{\boldsymbol{B}}_1 \in \mathbb{R}^{k\times k}$, $\hat{\boldsymbol{A}}_{12} \in \mathbb{R}^{k\times(n-k)}$, $\hat{\boldsymbol{A}}_{22} \in \mathbb{R}^{(n-k)\times(n-k)}$.

The *Kalman decomposition* defines $\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{x} = \begin{bmatrix}\boldsymbol{z}^1 \\ \boldsymbol{z}^2\end{bmatrix}^T$ where $\boldsymbol{z}^1 \in \mathbb{R}^k, \boldsymbol{z}^2 \in \mathbb{R}^{n-k}$, so the system is decomposed as

$$\dot{\boldsymbol{z}}^1 = \hat{\boldsymbol{A}}_{11}\boldsymbol{z}^1 + \hat{\boldsymbol{A}}_{12}\boldsymbol{z}^2 + \hat{\boldsymbol{B}}_1\boldsymbol{u}$$
$$\dot{\boldsymbol{z}}^2 = \hat{\boldsymbol{A}}_{22}\boldsymbol{z}^2$$

- We can show that $(\hat{\boldsymbol{A}}_{11}, \hat{\boldsymbol{B}}_1)$ is completely controllable
    - $\boldsymbol{P}^{-1}\boldsymbol{Q}_c = \begin{bmatrix}\boldsymbol{P}^{-1}\boldsymbol{B} & \cdots \\ \boldsymbol{P}^{-1}\boldsymbol{A}^{n-1}\boldsymbol{B} & =\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{B}}_1 & \hat{\boldsymbol{A}}_{11}\hat{\boldsymbol{B}}_1 & \cdots & \hat{\boldsymbol{A}}_{11}^{n-1}\hat{\boldsymbol{B}}_1 \\ 0 & 0 & \cdots & 0\end{bmatrix}$
    - Note $k = \text{rank}(\boldsymbol{Q}_c)$, and since $\boldsymbol{P}$ is invertible, $\text{rank}(\boldsymbol{P}^{-1}Q_c) = k$
    - Since the zeros at the bottom don't affect rank, $\text{rank}(\begin{bmatrix}\hat{\boldsymbol{B}}_1 & \hat{\boldsymbol{A}}_{11}\hat{\boldsymbol{B}}_1 & \cdots & \hat{\boldsymbol{A}}_{11}^{n-1}\hat{\boldsymbol{B}}_1\end{bmatrix}) = k$
        * We're not done yet because we want the last power of $\hat{\boldsymbol{A}}_{11}$ to be $k-1$
    - By Cayley-Hamilton, we know $\hat{\boldsymbol{A}}_{11}^k, \hat{\boldsymbol{A}}_{11}^{k+1}, \ldots, \hat{\boldsymbol{A}}_{11}^{n-1}$ can all be written as a linear combination of $\hat{\boldsymbol{A}}_{11}, \ldots, \hat{\boldsymbol{A}}_{11}^{k-1}$, because $\hat{\boldsymbol{A}}_{11} \in \mathbb{R}^{k\times k}$
    - This means $\text{rank}(\begin{bmatrix}\hat{\boldsymbol{B}}_1 & \hat{\boldsymbol{A}}_{11}\hat{\boldsymbol{B}}_1 & \cdots & \hat{\boldsymbol{A}}_{11}^{n-1}\hat{\boldsymbol{B}}_1\end{bmatrix}) = \text{rank}(\begin{bmatrix}\hat{\boldsymbol{B}}_1 & \hat{\boldsymbol{A}}_{11}\hat{\boldsymbol{B}}_1 & \cdots & \hat{\boldsymbol{A}}_{11}^{k-1}\hat{\boldsymbol{B}}_1\end{bmatrix}) = k$
        * This is a simple extension of what we proved in Assignment 3
- Because $\hat{\boldsymbol{A}}$ is a block-upper-triangular matrix, its eigenvalues are the union of eigenvalues of $\hat{\boldsymbol{A}}_{11}, \hat{\boldsymbol{A}}_{22}$; furthermore, the similarity transform by $\boldsymbol{P}$ does not affect eigenvalues, so the eigenvalues of $\boldsymbol{A}$ are also this same set
    - We can control the eigenvalues of the $\hat{\boldsymbol{A}}_{11}$ subsystem; these are known as the *controllable modes/eigenvalues*
    - However we can't control the eigenvalues of $\hat{\boldsymbol{A}}_{22}$, so these are the *uncontrollable modes/eigenvalues*
    - Intuitively we can see this because the control $\boldsymbol{u}$ applies to $\hat{\boldsymbol{A}}_{11}$ but not $\hat{\boldsymbol{A}}_{22}$
- Intuition: The rank of the controllability matrix is the number of states that are controllable; therefore if the rank is $n$, then all states are controllable, but if the rank is less than $n$, some states will not be

controllable and so it might not be possible to stabilize the system
- Practically, to compute the Kalman decomposition, we need to select a basis for $\mathcal{R}(\boldsymbol{Q}_c)$ (e.g. by picking independent columns), and then select $n-k$ other linearly independent vectors that together form a basis for $\mathbb{R}^n$; then we can form $\boldsymbol{P}$ and compute $\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}$
  - The choice of basis does affect the form of $\hat{\boldsymbol{A}}_{11}$ and $\hat{\boldsymbol{A}}_{22}$, however it does not change the controllable and uncontrollable eigenvalues

**Controllable Canonical Form**

- Consider a single input system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}u$ where $(\boldsymbol{A}, \boldsymbol{b})$ is completely controllable; by choosing a special basis, we can write this in a standard form known as the *controllable canonical form*
- Let $X_A(s) = \det(s\boldsymbol{I} - \boldsymbol{A}) = s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0$, the characteristic polynomial of $\boldsymbol{A}$
- Define our series of basis vectors:
  - $\boldsymbol{v}^n = \boldsymbol{b}$
  - $\boldsymbol{v}^{n-1} = \boldsymbol{A}\boldsymbol{v}^n + a_{n-1}\boldsymbol{v}^n = \boldsymbol{A}\boldsymbol{b} + a_{n-1}\boldsymbol{b}$
  - $\boldsymbol{v}^{n-2} = \boldsymbol{A}\boldsymbol{v}^{n-1} + a_{n-2}\boldsymbol{v}^n = \boldsymbol{A}^2\boldsymbol{b} + a_{n-1}\boldsymbol{A}\boldsymbol{b} + a_{n-2}\boldsymbol{b}$
  - $\cdots$
  - $\boldsymbol{v}^1 = \boldsymbol{A}\boldsymbol{v}^2 + a_1\boldsymbol{v}^n = \boldsymbol{A}^{n-1}\boldsymbol{b} + a_{n-1}\boldsymbol{A}^{n-2}\boldsymbol{b} + \cdots + a_1\boldsymbol{b}$
- Note that $\boldsymbol{A}\boldsymbol{v}^i = \boldsymbol{v}^{i-1} - a_{i-1}\boldsymbol{v}^n$ and $\boldsymbol{A}\boldsymbol{v}^1 + a_0\boldsymbol{v}^n = 0$
  - By Cayley-Hamilton, $\boldsymbol{A}(\boldsymbol{A}^{n-1} + a_{n-1}\boldsymbol{A}^{n-2} + \cdots + a_1\boldsymbol{I}) + a_0\boldsymbol{I} = 0$
    $\implies \boldsymbol{A}(\boldsymbol{A}^{n-1} + a_{n-1}\boldsymbol{A}^{n-2} + \cdots + a_1\boldsymbol{I})\boldsymbol{b} + a_0\boldsymbol{b} = 0$
  - Notice that $\begin{aligned}\boldsymbol{A}\boldsymbol{v}^1 + a_0\boldsymbol{v}^n &= \boldsymbol{A}(\boldsymbol{A}^{n-1}\boldsymbol{b} + a_{n-1}\boldsymbol{A}^{n-2}\boldsymbol{b} + \cdots + a_1\boldsymbol{b}) + a_0\boldsymbol{b} \\ &= \boldsymbol{A}(\boldsymbol{A}^{n-1} + a_{n-1}\boldsymbol{A}^{n-2} + \cdots + a_1\boldsymbol{I})\boldsymbol{b} + a_0\boldsymbol{b} \\ &= 0\end{aligned}$
- To show that $\{\boldsymbol{v}^1, \ldots, \boldsymbol{v}^n\}$ is linearly independent:
  - $\begin{bmatrix}\boldsymbol{v}^1 & \cdots & \boldsymbol{v}^n\end{bmatrix} = \begin{bmatrix}\boldsymbol{b} & \boldsymbol{A}\boldsymbol{b} & \cdots & \boldsymbol{A}^{n-1}\boldsymbol{b}\end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & 1 \\ a_2 & a_3 & \cdots & a_{n-1} & 1 & 0 \\ a_3 & \vdots & \ddots & & & \\ \vdots & a_{n-1} & & & & \\ a_{n-1} & 1 & & & & \\ 1 & 0 & & & & \end{bmatrix} = \boldsymbol{Q}_c\boldsymbol{T}$
  - Due to this structure, $\det(\boldsymbol{T}) = (-1)^{n-1}$; since the system is controllable, we know $\boldsymbol{Q}_c$ is invertible, and therefore $\begin{bmatrix}v^1 & \cdots & v^n\end{bmatrix}$ is also invertible
- Let $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{v}^1 & \cdots & \boldsymbol{v}^n\end{bmatrix}$ and $\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{x} \implies \dot{\boldsymbol{z}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{b}\boldsymbol{u} = \tilde{\boldsymbol{A}}\boldsymbol{z} + \tilde{\boldsymbol{b}}\boldsymbol{u}$
  - $\tilde{\boldsymbol{b}} = \boldsymbol{P}^{-1}\boldsymbol{b} \implies \boldsymbol{P}\tilde{\boldsymbol{b}} = \boldsymbol{b} = \boldsymbol{v}^n$, so $\tilde{\boldsymbol{b}} = \begin{bmatrix}0 \\ \vdots \\ 0 \\ 1\end{bmatrix}$
  - $\tilde{\boldsymbol{A}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} \implies \boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\tilde{\boldsymbol{A}}$
  - $\begin{aligned}\boldsymbol{A}\boldsymbol{P} &= \begin{bmatrix}\boldsymbol{A}\boldsymbol{v}^1 & \boldsymbol{A}\boldsymbol{v}^2 & \cdots & \boldsymbol{A}\boldsymbol{v}^n\end{bmatrix} \\ &= \begin{bmatrix}-a_0\boldsymbol{v}^n & \boldsymbol{v}^1 - a_1\boldsymbol{v}^n & \boldsymbol{v}^2 - a_2\boldsymbol{v}^n & \cdots & \boldsymbol{v}^{n-1} - a_{n-1}\boldsymbol{v}^n\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{v}^1 & \cdots & \boldsymbol{v}^n\end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \\ &= \boldsymbol{P}\tilde{\boldsymbol{A}}\end{aligned}$
- We have just proven that a completely controllable system can be written in controllable canonical form; it turns out that the reverse is also true, i.e. if a system can be written in controllable canonical

form, it is always controllable

**Theorem**

A single-input system $(\boldsymbol{A}, \boldsymbol{b})$ is completely controllable if and only if there exists a nonsingular matrix $\boldsymbol{P}$, such that

$$\dot{\boldsymbol{z}} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{z} + \boldsymbol{P}^{-1}\boldsymbol{b}\boldsymbol{u} = \tilde{\boldsymbol{A}}\boldsymbol{z} + \tilde{\boldsymbol{b}}\boldsymbol{u}$$

where

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-1} \end{bmatrix} \qquad \tilde{\boldsymbol{b}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

This is known as the *controllable canonical form*.

# Lecture 15, Nov 7, 2025

## Stabilization and Pole Assignment

- The *stabilization problem* is to design a controller $\boldsymbol{u} = \boldsymbol{K}\boldsymbol{x}$ so that the resulting system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} = (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})\boldsymbol{x}$ is asymptotically stable
  - For a nonlinear system, we want $\boldsymbol{x}(t) \to \boldsymbol{x}^*(t)$ where $\boldsymbol{x}^*(t)$ is an equilibrium condition, in which case $\tilde{\boldsymbol{u}} = \boldsymbol{u} - \boldsymbol{u}^* = \boldsymbol{K}(\boldsymbol{x} - \boldsymbol{x}^*)$ and the system is $\dot{\tilde{\boldsymbol{x}}} = (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})\tilde{\boldsymbol{x}}$
- The *pole assignment problem* is to find $\boldsymbol{K}$ such that the eigenvalues of $\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K}$ are in designed locations of $\mathbb{C}$

**Theorem**

If the single-input system $(\boldsymbol{A}, \boldsymbol{b})$ is controllable, then the pole assignment problem is solvable, i.e. the eigenvalues of $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})$ can be placed arbitrarily, as long as they are in conjugate pairs.

- Proof: We can put $(\boldsymbol{A}, \boldsymbol{b})$ in controllable canonical form, $\dot{\boldsymbol{z}} = \hat{\boldsymbol{A}}\boldsymbol{z} + \hat{\boldsymbol{b}}\boldsymbol{u}$
  - Let $\hat{\boldsymbol{k}} = \begin{bmatrix} \hat{k}_1 & \cdots & \hat{k}_n \end{bmatrix}$
  - The closed-loop system looks like $\boldsymbol{A} + \hat{\boldsymbol{b}}\hat{\boldsymbol{k}} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \hat{k}_1 - a_0 & \hat{k}_2 - a_1 & \hat{k}_3 - a_2 & \cdots & \hat{k}_n - a_{n-1} \end{bmatrix}$
  - We can show $\det(s\boldsymbol{I} - (\boldsymbol{A} + \hat{\boldsymbol{b}}\hat{\boldsymbol{k}})) = s^n + (a_{n-1} - \hat{k}_n)s^{n-1} + \cdots + (a_1 - \hat{k}_2)s + (a_0 - \hat{k}_1)$
  - Therefore if we want some set of poles $\lambda_1, \ldots, \lambda_n$, we can expand $(s - \lambda_1)(s - \lambda_2)\cdots(s - \lambda_n) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$, then take $\hat{k}_i = a_{i-1} - \alpha_{i-1}$
  - To get $\boldsymbol{k}$ for the original system (if it was not originally in controllable canonical form), note $\boldsymbol{k}\boldsymbol{x} = \boldsymbol{u} = \hat{\boldsymbol{k}}\boldsymbol{z} = \hat{\boldsymbol{k}}\boldsymbol{P}^{-1}\boldsymbol{x}$, so $\boldsymbol{k} = \hat{\boldsymbol{k}}\boldsymbol{P}^{-1}$

**Theorem**

*Wonham's Pole Assignment Theorem*: Any system $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable if and only if the poles of $\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K}$ can be freely assigned, i.e. the pole assignment problem is solvable.

- Lemma: If $(\boldsymbol{A}, \boldsymbol{B})$ is completely controllable, then $\forall \boldsymbol{b} \in \mathcal{R}(\boldsymbol{B}), \exists \boldsymbol{F} \in \mathbb{R}^{m \times n}$ such that $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{F}, \boldsymbol{b})$ is controllable

- This essentially transforms the multi-input case into the single-input case
- We will not prove this in lecture, but we'll use this for the proof of Wonham's pole assignment theorem

**Summary**

To place the closed-loop poles of $(\boldsymbol{A}, \boldsymbol{b})$, i.e. set the eigenvalues of $\boldsymbol{A} + \boldsymbol{kb}$ to $\{ \lambda_1, \ldots, \lambda_n \}$:

1. Expand the desired characteristic polynomial:

$$p_{des}(s) = (s - \lambda_1) \cdots (s - \lambda_n) = s^n + \alpha_{n-1} s^{n-1} + \cdots + \alpha_1 s + \alpha_0$$

2. Expand the characteristic polynomial of $\boldsymbol{A}$:

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

3. Let $\hat{\boldsymbol{k}} = \begin{bmatrix} a_0 - \alpha_0 & \cdots & a_{n-1} - \alpha_{n-1} \end{bmatrix}$
4. Determine $\boldsymbol{P}$ required to put the system into controllable canonical form:

$$\boldsymbol{P} = \boldsymbol{Q}_c \boldsymbol{T} \qquad \boldsymbol{T} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{n-1} & 1 \\ a_2 & a_3 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1} & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

5. Let $\boldsymbol{k} = \hat{\boldsymbol{k}} \boldsymbol{P}^{-1}$ and $\boldsymbol{u} = \boldsymbol{kx}$ solves the pole assignment problem.

# Lecture 16, Nov 7, 2025

## Pole Assignment for Multi-Input Systems

- Proof of Wonham's pole assignment theorem:
  - Assume $(\boldsymbol{A}, \boldsymbol{B})$ controllable, show that we can assign the poles to $\Lambda = \{ \lambda_1, \ldots, \lambda_n \}$
    - Let some nonzero $\boldsymbol{b} = \boldsymbol{Bg}$, then by the helper lemma from the previous lecture, there exists an $\boldsymbol{F}$ such that $(\boldsymbol{A} + \boldsymbol{BF}, \boldsymbol{b})$ is controllable
    - Since this is a single-input system and controllable, we know pole assignment is solvable, i.e. $\exists \boldsymbol{H} \in \mathbb{R}^{1 \times n}$ such that $(\boldsymbol{A} + \boldsymbol{BF} + \boldsymbol{bH})$ has eigenvalues $\Lambda$
    - $\begin{aligned} \boldsymbol{A} + \boldsymbol{BF} + \boldsymbol{bH} &= \boldsymbol{A} + \boldsymbol{BF} + \boldsymbol{BgH} \\ &= \boldsymbol{A} + \boldsymbol{B}(\boldsymbol{F} + \boldsymbol{gH}) \\ &= \boldsymbol{A} + \boldsymbol{BK} \end{aligned}$
    - Therefore if we choose $\boldsymbol{K} = \boldsymbol{F} + \boldsymbol{gH}$, then the closed-loop system will have eigenvalues $\Lambda$
  - To show the reverse direction, take the contrapositive, i.e. assume that $(\boldsymbol{A}, \boldsymbol{B})$ is not controllable, show that the pole assignment problem is not solvable
    - Use the Kalman decomposition $\begin{bmatrix} \dot{\boldsymbol{z}}^1 \\ \dot{\boldsymbol{z}}^2 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ 0 & \hat{\boldsymbol{A}}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{B}}_1 \\ 0 \end{bmatrix} \boldsymbol{u}$
    - Let $\hat{\boldsymbol{K}} = \begin{bmatrix} \hat{\boldsymbol{K}}_1 & \hat{\boldsymbol{K}}_2 \end{bmatrix}$ so that $\boldsymbol{u} = \hat{\boldsymbol{K}} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix} = \hat{\boldsymbol{K}}_1 \boldsymbol{z}^1 + \hat{\boldsymbol{K}}_2 \boldsymbol{z}^2$
    - The closed-loop system is $\begin{bmatrix} \dot{\boldsymbol{z}}^1 \\ \dot{\boldsymbol{z}}^2 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{A}}_{11} + \hat{\boldsymbol{B}}_1 \hat{\boldsymbol{K}}_1 & \hat{\boldsymbol{A}}_{12} + \hat{\boldsymbol{B}}_1 \hat{\boldsymbol{K}}_2 \\ 0 & \hat{\boldsymbol{A}}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix} = \hat{\boldsymbol{M}} \boldsymbol{z}$
    - Since $\hat{\boldsymbol{M}}$ is block-upper-triangular, its eigenvalues are the union of the eigenvalues of $\hat{\boldsymbol{A}}_{11} + \hat{\boldsymbol{B}}_1 \hat{\boldsymbol{K}}_1$ and those of $\hat{\boldsymbol{A}}_{22}$
    - However, we cannot affect the eigenvalues of $\hat{\boldsymbol{A}}_{22}$ through $\hat{\boldsymbol{K}}$ at all, so in general pole assignment is not solvable

- Wonham's pole assignment theorem also leads to an eigenvalue assignment algorithm for multi-input systems, by converting it into a single-input system
  - Like in the proof, we convert the system to a single-input one, $(\boldsymbol{A}+\boldsymbol{BF}, \boldsymbol{b})$, then use the single-input pole assignment algorithm to find $\boldsymbol{H}$, and finally take $\boldsymbol{K} = \boldsymbol{F} + \boldsymbol{gH}$

# Lecture 17, Nov 7, 2025

## Stabilizability

- Last lecture we showed that we can place the poles of a closed-loop system arbitrarily if it is controllable; what if the system is not controllable? Can we make it stable?

**Definition**

$(\boldsymbol{A}, \boldsymbol{B})$ is *stabilizable* if there exists some $\boldsymbol{K}$ such that all the eigenvalues of $(\boldsymbol{A} + \boldsymbol{BK})$ have negative real part, i.e. with control law $\boldsymbol{u} = \boldsymbol{Kx}$, the resulting system is asymptotically stable.

- Stabilizability is a weaker condition than controllability, i.e. controllability implies stabilizability, but not the other way around
- Example: $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
  - Notice this is already in the Kalman decomposition form, so we can tell that the system is not controllable
  - This has eigenvalues $\{1, -1\}$, where we cannot affect the $-1$ since it is in $\hat{\boldsymbol{A}}_{22}$; however we can affect the other eigenvalue of 1 to bring it into the open left half plane
  - Consider $\boldsymbol{u} = \begin{bmatrix} k_1 & k_2 \end{bmatrix} \boldsymbol{x}$, so the closed-loop system is $\boldsymbol{A} + \boldsymbol{bk} = \begin{bmatrix} 1+k_1 & 1+k_2 \\ 0 & -1 \end{bmatrix}$
  - Therefore we can choose any $k_2$, and choose a $k_1$ such that $k_1 < -1$, so that $1 + k_1$ (the first eigenvalue) has negative real part
  - The speed of convergence is capped by the uncontrollable eigenvalue of $-1$, so regardless of our choice of $k_1$, the system cannot possibly converge faster than $e^{-t}$

**Definition**

For a system $(\boldsymbol{A}, \boldsymbol{B})$, in its Kalman decomposition, the eigenvalues of $\hat{\boldsymbol{A}}_{22}$ are the *uncontrollable eigenvalues*; the other eigenvalues, i.e. the eigenvalues of $\hat{\boldsymbol{A}}_{11}$, are the *controllable eigenvalues*. Note the eigenvalues of $\boldsymbol{A}$ (equivalently the eigenvalues of $\hat{\boldsymbol{A}}$) is the union of the eigenvalues of $\hat{\boldsymbol{A}}_{11}, \hat{\boldsymbol{A}}_{22}$.

- An equivalent definition for stabilizability is to have all the uncontrollable eigenvalues have negative real part, or equivalently all the nonnegative eigenvalues are controllable

**Theorem**

*PBH Test for Stabilizability*: $\lambda$ is a controllable eigenvalue of $(\boldsymbol{A}, \boldsymbol{B})$ (equivalently, $\lambda$ is not an eigenvalue of $\hat{\boldsymbol{A}}_{22}$) if and only if

$$\operatorname{rank}(\begin{bmatrix} \lambda \boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}) = n$$

Equivalently, $\lambda$ is an uncontrollable eigenvalue if and only if $\operatorname{rank}(\begin{bmatrix} \lambda \boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}) < n$. Therefore a system is stabilizable if and only if this matrix has rank $n$ for all non-negative eigenvalues of $\boldsymbol{A}$.

- Proof of forward direction ($\lambda$ not an eigenvalue of $\hat{\boldsymbol{A}}_{22} \implies \operatorname{rank}(\begin{bmatrix} \lambda \boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix}) = n$):
  - $\operatorname{rank}(\begin{bmatrix} s\boldsymbol{I} - \hat{\boldsymbol{A}} & \hat{\boldsymbol{B}} \end{bmatrix}) = \operatorname{rank}(\begin{bmatrix} s\boldsymbol{I} - \boldsymbol{A} & \boldsymbol{B} \end{bmatrix})$ because the two matrices are related through a matrix multiplication by a non-singular matrix
  - $\begin{bmatrix} \lambda \boldsymbol{I} - \hat{\boldsymbol{A}} & \hat{\boldsymbol{B}} \end{bmatrix} = \begin{bmatrix} \lambda \boldsymbol{I} - \hat{\boldsymbol{A}}_{11} & -\hat{\boldsymbol{A}}_{12} & \hat{\boldsymbol{B}}_1 \\ 0 & \lambda \boldsymbol{I} - \hat{\boldsymbol{A}}_{22} & 0 \end{bmatrix}$

  - If $\lambda$ is not an eigenvalue of $\hat{\boldsymbol{A}}_{22}$, then $\lambda \boldsymbol{I} - \hat{\boldsymbol{A}}_{22}$ is full-rank and therefore the bottom $n-k$ rows are linearly independent, so we only need to look at the top $k$ rows (where $k = \text{rank}(\boldsymbol{Q}_c)$)
  - We showed in lecture that the subsystem $(\hat{\boldsymbol{A}}_{11}, \hat{\boldsymbol{B}}_1)$ is completely controllable, and therefore $\text{rank}(\begin{bmatrix}\lambda \boldsymbol{I} - \hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{B}}_1\end{bmatrix}) = k$ by the PBH controllability test
  - Since adding the extra columns in $-\hat{\boldsymbol{A}}_{12}$ cannot possibly make the first $k$ rows dependent, we conclude that the first $k$ rows are linearly independent, so the overall matrix has rank $n$
- In general, if we have a system that is stabilizable but not controllable, we can find its Kalman decomposition, and design a controller to stabilize the controllable subsystem only, and then transform back
  - For the uncontrollable subsystem the gain would be arbitrary, so we usually just append zeros

# Lecture 18, Nov 14, 2025

## Observability, State Estimation, and Output Feedback Control

- Previously we discussed how to design controllers to stabilize a system given the full system state $\boldsymbol{x}$; however in practice we rarely have the full system state, so we have to estimate $\boldsymbol{x}$ using the system output $\boldsymbol{y}$
  - Given $\boldsymbol{y}(t), \boldsymbol{u}(t)$ for $0 \le t \le T$, we want to estimate $\boldsymbol{x}(t)$
- Recall that the solution is $\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0 + \int_0^t e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\mathrm{d}\tau$
  - The output is then $\boldsymbol{y}(t) = \boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0 + \int_0^t \boldsymbol{C}e^{\boldsymbol{A}(t-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\mathrm{d}\tau + \boldsymbol{D}\boldsymbol{u}(t)$
  - Notice that we know every quantity except $\boldsymbol{x}_0$, so we can solve for $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0$

**Definition**

For an LTI system, given $\boldsymbol{y}(t), \boldsymbol{u}(t), 0 \le t \le T$, the *State Estimation Problem* is to estimate $\boldsymbol{x}(t)$ for $0 \le t \le T$. Equivalently, given $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0, 0 \le t \le T$, estimate $\boldsymbol{x}_0$.

- Therefore an equivalent problem is: given $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0$ for $0 \le t \le T$, estimate $\boldsymbol{x}_0$
  - Let $L_o : \mathbb{R}^n \mapsto \mathcal{C}([0,\infty], \mathbb{R}^p)$ such that $L_o(\boldsymbol{x}_0) = \boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0$, i.e. a function mapping initial conditions to functions
    * This is a mapping that goes from $\mathbb{R}^n$ to continuous functions of time outputting $\mathbb{R}^p$
    * Note this is a linear map, but it does not have a matrix representation because the output space is infinite dimensional
  - Now given $L_o(\boldsymbol{x}_0)$, under what conditions on $L_o$ can we recover $\boldsymbol{x}_0$?
- Given vector spaces $\mathcal{X}, \mathcal{Y}$ over the field $\mathbb{R}$, and let $f : \mathcal{X} \mapsto \mathcal{Y}$ be a linear map; $f(\boldsymbol{x}) = \boldsymbol{y}$ has a unique solution if and only if $\mathcal{N}(f)$ contains only the zero vector, i.e. $f$ is injective
  - Therefore, obtaining $\boldsymbol{x}_0$ is possible if and only if $L_o$ is injective ($\mathcal{N}(L_o)$ contains only the zero vector)
  - Note in this case, a function in $\mathcal{N}(L_o)$ needs to be zero for all time

**Theorem**

Let $f : \mathcal{X} \mapsto \mathcal{Y}$ be a linear map and let $\boldsymbol{y} \in \mathcal{R}(f)$; then $f(\boldsymbol{x}) = \boldsymbol{y}$ has a unique solution if and only if $\mathcal{N}(f) = \{\bar{0}\}$, i.e. the null space is trivial. This is equivalent to $f$ being injective.

- Proof:
  - Forward direction: take contrapositive: $\mathcal{N}(f) \neq \{\bar{0}\} \implies f(\boldsymbol{x}) = \boldsymbol{y}$ does not have a unique solution
    * Let $\boldsymbol{x} \in \mathcal{X}$ be a solution, i.e. $f(\boldsymbol{x}) = \boldsymbol{y}$
    * Let $\boldsymbol{u} \in \mathcal{N}(f)$ and $\boldsymbol{u} \neq \bar{0}$, so that $f(\boldsymbol{u}) = \bar{0}$
    * Then $f(\boldsymbol{x} + \boldsymbol{u}) = f(\boldsymbol{x}) + f(\boldsymbol{u}) = \boldsymbol{y} + \bar{0} = \boldsymbol{y}$

    * Therefore both $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{u}$ are solutions, and so the solution is not unique
  - Reverse direction: again take the contrapositive: $f(\boldsymbol{x}) = \boldsymbol{y}$ does not have a unique solution $\implies \mathcal{N}(f) \neq \{\,\bar{0}\,\}$
    * Let $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$ and $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2)$
    * Then $\bar{0} = f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2) = f(\boldsymbol{x}_1 - \boldsymbol{x}_2)$
    * Therefore $\boldsymbol{x}_1 - \boldsymbol{x}_2 \in \mathcal{N}(f)$, and $\boldsymbol{x}_1 - \boldsymbol{x}_2 \neq \bar{0}$, so $\mathcal{N}(f)$ is nontrivial
- Note the relation between controllability and observability:
  - In controllability:
    * $L_c(\boldsymbol{u}(\cdot)) = \displaystyle\int_0^T e^{\boldsymbol{A}(T-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\,\mathrm{d}\tau$
    * We want $L_c$ to be surjective, so that we can find an input to command the system to any state we want
    * We form $\boldsymbol{Q}_c$ from $(\boldsymbol{A}, \boldsymbol{B})$ and use the rank of this matrix to test for controllability
  - In observability:
    * $L_o(\boldsymbol{x}_0) = \boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0$
    * We want $L_o$ to be injective, so that for any output we want to be able to find a unique initial condition (and therefore system state)
    * Likewise, we will formulate a $\boldsymbol{Q}_o$ from $(\boldsymbol{A}, \boldsymbol{C})$ and check its rank

**Definition**

The system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ is observable if $L_o : \mathbb{R}^n \mapsto \mathcal{C}([0, \infty], \mathbb{R}^p)$ is injective, or equivalently $\mathcal{N}(L_o) = \{\,\bar{0}\,\}$, where $L_o(\boldsymbol{x}_0) = \boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0$.

- Example: consider $\dot{\boldsymbol{x}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\boldsymbol{x}$ and $y = \begin{bmatrix} 1 & 0 \end{bmatrix}\boldsymbol{x}$
  - We have $\dot{x}_1(t) = x_1(t) \implies x_1(t) = e^t x_1(0)$
  - $y(t) = x_1(t) = e^t x_1(0)$
  - Intuitively we know this system is not observable since we don't have any information about $x_2$
  - Consider the case of $x_2(0) = 1$ and $x_2(0) = 2$; in both cases we have the exact same $y(t)$ as long as $x_1(0)$ is the same, so we can't recover $x_2(0)$ and therefore the system is not observable
  - In this case, $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$ are both in $\mathcal{N}(L_o)$, since both of them result in a zero $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}_0$ (which is equivalent to $y(t)$ in this case)
- Practically speaking, if we discover that our system is not observable or detectable, it means that we either need to add more sensors to measure more outputs, or reduce the model complexity, because this indicates that our model is overly complex and contains unuseful states that we cannot determine anyway

**Theorem**

Let the *observability matrix* be defined as

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

The observability matrix has the same null space as $L_o$, i.e. $\mathcal{N}(L_o) = \mathcal{N}(\boldsymbol{Q}_o)$, therefore observability is equivalent to $\mathcal{N}(\boldsymbol{Q}_o) = \{\,0\,\}$, i.e. $\operatorname{rank}(\boldsymbol{Q}_o) = n$.

- Proof:
  - $\mathcal{N}(\boldsymbol{Q}_o) \subseteq \mathcal{N}(L_o)$
    * Let $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Q}_o)$, then $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{x} = \cdots = \boldsymbol{C}\boldsymbol{A}^{n-1}\boldsymbol{x} = 0$

    * We want to show that $L_o(\boldsymbol{x}) = \boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x} = 0$
    * $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x} = \boldsymbol{C}\left(\sum_{k=0}^{\infty}\frac{\boldsymbol{A}^k}{k!}\right)\boldsymbol{x}$
    $$= \sum_{k=0}^{\infty}\frac{1}{k!}\boldsymbol{C}\boldsymbol{A}^k\boldsymbol{x}$$
    * We already know that $\boldsymbol{C}\boldsymbol{A}^k\boldsymbol{x} = 0$ for $k = 0, \ldots, n-1$ because $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Q}_o)$
    * By Cayley-Hamilton, any higher power of $\boldsymbol{A}$ can be expressed as a linear combination of $\boldsymbol{A}^i$ for $i = 0, \ldots, n-1$, so we can show that $\boldsymbol{C}\boldsymbol{A}^k\boldsymbol{x} = 0$ holds for all $k$ by an inductive proof
    * Therefore $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}$ and $\boldsymbol{x} \in \mathcal{N}(L_o)$
  - $\mathcal{N}(L_o) \subseteq \mathcal{N}(\boldsymbol{Q}_o)$
    * Let $\boldsymbol{x} \in (L_o)$, then $\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x} = 0$ for all $t$
    * Set $t = 0$ to get $\boldsymbol{C}\boldsymbol{x} = 0$
    * Take the derivative, $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{C}e^{\boldsymbol{A}t}\boldsymbol{x}) = \boldsymbol{C}\boldsymbol{A}e^{\boldsymbol{A}t}\boldsymbol{x} = 0$, and evaluate at $t = 0$ again to get $\boldsymbol{C}\boldsymbol{A}\boldsymbol{x} = 0$
    * Therefore we can show that $\boldsymbol{C}\boldsymbol{A}^i\boldsymbol{x}$ for $i = 0, \ldots, n-1$, and therefore $\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{C}\boldsymbol{A}\\ \vdots\\ \boldsymbol{C}\boldsymbol{A}^{n-1}\end{bmatrix}\boldsymbol{x} = 0$

**Kalman Decomposition for Observability**

- Similar to the controllability case, we can show that $\mathcal{N}(\boldsymbol{Q}_o)$ is $\boldsymbol{A}$-invariant, and $\mathcal{N}(\boldsymbol{Q}_o) \subseteq \mathcal{N}(\boldsymbol{C})$
  - Let $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Q}_o)$, then $\boldsymbol{C}\boldsymbol{x} = \boldsymbol{C}\boldsymbol{A}\boldsymbol{x} = \cdots = \boldsymbol{C}\boldsymbol{A}^{n-1}\boldsymbol{x} = 0$
  - Consider $\boldsymbol{Q}_o\boldsymbol{A}\boldsymbol{x}$, this has rows $\boldsymbol{C}\boldsymbol{A}\boldsymbol{x}, \ldots, \boldsymbol{C}\boldsymbol{A}^n\boldsymbol{x}$; we already know that $\boldsymbol{C}\boldsymbol{A}^k\boldsymbol{x} = 0$ up to $k = n-1$
  - For $\boldsymbol{C}\boldsymbol{A}^n\boldsymbol{x}$, use Cayley-Hamilton to expand $\boldsymbol{A}^n$ as a sum of lower powers, then we can show $\boldsymbol{C}\boldsymbol{A}^n\boldsymbol{x} = 0$
  - Therefore $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Q}_o) \implies \boldsymbol{A}\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Q}_o)$ and so $\mathcal{N}(\boldsymbol{Q}_o)$ is $\boldsymbol{A}$-invariant
  - Also, since $\boldsymbol{C}\boldsymbol{x} = 0$ from the first row of $\boldsymbol{Q}_o\boldsymbol{x} = 0$, $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{C})$, so $\mathcal{N}(\boldsymbol{Q}_o) \subseteq \mathcal{N}(\boldsymbol{C})$
- By the representation theorem, we can find a coordinate transformation $\boldsymbol{P}$ by taking the first $k = n - \operatorname{rank}(\boldsymbol{Q}_o)$ vectors as a basis for $\mathcal{N}(\boldsymbol{Q}_o)$, then the rest such that $\boldsymbol{P}$ is invertible, and let $\boldsymbol{z} = \boldsymbol{P}^{-1}\boldsymbol{x}$
- Then we get $\dot{\boldsymbol{z}} = \begin{bmatrix}\dot{\boldsymbol{z}}^1\\ \dot{\boldsymbol{z}}^2\end{bmatrix} = \begin{bmatrix}\hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12}\\ 0 & \hat{\boldsymbol{A}}_{22}\end{bmatrix}\boldsymbol{z} + \begin{bmatrix}\hat{\boldsymbol{B}}_1\\ \hat{\boldsymbol{B}}_2\end{bmatrix}, \boldsymbol{y} = \begin{bmatrix}0 & \hat{\boldsymbol{C}}_2\end{bmatrix}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u}$
  - The subsystem $\dot{\boldsymbol{z}}^2 = \hat{\boldsymbol{A}}_{22}\boldsymbol{z}^2 + \hat{\boldsymbol{B}}_2\boldsymbol{u}, \boldsymbol{y} = \hat{\boldsymbol{C}}_2\boldsymbol{z}^2 + \boldsymbol{D}\boldsymbol{u}$ is observable, while the subsystem pertaining to $\boldsymbol{z}^1$ is unobservable

**Kalman Decomposition for Controllability and Observability**

- Now we will combine what we know about observability and controllability
- Lemma: The intersection of two $\boldsymbol{A}$-invariant subspaces is also $\boldsymbol{A}$-invariant
- Consider a system that is neither controllable nor observable
  - Let $\mathcal{V}_{c\bar{o}} = \mathcal{R}(\boldsymbol{Q}_c) \cap \mathcal{N}(\boldsymbol{Q}_o)$ be the controllable but not observable subspace; let its dimension be $n_{c\bar{o}}$
    * This is $\boldsymbol{A}$-invariant because it's the intersection of two $\boldsymbol{A}$-invariant subspaces
  - Let $\mathcal{V}_{co} = \mathcal{R}(\boldsymbol{Q}_c) \setminus \mathcal{V}_{c\bar{o}}$ be the controllable and observable subspace; let its dimension be $n_{co}$
    * $\mathcal{V}_{c\bar{o}} \oplus \mathcal{V}_{co} = \mathcal{R}(\boldsymbol{Q}_c)$
  - Let $\mathcal{V}_{\bar{c}\bar{o}} = \mathcal{N}(\boldsymbol{Q}_o) \setminus \mathcal{V}_{c\bar{o}}$ be the not controllable and not observable subspace; let its dimension be $n_{\bar{c}\bar{o}}$
    * $\mathcal{V}_{c\bar{o}} \oplus \mathcal{V}_{\bar{c}\bar{o}} = \mathcal{N}(\boldsymbol{Q}_o)$
  - Let $\mathcal{V}_{\bar{c}o} = \mathbb{R}^n \setminus (\mathcal{V}_{c\bar{o}} \oplus \mathcal{V}_{co} \oplus \mathcal{V}_{\bar{c}\bar{o}})$ be the not controllable but observable subspace; let its dimension be $n_{\bar{c}o}$
    * Note that the direct sum of all subspaces is $\mathbb{R}^n$
- Let the matrix $\boldsymbol{P}$ contain the basis for $\mathcal{V}_{c\bar{o}}, \mathcal{V}_{co}, \mathcal{V}_{\bar{c}\bar{o}}, \mathcal{V}_{\bar{c}o}$ in this exact order
  - The first $n_{c\bar{o}} + n_{co}$ columns form a basis for $\mathcal{R}(\boldsymbol{Q}_c)$

- Let $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \\ \boldsymbol{z}^3 \\ \boldsymbol{z}^4 \end{bmatrix} = \boldsymbol{P}^{-1}\boldsymbol{x}$, where each of the 4 components corresponds to the 4 subspaces in the order above
- Applying the representation theorem, $\dot{\boldsymbol{z}} = \begin{bmatrix} \hat{\boldsymbol{A}}_{c\bar{o}} & * & * & * \\ 0 & \hat{\boldsymbol{A}}_{co} & 0 & * \\ 0 & 0 & \hat{\boldsymbol{A}}_{\bar{c}\bar{o}} & * \\ 0 & 0 & 0 & \hat{\boldsymbol{A}}_{\bar{c}o} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \\ \boldsymbol{z}^3 \\ \boldsymbol{z}^4 \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{B}}_{c\bar{o}} \\ \hat{\boldsymbol{B}}_{co} \\ 0 \\ 0 \end{bmatrix} \boldsymbol{u}$

$$\boldsymbol{y} = \begin{bmatrix} 0 & \hat{\boldsymbol{C}}_{co} & 0 & \hat{\boldsymbol{C}}_{\bar{c}o} \end{bmatrix} \boldsymbol{z} + \boldsymbol{D}\boldsymbol{u}$$

  - The entries marked with $*$ are nonzero in general, which don't affect our analysis
  - We first apply the Kalman Decomposition for controllability, which gives us $\hat{\boldsymbol{A}}_{11} \in \mathbb{R}^{(n_{c\bar{o}}+n_{co})\times(n_{c\bar{o}}+n_{co})}$ and so on
  - We can verify the zero entries by checking which subspace is contained in the others
- If we start in the controllable subspace, i.e. $\boldsymbol{x}(0) \in \mathcal{R}(\boldsymbol{Q}_c)$, then $\boldsymbol{z}^3(t) = \boldsymbol{z}^4(t) = 0$ for all $t$
  - $\boldsymbol{x}(0)$ is a linear combination of the basis vectors of $\mathcal{V}_{c\bar{o}}$ and $\mathcal{V}_{co}$, since those two subspaces together make up $\mathcal{R}(\boldsymbol{Q}_c)$; it doesn't have any component in the basis vectors of the other two subspaces, since those do not intersect $\mathcal{R}(\boldsymbol{Q}_c)$
  - From $\boldsymbol{x}(0) = \boldsymbol{P}\boldsymbol{z}(0)$ we see that $\boldsymbol{z}^3(0) = \boldsymbol{z}^4(0) = 0$ as a result, and this gives $\dot{\boldsymbol{z}}^3 = \dot{\boldsymbol{z}}^4 = 0$ for all time
  - Then $\begin{bmatrix} \dot{\boldsymbol{z}}^1 \\ \dot{\boldsymbol{z}}^2 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{A}}_{c\bar{o}} & * \\ 0 & \hat{\boldsymbol{A}}_{co} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{B}}_{c\bar{o}} \\ \hat{\boldsymbol{B}}_{co} \end{bmatrix} \boldsymbol{u}$

    $$\boldsymbol{y} = \begin{bmatrix} 0 & \hat{\boldsymbol{C}}_{co} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix} + \boldsymbol{D}\boldsymbol{u}$$

    * This subsystem is controllable (since we get it from the Kalman decomposition for controllability) but not observable
    * From this we can extract the controllable and observable subsystem $\dot{\boldsymbol{z}}^2 = \hat{\boldsymbol{A}}_{co}\boldsymbol{z}^2 + \hat{\boldsymbol{B}}_{co}\boldsymbol{u}$

      $$\boldsymbol{y} = \hat{\boldsymbol{C}}_{co}\boldsymbol{z}^2 + \boldsymbol{D}\boldsymbol{u}$$

- The observable subsystem is $\begin{bmatrix} \dot{\boldsymbol{z}}^2 \\ \dot{\boldsymbol{z}}^4 \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{A}}_{co} & * \\ 0 & \hat{\boldsymbol{A}}_{\bar{c}o} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^2 \\ \boldsymbol{z}^4 \end{bmatrix} + \hat{\boldsymbol{B}}_{co}\boldsymbol{u}$

$$\boldsymbol{y} = \begin{bmatrix} \hat{\boldsymbol{C}}_{co} & \hat{\boldsymbol{C}}_{\bar{c}o} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}^2 \\ \boldsymbol{z}^4 \end{bmatrix} + \boldsymbol{D}\boldsymbol{u}$$

**Theorem**

The system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ has the same transfer function as the controllable and observable subsystem, $(\hat{\boldsymbol{A}}_{co}, \hat{\boldsymbol{B}}_{co}, \hat{\boldsymbol{C}}_{co}, \hat{\boldsymbol{D}})$, which is known as the *minimal realization* of the transfer function, as it is the smallest (lowest number of states) system that results in this transfer function.

- Let $\boldsymbol{G} : \mathbb{C} \mapsto \mathbb{R}^{p\times m}$ be a transfer function matrix, and let $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ be any state space realization of the same system and $(\hat{\boldsymbol{A}}_{co}, \hat{\boldsymbol{B}}_{co}, \hat{\boldsymbol{C}}_{co}, \hat{\boldsymbol{D}})$ be the controllable and observable subsystem
  - If $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ is uncontrollable or unobservable, then the minimal realization has $k < n$ states, and so it will have $k$ poles
  - The transfer function $\boldsymbol{G}$ will also have $k$ poles, and therefore there are $n - k$ pole-zero cancellations

**Note**

Key insight: If a transfer function has one or more pole-zero cancellations, then its state space realization is either uncontrollable or unobservable.

- Example: Consider $G(s) = \dfrac{s+1}{(s+1)(s+3)}$, which has one pole-zero cancellation

- Using the uncancelled transfer function, one choice of states is $\boldsymbol{x} = \begin{bmatrix} y \\ \dot{y} - u \end{bmatrix}$, we get the realization $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \boldsymbol{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}$
  - $\boldsymbol{Q}_c = \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}$, which has rank 1, so the system is not controllable
- Using $G(s) = \dfrac{1}{s+3}$, choose $x = y$, which results in $A = -3, B = 1, C = 1$
  - This is now the minimal realization of the system
  - $Q_c = B = 1$ so clearly this system is controllable
  - $Q_o = C = 1$ so the system is observable as well

**State Observers**

- Given a control system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$, we want to design an *observer*, which estimates ("observes") the state $\boldsymbol{x}(t)$, given the known $\boldsymbol{y}(t)$ and $\boldsymbol{u}(t)$
- The observer predicts a state estimate $\hat{\boldsymbol{x}}$ as $\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{L}(\boldsymbol{y} - \hat{\boldsymbol{y}})$
$$\hat{\boldsymbol{y}} = \boldsymbol{C}\hat{\boldsymbol{x}} + \boldsymbol{D}\boldsymbol{u}$$
  - The observer is an LTI system itself which tries to simulate the system dynamics and corrects its estimate based on observations
  - $\boldsymbol{L}$ is a matrix to be designed, which corrects the estimated state based on the difference of the predicted output versus the actual measured output
- Define the *estimation error* $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$, which we can show to have dynamics $\dot{\boldsymbol{e}} = (\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C})\boldsymbol{e}$
  - Therefore we just need to choose $\boldsymbol{L}$ such that this system is asymptotically stable, i.e. make the eigenvalues of $\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C}$ have negative real part

**Definition**

The *state estimation problem* is to find a matrix $\boldsymbol{L}$ for a system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ such that the eigenvalues of $\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C}$ all have real part less than zero, so the estimation error is asymptotically stable.

A system is called *detectable* if the state estimation problem is solvable.

**Theorem**

A system $(\boldsymbol{C}, \boldsymbol{A})$ is detectable if and only if $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ is stabilizable.

- This is because the eigenvalues of $\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C}$ is the same as $(\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C})^T = \boldsymbol{A}^T - \boldsymbol{C}^T\boldsymbol{L}^T$, so if we can find $\boldsymbol{K}$ to stabilize $(\boldsymbol{A}^T, \boldsymbol{C}^T)$, then taking $\boldsymbol{L} = \boldsymbol{K}^T$ solves the state estimation problem

**Theorem**

*Duality Theorem*: Controllability and detectability are duals, i.e.
1. $(\boldsymbol{C}, \boldsymbol{A})$ is observable if and only if $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ is controllable.
2. $(\boldsymbol{C}, \boldsymbol{A})$ is detectable if and only if $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ is stabilizable.

$(\boldsymbol{A}^T, \boldsymbol{C}^T)$ is known as the *dual system* of $(\boldsymbol{C}, \boldsymbol{A})$.

- We can prove this by noticing the similarity between $\boldsymbol{Q}_o$ and $\boldsymbol{Q}_c^T$; if we take $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ and form its controllability matrix, we get $\boldsymbol{Q}_o^T$, which has the same rank as $\boldsymbol{Q}_o$, so the system is observable if and only if its dual system is controllable

**Theorem**

*PBH Test for Detectability*: $(\boldsymbol{C}, \boldsymbol{A})$ is detectable if and only if for all eigenvalues $\lambda$ of $\boldsymbol{A}$,

$$\text{Re}(\lambda) \geq 0 \implies \text{rank}\left(\begin{bmatrix}\lambda \boldsymbol{I} - \boldsymbol{A} \\ \boldsymbol{C}\end{bmatrix}\right) = n$$

- This is analogous to applying the PBH test for stabilizability to the dual system
- Proof:
  - $(\boldsymbol{C}, \boldsymbol{A})$ is detectable if and only if $(\boldsymbol{A}^T, \boldsymbol{C}^T)$ is stabilizable, which is true if and only if for all eigenvalues of $\boldsymbol{A}^T$, $\text{Re}(\lambda) \geq 0 \implies \text{rank}(\begin{bmatrix}\lambda \boldsymbol{I} - \boldsymbol{A}^T & \boldsymbol{C}^T\end{bmatrix}) = n$ (PBH stabilizability test)
  - Since transpose does not change rank or eigenvalues, this is the same as rank $\left(\begin{bmatrix}\lambda \boldsymbol{I} - \boldsymbol{A} \\ \boldsymbol{C}\end{bmatrix}\right) = n$
- Suppose $(\boldsymbol{C}, \boldsymbol{A})$ is not observable, then by applying the Kalman decomposition we get $\dot{\boldsymbol{z}} = \begin{bmatrix}\hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ 0 & \hat{\boldsymbol{A}}_{22}\end{bmatrix} \boldsymbol{z} + \begin{bmatrix}\hat{\boldsymbol{B}}_1 \\ \hat{\boldsymbol{B}}_2\end{bmatrix} \boldsymbol{u}$ and $\boldsymbol{y} = \begin{bmatrix}0 & \hat{\boldsymbol{C}}_1\end{bmatrix} \boldsymbol{z} + \boldsymbol{D}\boldsymbol{u}$ where $(\hat{\boldsymbol{C}}_1, \hat{\boldsymbol{A}}_{22})$ is the observable subsystem
- Let $\hat{\boldsymbol{L}} = \begin{bmatrix}\hat{\boldsymbol{L}}_1 \\ \hat{\boldsymbol{L}}_2\end{bmatrix}$ then $\hat{\boldsymbol{A}} - \hat{\boldsymbol{L}}\hat{\boldsymbol{C}} = \begin{bmatrix}\hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} - \hat{\boldsymbol{L}}_1\hat{\boldsymbol{C}}_1 \\ 0 & \hat{\boldsymbol{A}}_{22} - \hat{\boldsymbol{L}}_2\hat{\boldsymbol{C}}_1\end{bmatrix}$
  - This shows that we can affect the eigenvalues of the $\hat{\boldsymbol{A}}_{22}$ subsystem but not the $\hat{\boldsymbol{A}}_{11}$ subsystem
  - The eigenvalues of $\hat{\boldsymbol{A}}_{11}$ are the *unobservable eigenvalues*, and the eigenvalues of $\hat{\boldsymbol{A}}_{22}$ are the *observable eigenvalues*; together these make up all eigenvalues of $\boldsymbol{A}$
  - Just like the controllability case, by the PBH test, an eigenvalue is unobservable if and only if rank $\left(\begin{bmatrix}\lambda \boldsymbol{I} - \boldsymbol{A} \\ C\end{bmatrix}\right) < n$, and observable if and only if rank $\left(\begin{bmatrix}\lambda \boldsymbol{I} - \boldsymbol{A} \\ C\end{bmatrix}\right) = n$

**Output Feedback Stabilization**

- Putting it all together, how do we stabilize a system if we only know the output $\boldsymbol{y}(t)$ and input $\boldsymbol{u}(t)$ but not the state?
- Given a stabilizable and detectable system, our goal is to design $\boldsymbol{K}$ and $\boldsymbol{L}$ such that the eigenvalues of $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})$ and $(\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C})$ have real part less than zero, then we can implement the control law $\boldsymbol{u} = \boldsymbol{K}\hat{\boldsymbol{x}}$ and observer $\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{L}(\boldsymbol{y} - \hat{\boldsymbol{y}})$
$$\hat{\boldsymbol{y}} = \boldsymbol{C}\hat{\boldsymbol{x}} + \boldsymbol{D}\boldsymbol{u}$$
- We will show that this indeed results in an asymptotically stable system
  - Let $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$
  - $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$
    $= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{K}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{K}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{K}\boldsymbol{x}$
    $= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{K}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{K}(\boldsymbol{x} - \hat{\boldsymbol{x}})$
    $= (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})\boldsymbol{x} - \boldsymbol{B}\boldsymbol{K}\boldsymbol{e}$
  - Also, $\dot{\boldsymbol{e}} = (\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C})\boldsymbol{e}$ (shown previously)
  - Therefore $\begin{bmatrix}\dot{\boldsymbol{x}} \\ \dot{\boldsymbol{e}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K} & -\boldsymbol{B}\boldsymbol{K} \\ 0 & \boldsymbol{A} - \boldsymbol{L}\boldsymbol{C}\end{bmatrix} \begin{bmatrix}\boldsymbol{x} \\ \boldsymbol{e}\end{bmatrix}$, so if we have the eigenvalues of $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K})$ and $(\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C})$ both negative, then the overall system is stable

**Theorem**

*Separation Principle*: To stabilize a system $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D})$ through output feedback, we can design separately an asymptotically stable state feedback controller to place the eigenvalues of $\boldsymbol{A} + \boldsymbol{B}\boldsymbol{K}$, and an asymptotically stable observer to place the eigenvalues of $\boldsymbol{A} - \boldsymbol{L}\boldsymbol{C}$, then using the observer estimate for state feedback. The resulting control law $\boldsymbol{u} = \boldsymbol{K}\hat{\boldsymbol{x}}$ makes the system asymptotically stable.

# Lecture 19, Nov 26, 2025

## Linear Quadratic Optimal Control

- Consider a spring-mass system with an input force, $u - kx_1 = m\ddot{x}_1$ and let $x_2 = \dot{x}_1$, so the system is described by $\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 1 \\ -k/m & 0 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ 1/m \end{bmatrix} \boldsymbol{u}$
  - The energy of the system is $E = \frac{1}{2}kx_1^2 + \frac{1}{2}mx_2^2$
  - We can try to minimize the energy over time, $J = \int_0^\infty \frac{1}{2}kx_1^2(\tau) + \frac{1}{2}mx_2^2(\tau) + \frac{1}{2}u^2(\tau)\,\mathrm{d}\tau$
  - Note we need to include the control input in the integral to make the problem well-defined, otherwise this will just cause infinite control inputs
- More generally, we can have any cost function quadratic in the system state and control input

**Definition**

For a stabilizable system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$, let $\mathcal{U}$ be the set of admissible control laws of the form $\boldsymbol{u}(t) = \boldsymbol{\phi}(t), t \geq 0$ such that $\boldsymbol{\phi} : [0, \infty) \mapsto \mathbb{R}^m$ is continuous and asymptotically stabilizes the closed-loop system.
The *Linear Quadratic Optimal Control* problem is to find some $\boldsymbol{\phi} \in \mathcal{U}$ which minimizes a quadratic cost function,

$$J(\boldsymbol{x}, \boldsymbol{\phi}) = \int_0^\infty \boldsymbol{x}(t)^T \boldsymbol{Q}\boldsymbol{x}(t) + \boldsymbol{\phi}(t)^T \boldsymbol{R}\boldsymbol{\phi}(t)\,\mathrm{d}t$$

where $\boldsymbol{Q}$ is symmetric positive semidefinite, and $\boldsymbol{R}$ is symmetric positive definite.

- Intuitively, the reason we choose this cost function is that it usually relates to the "energy" of the system, and penalizes the state being far from the origin and the control effort being large
- Let $L(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u}$ be the instantaneous cost for the current values of $\boldsymbol{x}, \boldsymbol{u}$
- Define the optimal value function $V(\boldsymbol{x}) = \inf_{\boldsymbol{\phi} \in \mathcal{U}} J(\boldsymbol{x}, \boldsymbol{\phi}) = \inf_{\boldsymbol{\phi} \in \mathcal{U}} \int_0^\infty \boldsymbol{x}(t)^T\boldsymbol{Q}\boldsymbol{x}(t) + \boldsymbol{u}(t)^T\boldsymbol{R}\boldsymbol{u}(t)\,\mathrm{d}t$ such that $\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{u}(t), \boldsymbol{x}(0) = \boldsymbol{x}$, i.e. $V(\boldsymbol{x})$ is the lowest possible cost if we start with an initial state of $\boldsymbol{x}$
  - Note the infimum inf denotes the greatest lower bound (we can't simply use min since we don't know if one exists yet), e.g. $\inf_{n \in \mathbb{N}} \frac{1}{n} = 0$, but 0 is not in the set
- We can show that if an optimal control $\boldsymbol{\phi}^* \in \mathcal{U}$ exists, i.e. $V(x) = \min_{\boldsymbol{\phi}^* \in \mathcal{U}} J(\boldsymbol{x}, \boldsymbol{\phi}) = J(\boldsymbol{x}, \boldsymbol{\phi}^*)$, then $(V, \boldsymbol{\phi}^*)$ satisfies the *Hamilton-Jacobi-Bellman* (HJB) equation, $\min_{\boldsymbol{u} \in \mathbb{R}^m} \frac{\partial V}{\partial \boldsymbol{x}}(\boldsymbol{x}) \cdot (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}) + \boldsymbol{L}(\boldsymbol{x}, \boldsymbol{u}) = 0$
  - Intuitively, since the first term is essentially $\frac{\partial V}{\partial t}$, this means that the cost does not change instantaneously
  - Let $\tau > 0$ and suppose $\boldsymbol{\phi}^*$ exists, then $V(\boldsymbol{x}) = \int_0^\tau L(\boldsymbol{x}(t), \boldsymbol{\phi}^*(t))\,\mathrm{d}t + \int_\tau^\infty L(\boldsymbol{x}(t), \boldsymbol{\phi}^*(t))\,\mathrm{d}t$
  - The second term must minimize the cost on $[\tau, \infty)$, i.e. $\int_\tau^\infty L(\boldsymbol{x}(t), \boldsymbol{\phi}^*(t))\,\mathrm{d}t = \min_{\boldsymbol{u} \in \mathcal{U}} \int_\tau^\infty L(\boldsymbol{x}(t), \boldsymbol{u}(t))\,\mathrm{d}t$, since if this is not the case, then if we decrease $\tau$, then $V(x)$ will decrease, violating our assumption that it is optimal
    - This means that the control law must be optimal at each point in time, so we can consider only a very small initial $\tau$ (i.e. the initial step) and ignore the rest
    - This is known as *Bellman's principle of optimality*: an optimal policy has the property that, whatever the initial state and decisions so far are, the remaining decisions must constitute and optimal policy with regard to the state resulting from the decisions so far
  - Therefore $V(\boldsymbol{x}) = \min_{\boldsymbol{u} \in \mathcal{U}} \int_0^\tau L(\boldsymbol{x}(t), \boldsymbol{u}(t))\,\mathrm{d}t + V(\boldsymbol{x}(\tau))$, where $V(\boldsymbol{x}(\tau))$ is the cost from time $\tau$

onwards and known as the *cost-to-go*
  - For small $\tau$, $\int_0^\tau L(\boldsymbol{x}(t), \boldsymbol{u}(t))\,\mathrm{d}t = \tau L(\boldsymbol{x}, \boldsymbol{u}) + O(\tau)$ where $O$ has the property that $\lim_{s\to 0} \frac{O(s)}{s} = 0$
  - Using a first-order Taylor series at $\tau = 0$, $V(\boldsymbol{x}(\tau)) = V(\boldsymbol{x}) + \frac{\mathrm{d}}{\mathrm{d}t}V(\boldsymbol{x}(t))\Big|_{t=0} \tau + O(\tau)$
    * Note $\frac{\mathrm{d}}{\mathrm{d}t}V(\boldsymbol{x}(t)) = \begin{bmatrix}\frac{\partial V}{\partial x_1}(x) & \cdots & \frac{\partial V}{\partial x_n}(x)\end{bmatrix} \cdot \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \frac{\partial V}{\partial \boldsymbol{x}}(\boldsymbol{x}) \cdot (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u})$
  - Substituting both expansions into the expression earlier and some manipulation later we get the HJB equation
- The HJB equation is quadratic in $\boldsymbol{u}$, and since $\boldsymbol{R}$ is positive definite there is a unique minimum, i.e. the problem is convex, with the unique solution given by $\boldsymbol{u}^* = -\frac{1}{2}\boldsymbol{R}^{-1}\boldsymbol{B}^T\left(\frac{\partial V}{\partial \boldsymbol{x}}(\boldsymbol{x})\right)^T$
- Propose a trial solution $V(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}$, where $\boldsymbol{P}^T = \boldsymbol{P}$ is positive semidefinite, so $\frac{\partial V}{\partial \boldsymbol{x}}(\boldsymbol{x}) = 2\boldsymbol{x}^T\boldsymbol{P}$ and $\boldsymbol{u}^* = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}$
  - Substitute into the HJB equation to get $\boldsymbol{x}^T(-\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{Q})\boldsymbol{x} = 0$
  - Since this holds for any arbitrary $\boldsymbol{x} \in \mathbb{R}^n$, this means that $-\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{Q} = 0$
  - $-\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{Q} = 0$ is known as the *algebraic Ricatti equation*
- Now we need to show that our candidate optimal control law $\boldsymbol{\phi}(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t)$ does lead to the lowest cost, and stabilizes the system so it is admissible
- We will use the result that if $\boldsymbol{Q}$ is symmetric positive semidefinite, then there exists a matrix square root $\boldsymbol{Q}^{\frac{1}{2}}$ such that $(\boldsymbol{Q}^{\frac{1}{2}})^T\boldsymbol{Q}^{\frac{1}{2}} = \boldsymbol{Q}$
  - Since $\boldsymbol{Q}$ is PSD, $\boldsymbol{Q} = \boldsymbol{M}^T\boldsymbol{\Lambda}\boldsymbol{M}$ where $\boldsymbol{M}$ is an orthogonal matrix, so $\boldsymbol{Q}^{\frac{1}{2}} = \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{M}$ where $\boldsymbol{\Lambda}^{\frac{1}{2}}$ has square roots on the diagonal
- Also, all principal sub-matrices of $\boldsymbol{Q}$, i.e. any sub-matrix on the diagonal of $\boldsymbol{Q}$, must also be positive semidefinite (this can be shown by noting $\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \geq 0$ and we can choose $\boldsymbol{x}$ to isolate any sub-matrix)

**Theorem**

If $(\boldsymbol{A}, \boldsymbol{B})$ is stabilizable and $(\boldsymbol{Q}^{\frac{1}{2}}, \boldsymbol{A})$ is detectable, then the algebraic Ricatti equation

$$-\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{Q} = 0$$

has a unique symmetric positive semidefinite solution $\boldsymbol{P}$, and the control policy

$$\boldsymbol{\phi}^*(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t)$$

is the optimal controller for the linear quadratic optimal control problem, and the optimal cost is given by $V(\boldsymbol{x}) = J(\boldsymbol{x}, \boldsymbol{\phi}^*) = \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}$.

- Proof: Let $\boldsymbol{\phi} \in \mathcal{U}$ and $\boldsymbol{x}(0) = \boldsymbol{x}$, we substitute the algebraic Ricatti equation for $\boldsymbol{Q}$ into the cost, then complete the square and use $\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{B}\boldsymbol{\phi}(t)$
  - $$\begin{aligned} J(\boldsymbol{x}, \boldsymbol{\phi}) &= \int_0^\infty \boldsymbol{x}(t)^T\boldsymbol{Q}\boldsymbol{x}(t) + \boldsymbol{\phi}(t)^T\boldsymbol{R}\boldsymbol{\phi}(t)\,\mathrm{d}t \\ &= \int_0^\infty \boldsymbol{x}(t)^T(\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} - \boldsymbol{P}\boldsymbol{A} - \boldsymbol{A}^T\boldsymbol{P})\boldsymbol{x}(t) + \boldsymbol{\phi}(t)^T\boldsymbol{R}\boldsymbol{\phi}(t)\,\mathrm{d}t \\ &= \int_0^\infty (\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t))^T\boldsymbol{R}(\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t)) - \left(\dot{\boldsymbol{x}}(t)^T\boldsymbol{P}\boldsymbol{x}(t) + \boldsymbol{x}(t)\boldsymbol{P}\dot{\boldsymbol{x}}(t)\right)\,\mathrm{d}t \\ &= \int_0^\infty (\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t))^T\boldsymbol{R}(\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t))\,\mathrm{d}t - \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{x}(t)^T\boldsymbol{P}\boldsymbol{x}(t))\,\mathrm{d}t \\ &= \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} + \int_0^\infty (\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t))^T\boldsymbol{R}(\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t))\,\mathrm{d}t \end{aligned}$$
  - By the positive semidefiniteness of $\boldsymbol{R}$, the term inside the integral is always nonnegative, and it is

minimized when $\boldsymbol{\phi}(t) + \boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t)$, i.e. $\boldsymbol{\phi}^*(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t)$, which gives us the optimal cost $\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}$

**Summary**

For a system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}$, linear quadratic optimal control seeks to optimize a cost functional

$$J(\boldsymbol{x}, \boldsymbol{\phi}) = \int_0^\infty \boldsymbol{x}(t)^T\boldsymbol{Q}\boldsymbol{x}(t) + \boldsymbol{\phi}(t)^T\boldsymbol{R}\boldsymbol{\phi}(t)\mathrm{d}t$$

where $\boldsymbol{u} = \boldsymbol{\phi}, \boldsymbol{x}(0) = \boldsymbol{x}$, assuming that $(\boldsymbol{A}, \boldsymbol{B})$ is stabilizable and $(\boldsymbol{Q}^{\frac{1}{2}}, \boldsymbol{A})$ is detectable. Then the optimal control law is given by

$$\boldsymbol{\phi}^*(t) = -\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{x}(t)$$

where $\boldsymbol{P}$ is the unique positive semidefinite solution to the algebraic Ricatti equation,

$$-\boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{Q} = 0$$