

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

Definition

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

- Therefore an equivalent problem is: given $\mathbf{C} e^{\mathbf{A}t} \mathbf{x}_0$ for $0 \leq t \leq T$, estimate \mathbf{x}_0
 - Let $L_o : \mathbb{R}^n \mapsto \mathcal{C}([0, \infty], \mathbb{R}^p)$ such that $L_o(\mathbf{x}_0) = \mathbf{C} e^{\mathbf{A}t} \mathbf{x}_0$, 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 $L_o(\mathbf{x}_0)$, under what conditions on L_o can we recover \mathbf{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; $f(\mathbf{x}) = \mathbf{y}$ has a unique solution if and only if $\mathcal{N}(f)$ contains only the zero vector, i.e. f is injective
 - Therefore, obtaining \mathbf{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 $\mathbf{y} \in \mathcal{R}(f)$; then $f(\mathbf{x}) = \mathbf{y}$ 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: $\mathcal{N}(f) \neq \{\bar{0}\} \implies f(\mathbf{x}) = \mathbf{y}$ does not have a unique solution
 - Let $\mathbf{x} \in \mathcal{X}$ be a solution, i.e. $f(\mathbf{x}) = \mathbf{y}$
 - Let $\mathbf{u} \in \mathcal{N}(f)$ and $\mathbf{u} \neq \bar{0}$, so that $f(\mathbf{u}) = \bar{0}$
 - Then $f(\mathbf{x} + \mathbf{u}) = f(\mathbf{x}) + f(\mathbf{u}) = \mathbf{y} + \bar{0} = \mathbf{y}$
 - Therefore both \mathbf{x} and $\mathbf{x} + \mathbf{u}$ are solutions, and so the solution is not unique
 - Reverse direction: again take the contrapositive: $f(\mathbf{x}) = \mathbf{y}$ does not have a unique solution $\implies \mathcal{N}(f) \neq \{\bar{0}\}$
 - Let $\mathbf{x}_1 \neq \mathbf{x}_2$ and $f(\mathbf{x}_1) = f(\mathbf{x}_2)$
 - Then $\bar{0} = f(\mathbf{x}_1) - f(\mathbf{x}_2) = f(\mathbf{x}_1 - \mathbf{x}_2)$
 - Therefore $\mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{N}(f)$, and $\mathbf{x}_1 - \mathbf{x}_2 \neq \bar{0}$, so $\mathcal{N}(f)$ is nontrivial
- Note the relation between controllability and observability:
 - In controllability:
 - * $L_c(\mathbf{u}(\cdot)) = \int_0^T e^{\mathbf{A}(T-\tau)} \mathbf{B} \mathbf{u}(\tau) d\tau$
 - * 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 \mathbf{Q}_o from (\mathbf{A}, \mathbf{B}) and use the rank of this matrix to test for controllability
- In observability:
 - * $L_o(\mathbf{x}_0) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}_0$
 - * 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 \mathbf{Q}_o from (\mathbf{A}, \mathbf{C}) and check its rank

Definition

The system $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ 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 $L_o(\mathbf{x}_0) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}_0$.

- Example: consider $\dot{\mathbf{x}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \mathbf{x}$ and $y = [1 \ 0] \mathbf{x}$
 - 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 $\mathbf{C}e^{\mathbf{A}t}\mathbf{x}_0$ (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

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

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

- Proof:
 - $\mathcal{N}(\mathbf{Q}_o) \subseteq \mathcal{N}(L_o)$
 - * Let $\mathbf{x} \in \mathcal{N}(\mathbf{Q}_o)$, then $\mathbf{C}\mathbf{x} = \mathbf{CA}\mathbf{x} = \dots = \mathbf{CA}^{n-1}\mathbf{x} = 0$
 - * We want to show that $L_o(\mathbf{x}) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x} = 0$
 - * $\mathbf{C}e^{\mathbf{A}t}\mathbf{x} = \mathbf{C} \left(\sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!} \right) \mathbf{x}$
 $= \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{CA}^k \mathbf{x}$
 - * We already know that $\mathbf{CA}^k \mathbf{x} = 0$ for $k = 0, \dots, n-1$ because $\mathbf{x} \in \mathcal{N}(\mathbf{Q}_o)$
 - * By Cayley-Hamilton, any higher power of \mathbf{A} can be expressed as a linear combination of \mathbf{A}^i for $i = 0, \dots, n-1$, so we can show that $\mathbf{CA}^k \mathbf{x} = 0$ holds for all k by an inductive proof
 - * Therefore $\mathbf{C}e^{\mathbf{A}t}\mathbf{x} = 0$ and $\mathbf{x} \in \mathcal{N}(L_o)$
 - $\mathcal{N}(L_o) \subseteq \mathcal{N}(\mathbf{Q}_o)$
 - * Let $\mathbf{x} \in \mathcal{N}(L_o)$, then $\mathbf{C}e^{\mathbf{A}t}\mathbf{x} = 0$ for all t

- * Set $t = 0$ to get $\mathbf{C}\mathbf{x} = 0$
- * Take the derivative, $\frac{d}{dt}(\mathbf{C}e^{\mathbf{A}t}\mathbf{x}) = \mathbf{C}\mathbf{A}e^{\mathbf{A}t}\mathbf{x} = 0$, and evaluate at $t = 0$ again to get $\mathbf{C}\mathbf{A}\mathbf{x} = 0$
- * Therefore we can show that $\mathbf{C}\mathbf{A}^i\mathbf{x}$ for $i = 0, \dots, n-1$, and therefore $\begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{n-1} \end{bmatrix} \mathbf{x} = 0$

Kalman Decomposition for Observability

- Similar to the controllability case, we can show that $\mathcal{N}(\mathbf{Q}_o)$ is \mathbf{A} -invariant, and $\mathcal{N}(\mathbf{Q}_o) \subseteq \mathcal{N}(\mathbf{C})$
 - Let $\mathbf{x} \in \mathcal{N}(\mathbf{Q}_o)$, then $\mathbf{C}\mathbf{x} = \mathbf{C}\mathbf{A}\mathbf{x} = \dots = \mathbf{C}\mathbf{A}^{n-1}\mathbf{x} = 0$
 - Consider $\mathbf{Q}_o\mathbf{A}\mathbf{x}$, this has rows $\mathbf{C}\mathbf{A}\mathbf{x}, \dots, \mathbf{C}\mathbf{A}^n\mathbf{x}$; we already know that $\mathbf{C}\mathbf{A}^k\mathbf{x} = 0$ up to $k = n-1$
 - For $\mathbf{C}\mathbf{A}^n\mathbf{x}$, use Cayley-Hamilton to expand \mathbf{A}^n as a sum of lower powers, then we can show $\mathbf{C}\mathbf{A}^n\mathbf{x} = 0$
 - Therefore $\mathbf{x} \in \mathcal{N}(\mathbf{Q}_o) \implies \mathbf{A}\mathbf{x} \in \mathcal{N}(\mathbf{Q}_o)$ and so $\mathcal{N}(\mathbf{Q}_o)$ is \mathbf{A} -invariant
 - Also, since $\mathbf{C}\mathbf{x} = 0$ from the first row of $\mathbf{Q}_o\mathbf{x} = 0$, $\mathbf{x} \in \mathcal{N}(\mathbf{C})$, so $\mathcal{N}(\mathbf{Q}_o) \subseteq \mathcal{N}(\mathbf{C})$
- By the representation theorem, we can find a coordinate transformation \mathbf{P} by taking the first $k = n - \text{rank}(\mathbf{Q}_o)$ vectors as a basis for $\mathcal{N}(\mathbf{Q}_o)$, then the rest such that \mathbf{P} is invertible, and let $\mathbf{z} = \mathbf{P}^{-1}\mathbf{x}$
- Then we get $\dot{\mathbf{z}} = \begin{bmatrix} \dot{\mathbf{z}}^1 \\ \dot{\mathbf{z}}^2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ 0 & \hat{\mathbf{A}}_{22} \end{bmatrix} \mathbf{z} + \begin{bmatrix} \hat{\mathbf{B}}_1 \\ \hat{\mathbf{B}}_2 \end{bmatrix}, \mathbf{y} = [0 \ \hat{\mathbf{C}}_2] \mathbf{z} + \mathbf{D}\mathbf{u}$
 - The subsystem $\dot{\mathbf{z}}^2 = \hat{\mathbf{A}}_{22}\mathbf{z}^2 + \hat{\mathbf{B}}_2\mathbf{u}, \mathbf{y} = \hat{\mathbf{C}}_2\mathbf{z}^2 + \mathbf{D}\mathbf{u}$ is observable, while the subsystem pertaining to \mathbf{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 \mathbf{A} -invariant subspaces is also \mathbf{A} -invariant
- Consider a system that is neither controllable nor observable
 - Let $\mathcal{V}_{c\bar{o}} = \mathcal{R}(\mathbf{Q}_c) \cap \mathcal{N}(\mathbf{Q}_o)$ be the controllable but not observable subspace; let its dimension be $n_{c\bar{o}}$
 - * This is \mathbf{A} -invariant because it's the intersection of two \mathbf{A} -invariant subspaces
 - Let $\mathcal{V}_{co} = \mathcal{R}(\mathbf{Q}_c) \setminus \mathcal{V}_{c\bar{o}}$ be the controllable and observable subspace; let its dimension be n_{co}
 - * $\mathcal{V}_{c\bar{o}} \oplus \mathcal{V}_{co} = \mathcal{R}(\mathbf{Q}_c)$
 - Let $\mathcal{V}_{\bar{c}o} = \mathcal{N}(\mathbf{Q}_o) \setminus \mathcal{V}_{c\bar{o}}$ be the not controllable and not observable subspace; let its dimension be $n_{\bar{c}o}$
 - * $\mathcal{V}_{c\bar{o}} \oplus \mathcal{V}_{\bar{c}o} = \mathcal{N}(\mathbf{Q}_o)$
 - Let $\mathcal{V}_{\bar{c}\bar{o}} = \mathbb{R}^n \setminus (\mathcal{V}_{c\bar{o}} \oplus \mathcal{V}_{co} \oplus \mathcal{V}_{\bar{c}o})$ be the not controllable but observable subspace; let its dimension be $n_{\bar{c}\bar{o}}$
 - * Note that the direct sum of all subspaces is \mathbb{R}^n
- Let the matrix \mathbf{P} contain the basis for $\mathcal{V}_{c\bar{o}}, \mathcal{V}_{co}, \mathcal{V}_{\bar{c}o}, \mathcal{V}_{\bar{c}\bar{o}}$ in this exact order
 - The first $n_{c\bar{o}} + n_{co}$ columns form a basis for $\mathcal{R}(\mathbf{Q}_c)$
- Let $\mathbf{z} = \begin{bmatrix} \mathbf{z}^1 \\ \mathbf{z}^2 \\ \mathbf{z}^3 \\ \mathbf{z}^4 \end{bmatrix} = \mathbf{P}^{-1}\mathbf{x}$, where each of the 4 components corresponds to the 4 subspaces in the order above
- Applying the representation theorem, $\dot{\mathbf{z}} = \begin{bmatrix} \hat{\mathbf{A}}_{c\bar{o}} & * & * & * \\ 0 & \hat{\mathbf{A}}_{co} & 0 & * \\ 0 & 0 & \hat{\mathbf{A}}_{\bar{c}o} & * \\ 0 & 0 & 0 & \hat{\mathbf{A}}_{\bar{c}\bar{o}} \end{bmatrix} \begin{bmatrix} \mathbf{z}^1 \\ \mathbf{z}^2 \\ \mathbf{z}^3 \\ \mathbf{z}^4 \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{B}}_{c\bar{o}} \\ \hat{\mathbf{B}}_{co} \\ 0 \\ 0 \end{bmatrix} \mathbf{u}$

$$\mathbf{y} = [0 \ \hat{\mathbf{C}}_{co} \ 0 \ \hat{\mathbf{C}}_{\bar{c}o}] \mathbf{z} + \mathbf{D}\mathbf{u}$$
 - 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 $\hat{\mathbf{A}}_{11} \in \mathbb{R}^{(n_{co}+n_{co}) \times (n_{co}+n_{co})}$ 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. $\mathbf{x}(0) \in \mathcal{R}(\mathbf{Q}_c)$, then $\mathbf{z}^3(t) = \mathbf{z}^4(t) = 0$ for all t
 - $\mathbf{x}(0)$ is a linear combination of the basis vectors of \mathcal{V}_{co} and \mathcal{V}_{co} , since those two subspaces together make up $\mathcal{R}(\mathbf{Q}_c)$; it doesn't have any component in the basis vectors of the other two subspaces, since those do not intersect $\mathcal{R}(\mathbf{Q}_c)$
 - From $\mathbf{x}(0) = \mathbf{P}\mathbf{z}(0)$ we see that $\mathbf{z}^3(0) = \mathbf{z}^4(0) = 0$ as a result, and this gives $\dot{\mathbf{z}}^3 = \dot{\mathbf{z}}^4 = 0$ for all time
 - Then $\begin{bmatrix} \dot{\mathbf{z}}^1 \\ \dot{\mathbf{z}}^2 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}}_{co} & * \\ 0 & \hat{\mathbf{A}}_{co} \end{bmatrix} \begin{bmatrix} \mathbf{z}^1 \\ \mathbf{z}^2 \end{bmatrix} + \begin{bmatrix} \hat{\mathbf{B}}_{co} \\ \hat{\mathbf{B}}_{co} \end{bmatrix} \mathbf{u}$
 $\mathbf{y} = [0 \quad \hat{\mathbf{C}}_{co}] \begin{bmatrix} \mathbf{z}^1 \\ \mathbf{z}^2 \end{bmatrix} + \mathbf{D}\mathbf{u}$
 - * 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 $\dot{\mathbf{z}}^2 = \hat{\mathbf{A}}_{co}\mathbf{z}^2 + \hat{\mathbf{B}}_{co}\mathbf{u}$
 $\mathbf{y} = \hat{\mathbf{C}}_{co}\mathbf{z}^2 + \mathbf{D}\mathbf{u}$
- The observable subsystem is $\begin{bmatrix} \dot{\mathbf{z}}^2 \\ \dot{\mathbf{z}}^4 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}}_{co} & * \\ 0 & \hat{\mathbf{A}}_{co} \end{bmatrix} \begin{bmatrix} \mathbf{z}^2 \\ \mathbf{z}^4 \end{bmatrix} + \hat{\mathbf{B}}_{co}\mathbf{u}$
 $\mathbf{y} = [\hat{\mathbf{C}}_{co} \quad \hat{\mathbf{C}}_{co}] \begin{bmatrix} \mathbf{z}^2 \\ \mathbf{z}^4 \end{bmatrix} + \mathbf{D}\mathbf{u}$

Theorem

The system $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ has the same transfer function as the controllable and observable subsystem, $(\hat{\mathbf{A}}_{co}, \hat{\mathbf{B}}_{co}, \hat{\mathbf{C}}_{co}, \hat{\mathbf{D}})$, 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 $\mathbf{G} : \mathbb{C} \mapsto \mathbb{R}^{p \times m}$ be a transfer function matrix, and let $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ be any state space realization of the same system and $(\hat{\mathbf{A}}_{co}, \hat{\mathbf{B}}_{co}, \hat{\mathbf{C}}_{co}, \hat{\mathbf{D}})$ be the controllable and observable subsystem
 - If $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ is uncontrollable or unobservable, then the minimal realization has $k < n$ states, and so it will have k poles
 - The transfer function \mathbf{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 $G(s) = \frac{s+1}{(s+1)(s+3)}$, which has one pole-zero cancellation
 - Using the uncancelled transfer function, one choice of states is $\mathbf{x} = \begin{bmatrix} y \\ \dot{y} - u \end{bmatrix}$, we get the realization
 $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \mathbf{C} = [1 \quad 0]$
* $\mathbf{Q}_c = \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}$, which has rank 1, so the system is not controllable
 - Using $G(s) = \frac{1}{s+3}$, 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 $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$, we want to design an *observer*, which estimates (“observes”) the state $\mathbf{x}(t)$, given the known $\mathbf{y}(t)$ and $\mathbf{u}(t)$
- The observer predicts a state estimate $\hat{\mathbf{x}}$ as $\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{L}(\mathbf{y} - \hat{\mathbf{y}})$

$$\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}} + \mathbf{D}\mathbf{u}$$

- The observer is an LTI system itself which tries to simulate the system dynamics and corrects its estimate based on observations
- \mathbf{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* $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$, which we can show to have dynamics $\dot{\mathbf{e}} = (\mathbf{A} - \mathbf{LC})\mathbf{e}$
- Therefore we just need to choose \mathbf{L} such that this system is asymptotically stable, i.e. make the eigenvalues of $\mathbf{A} - \mathbf{LC}$ have negative real part

Definition

The *state estimation problem* is to find a matrix \mathbf{L} for a system $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ such that the eigenvalues of $\mathbf{A} - \mathbf{LC}$ 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 (\mathbf{C}, \mathbf{A}) is detectable if and only if $(\mathbf{A}^T, \mathbf{C}^T)$ is stabilizable.

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

Theorem

Duality Theorem: Controllability and detectability are duals, i.e.

- (\mathbf{C}, \mathbf{A}) is observable if and only if $(\mathbf{A}^T, \mathbf{C}^T)$ is controllable.
- (\mathbf{C}, \mathbf{A}) is detectable if and only if $(\mathbf{A}^T, \mathbf{C}^T)$ is stabilizable.

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

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

Theorem

PBH Test for Detectability: (\mathbf{C}, \mathbf{A}) is detectable if and only if for all eigenvalues λ of \mathbf{A} ,

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

- This is analogous to applying the PBH test for stabilizability to the dual system
- Proof:
 - (\mathbf{C}, \mathbf{A}) is detectable if and only if $(\mathbf{A}^T, \mathbf{C}^T)$ is stabilizable, which is true if and only if for all eigenvalues of \mathbf{A}^T , $\text{Re}(\lambda) \geq 0 \implies \text{rank}([\lambda \mathbf{I} - \mathbf{A}^T \quad \mathbf{C}^T]) = n$ (PBH stabilizability test)
 - Since transpose does not change rank or eigenvalues, this is the same as $\text{rank} \left(\begin{bmatrix} \lambda \mathbf{I} - \mathbf{A} \\ \mathbf{C} \end{bmatrix} \right) = n$
- Suppose (\mathbf{C}, \mathbf{A}) is not observable, then by applying the Kalman decomposition we get

$\dot{\mathbf{z}} = \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ 0 & \hat{\mathbf{A}}_{22} \end{bmatrix} \mathbf{z} + \begin{bmatrix} \hat{\mathbf{B}}_1 \\ \hat{\mathbf{B}}_2 \end{bmatrix} \mathbf{u}$ and $\mathbf{y} = [0 \ \hat{\mathbf{C}}_1] \mathbf{z} + \mathbf{D}\mathbf{u}$ where $(\hat{\mathbf{C}}_1, \hat{\mathbf{A}}_{22})$ is the observable subsystem

- Let $\hat{\mathbf{L}} = \begin{bmatrix} \hat{\mathbf{L}}_1 \\ \hat{\mathbf{L}}_2 \end{bmatrix}$ then $\hat{\mathbf{A}} - \hat{\mathbf{L}}\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} - \hat{\mathbf{L}}_1\hat{\mathbf{C}}_1 \\ 0 & \hat{\mathbf{A}}_{22} - \hat{\mathbf{L}}_2\hat{\mathbf{C}}_1 \end{bmatrix}$

- This shows that we can affect the eigenvalues of the $\hat{\mathbf{A}}_{22}$ subsystem but not the $\hat{\mathbf{A}}_{11}$ subsystem
- The eigenvalues of $\hat{\mathbf{A}}_{11}$ are the *unobservable eigenvalues*, and the eigenvalues of $\hat{\mathbf{A}}_{22}$ are the *observable eigenvalues*; together these make up all eigenvalues of \mathbf{A}
- Just like the controllability case, by the PBH test, an eigenvalue is unobservable if and only if $\text{rank} \left(\begin{bmatrix} \lambda\mathbf{I} - \mathbf{A} \\ \mathbf{C} \end{bmatrix} \right) < n$, and observable if and only if $\text{rank} \left(\begin{bmatrix} \lambda\mathbf{I} - \mathbf{A} \\ \mathbf{C} \end{bmatrix} \right) = n$

Output Feedback Stabilization

- Putting it all together, how do we stabilize a system if we only know the output $\mathbf{y}(t)$ and input $\mathbf{u}(t)$ but not the state?
- Given a stabilizable and detectable system, our goal is to design \mathbf{K} and \mathbf{L} such that the eigenvalues of $(\mathbf{A} + \mathbf{B}\mathbf{K})$ and $(\mathbf{A} - \mathbf{L}\mathbf{C})$ have real part less than zero, then we can implement the control law $\mathbf{u} = \mathbf{K}\hat{\mathbf{x}}$ and observer $\dot{\hat{\mathbf{x}}} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\mathbf{u} + \mathbf{L}(\mathbf{y} - \hat{\mathbf{y}})$

$$\hat{\mathbf{y}} = \mathbf{C}\hat{\mathbf{x}} + \mathbf{D}\mathbf{u}$$

- We will show that this indeed results in an asymptotically stable system

- Let $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$
- $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$

$$= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{K}\hat{\mathbf{x}} + \mathbf{B}\mathbf{K}\mathbf{x} - \mathbf{B}\mathbf{K}\mathbf{x}$$

$$= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{K}\mathbf{x} - \mathbf{B}\mathbf{K}(\mathbf{x} - \hat{\mathbf{x}})$$

$$= (\mathbf{A} + \mathbf{B}\mathbf{K})\mathbf{x} - \mathbf{B}\mathbf{K}\mathbf{e}$$

- Also, $\dot{\mathbf{e}} = (\mathbf{A} - \mathbf{L}\mathbf{C})\mathbf{e}$ (shown previously)

- Therefore $\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} + \mathbf{B}\mathbf{K} & -\mathbf{B}\mathbf{K} \\ 0 & \mathbf{A} - \mathbf{L}\mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix}$, so if we have the eigenvalues of $(\mathbf{A} + \mathbf{B}\mathbf{K})$ and $(\mathbf{A} - \mathbf{L}\mathbf{C})$ both negative, then the overall system is stable

Theorem

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