

Chapter 3

Controllability and Observability

In this chapter, we study the controllability and observability concepts. Controllability is concerned with whether one can design control input to steer the state to arbitrarily values. Observability is concerned with whether without knowing the initial state, one can determine the state of a system given the input and the output.

3.1 Some linear mapping concepts

The study of controllability and observability for linear systems essentially boils down to studying two linear maps: the reachability map L_r which maps, for zero initial state, the control input $u(\cdot)$ to the final state $x(t_1)$; and the observability map, L_o which maps, for zero input, from the initial state $x(t_0)$ to the output trajectory $y(\cdot)$. The *range* and the *null* spaces of these maps respectively are critical for their studies.

Let $L : \mathcal{X}(= \mathbb{R}^p) \rightarrow \mathcal{Y}(= \mathbb{R}^q)$ be a linear map: i.e.

$$L(\alpha x_1 + \beta x_2) = \alpha(Lx_1) + \beta \cdot (Lx_2)$$

$\mathcal{R}(L) \subset \mathcal{Y}$ denotes the range of L , given by:

$$\mathcal{R}(L) = \{y \in \mathcal{Y} | y = Lx \text{ for some } x \in \mathcal{X}\}$$

$\mathcal{N}(L) \subset \mathcal{X}$ denotes the null space of L , given by:

$$\mathcal{N}(L) = \{x \in \mathcal{X} | Lx = \mathbf{0}\}$$

Note: Both $\mathcal{R}(L)$ and $\mathcal{N}(L)$ are linear subspaces, i.e. for scalar α, β ,

$$\begin{aligned} y_1, y_2 \in \mathcal{R}(L) &\Rightarrow (\alpha y_1 + \beta y_2) \in \mathcal{R}(L) \\ x_1, x_2 \in \mathcal{N}(L) &\Rightarrow (\alpha x_1 + \beta x_2) \in \mathcal{N}(L) \end{aligned}$$

- L is called onto or surjective if $\mathcal{R}(L) = \mathcal{Y}$ (everything is within range).
- L is called into or one-to-one (cf. many-to-one) if $\mathcal{N}(L) = \{0\}$. Thus,

$$Lx_1 = Lx_2 \Leftrightarrow x_1 = x_2$$

3.1.1 The Finite Rank Theorem

Proposition 3.1.1 : Let $L \in \mathbb{R}^{n \times m}$, so that $L : u \in \mathbb{R}^m \mapsto y \in \mathbb{R}^n$.

1.

$$\mathcal{R}(L) = \mathcal{R}(LL^T)$$

2.

$$\mathcal{N}(L) = \mathcal{N}(L^T L)$$

3. Any point $u \in \mathbb{R}^n$ can be written as

$$u = L^T y_1 + u_n \in \mathbb{R}^n; \text{ where } y_1 \in \mathbb{R}^n, \quad Lu_n = 0.$$

In other words, u can always be written as a sum of $u_1 \in \mathcal{R}(L^T)$ and $u_n \in \mathcal{N}(L)$.

4. Similarly, any point $u \in \mathbb{R}^n$ can be written as

$$y = Lu_1 + y_n \in \mathbb{R}^n; \text{ where } u_1 \in \mathbb{R}^m, \quad L^T y_n = 0.$$

In other words, u can always be written as a sum of $y_1 \in \mathcal{R}(L)$ and $y_n \in \mathcal{N}(L^T)$.

Note:

- If $L \in \mathbb{R}^{n \times m}$ and $m > n$ (short and fat), then $LL^T \in \mathbb{R}^{n \times n}$ is both short and thin.
- To check whether the system is controllable we need to check if LL^T is full rank - e.g. check its determinant.
- If $m < n$ (tall and thin), then $L^T L \in \mathbb{R}^{n \times n}$ is also both short and thin.
- Checking $\mathcal{N}(L^T L)$ will be useful when we talk about observerability.

3.2 Controllability

Consider a state determined dynamical system with a transition map $s(t_1, t_0, x_0, u(\cdot))$ such that

$$x(t_1) = s(t_1, t_0, x(t_0), u(\cdot)).$$

This can be continuous time or discrete time.

Definition 3.2.1 (Controllability) : The dynamical system is controllable on $[t_0, t_1]$ if \forall initial and final states x_0, x_1 , $\exists u(\cdot)$ so that $s(t_1, t_0, x_0, u) = x_1$.

It is said to controllable at t_0 if $\forall x_0, x_1$, $\exists t_1 \geq t_0$ and $u(\cdot) \in \mathcal{U}$ so that $s(t_1, t_0, x_0, u) = x_1$

A system that is **not** controllable probably means the followings:

- If state space system is realized from input-output data, then the realization is redundant. i.e. some states have no relationship to either the input or the output.
- If states are meaningful (physical) variables that need to be controlled, then the design of the actuators are deficient.
- The effect of control is limited. There is also a possibility of instability

Remarks:

- The followings are equivalent:
 1. A system is controllable at $t = t_0$.
 2. The system can be transferred from any state at $t = t_0$ to any other state in finite time.
 3. For each $x_0 \in \Sigma$,

$$s(\cdot, t_0, x_0, \cdot) : [t_0, \infty) \times u(\cdot) \mapsto s(t_1, t_0, x_0, u(\cdot))$$

is surjective (onto).

- Controllability does **not** say that $x(t)$ remains at x_1 . e.g. for

$$\dot{x} = Ax + Bu$$

to make $x(t) = x_1 \forall t \geq t_0$, it is necessary that $Ax_1 \in \text{Range}(B)$. This is generally not true.

Example - An Uncontrollable System

- An linear actuator pushing 2 masses on each end of the actuator in space.

$$m_1 \ddot{x}_1 = u; \quad m_2 \ddot{x}_2 = -u$$

- State: $X = [x_1, \dot{x}_1, x_2, \dot{x}_2]^T$.
- Action and reaction are equal and opposite, and act on different bodies.
- Momentum is conserved:

$$m_1 \dot{x}_1 + m_2 \dot{x}_2 = \text{constant} = m_1 \dot{x}_1(t_0) + m_2 \dot{x}_2(t_0)$$

- E.g. if both masses are initially at rest: it is not possible to choose $u(\cdot)$ to make $\dot{x}_1 = 0$, $\dot{x}_2 = 1$.

3.2.1 Effects of feedback on controllability

Consider a system

$$\dot{x} = f(t, x(t), u(t)).$$

Suppose that the system is under **memoryless (static) state feedback**:

$$u(t) = v(t) - g(t, x(t))$$

with $v(t)$ being the new input.

- A system is controllable if and only if the system under memoryless feedback with $v(t)$ as the input is controllable.

Proof: Suppose that the original system is controllable. Given a pair of initial and final states, x_0 and x_1 , suppose that the control $u(\cdot)$ transfers the state from x_0 at t_0 to x_1 at t_1 . For the new system, which has also $x(t)$ as the state variables, we can use the control $v(t) = u(t) + g(t, x(t))$ to transfer the state from x_0 at t_0 to x_1 at t_1 . Therefore, the new system is also controllable.

Suppose that the feedback system is controllable, and $v(\cdot)$ is the control that can achieve the desired state transfer. Then for the original system, we can use the control $u(t) = v(t) - g(t, x(t))$ to achieve the same state transfer. Thus, the original system is also controllable. \diamond

Consider now the system under **dynamic state feedback**:

$$\begin{aligned}\dot{z}(t) &= \alpha(t, x(t), z(t)) \\ u(t) &= v(t) - g(t, x(t), z(t))\end{aligned}$$

where $v(t)$ is the new input and $z(t)$ is the state of the controller.

- If a system under **dynamic feedback** is controllable then the original system is.
- The converse is not true, i.e. the original system is controllable does not necessarily imply that the system under dynamic state feedback is.

Proof: The same proof for the memoryless state feedback works for the first statement. For the second statement, we give an example of the controller:

$$\dot{z} = z; \quad u(t) = v(t) - g(t, x(t), z(t)).$$

The state of the new system is (x, z) but no control can affect $z(t)$. Thus, the new system is not controllable. \diamond

These results show that feedback control cannot make an uncontrollable system become controllable, but can even make a controllable system, uncontrollable. To make a system controllable, one must redesign the actuators.

3.3 Controllability of Linear Systems

Continuous time system:

$$\begin{aligned}\dot{x}(t) &= A(t)x(t) + B(t)u(t) & x &\in \mathfrak{R}^n \\ x(t_1) &= \Phi(t_1, t_0)x_0 + \int_{t_0}^{t_1} \Phi(t_1, t)B(t)u(t)dt.\end{aligned}$$

The **reachability map** on $[t_0, t_1]$ is defined to be:

$$L_{r,[t_0,t_1]}(u(\cdot)) = \int_{t_0}^{t_1} \Phi(t_1, t)B(t)u(t)dt$$

Thus, it is *controllable on* $[t_0, t_1]$ if and only if $L_{r,[t_0,t_1]}(u(\cdot))$ is surjective (onto).

Notice that $L_{r,[t_0,t_1]}$ determines the set of states that can be reached from the origin at $t = t_1$.

The study of the range space of the linear map:

$$L_{r,[t_0,t_1]} : \{u(\cdot)\} \rightarrow \mathfrak{R}^n$$

is central to the study of controllability.

Proposition 3.3.1 *If a continuous time, possibly time varying, linear system is controllable on $[t_0, t_1]$ then it is controllable on $[t_0, t_2]$ where $t_2 \geq t_1$.*

Proof: To transfer from $x(t_0) = x_0$ to $x(t_2) = x_2$, define

$$x_1 = \Phi(t_1, t_2)x_2$$

Suppose that $u_{[t_0, t_1]}^*$ transfers $x(t_0) = x_0$ to $x(t_1) = x_1$. Choose

$$u(t) = \begin{cases} u^*(t) & t \in [t_0, t_1] \\ 0 & t \in (t_1, t_2] \end{cases}$$

◇

◇

This result does **not** hold for $t_0 < t_2 < t_1$. e.g. The system

$$\dot{x}(t) = B(t)u(t) \quad B(t) = \begin{cases} 0 & t \in [t_0, t_2] \\ \mathbf{I}_n & t \in (t_2, t_1] \end{cases}$$

is controllable on $[t_0, t_1]$ but $x(t_2) = x(t_0)$ for any control $u(\cdot)$.

3.3.1 Discrete time system

$$x(k+1) = A(k)x(k) + B(k)u(k), \quad x \in \mathfrak{R}^n, u \in \mathfrak{R}^m$$

$$x(k_1) = \Phi(k_1, k_0)x(k_0) + \sum_{k=k_0}^{k_1-1} \Phi(k_1, k+1)B(k)u(k)$$

The transition function of the discrete time system is given by:

$$x(k_1) = s(k_0, k_1, x(k_0), u(\cdot)) = \Phi(k_1, k_0)x(k_0) + \sum_{k=k_0}^{k_1-1} \Phi(k_1, k+1)B(k)u(k) \quad (3.1)$$

$$\Phi(k_f, k_i) = \prod_{k=k_i}^{k_f-1} A(k) \quad (3.2)$$

The reachability map of the discrete time system can be written as:

$$L_{r, [k_0, k_1]}(u(k_0), \dots, u(k_1-1)) = \sum_{k=k_0}^{k_1-1} \Phi(k_1, k+1)B(k)u(k) = L_r(k_0, k_1)\mathbf{U}$$

where $L_r(k_0, k_1) \in \mathfrak{R}^{n \times (k_1-k_0)m}$, $\mathbf{U} = \begin{pmatrix} u(k_0) \\ \vdots \\ u(k_1-1) \end{pmatrix}$.

Thus, the system is controllable on $[k_0, k_1]$ if and only if $L_{r, [k_0, k_1]}$ is surjective, which is true if and only if $L_r(k_0, k_1)$ has rank n .

Proposition 3.3.2 *Suppose that $A(k)$ is nonsingular for each $k_1 \leq k < k_2$, then the discrete time linear system is controllable on $[k_0, k_1]$ implies that it is controllable on $[k_0, k_2]$ where $k_2 \geq k_1$.*

Proof: : The proof is similar to the continuous case. However for a discrete time system, we need $A(k)$ nonsingular for $k_1 \leq k < k_2$ to ensure that $\Phi(k_2, k_1)$ is invertible. Q.E.D. ◇

Example

Consider a unit point mass under control of a force:

$$\frac{d}{dt} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u$$

This is the same as $\ddot{x} = u$.

Suppose that $x(0) = \dot{x}(0) = 0$, we would like to translate the mass to $x(T) = 1$, $\dot{x}(T) = 0$.

The reachability map $L : u(\cdot) \rightarrow x(T)$ is

$$x(T) = \int_0^T \Phi(T, t) B(t) u(t) dt$$

Let us try to solve this problem using piecewise constant control:

$$u(t) = \begin{cases} u_0 & 0 \leq t < T/10 \\ u_1 & T/10 \leq t < 2T/10 \\ \vdots & \\ u_{10} & 9T/10 \leq t \leq T \end{cases}$$

and find $U = [u_1, u_2, \dots, u_{10}]^T$.

The reachability map becomes:

$$x(T) = \underbrace{(L_1 \quad L_2 \quad \dots \quad L_{10})}_{L_r} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{10} \end{pmatrix}$$

where

$$\begin{aligned} L_1 &= \left[\int_0^{T/10} \Phi(T, t) B(t) dt \right] = \frac{T}{10} \begin{pmatrix} T \\ 1 \end{pmatrix}, \\ L_2 &= \left[\int_{T/10}^{2T/10} \Phi(T, t) B(t) dt \right] = \frac{T}{10} \begin{pmatrix} 0.9T \\ 1 \end{pmatrix}, \\ &\vdots \\ L_{10} &= \frac{T}{10} \begin{pmatrix} 0.1T \\ 1 \end{pmatrix} \end{aligned}$$

Whether the matrix denoted by $L_r \in \mathfrak{R}^{2 \times 10}$ is surjective (onto) (i.e. $\mathcal{R}(L_r) = \mathfrak{R}^2$) determines whether we can achieve our task using piecewise constant control.

In our case, let $T = 10$ for convenience,

$$L_r = \begin{pmatrix} 10 & 9 & \dots & 1 \\ 1 & 1 & \dots & 1 \end{pmatrix}$$

and

$$L_r L_r^T = \begin{pmatrix} 385 & 55 \\ 55 & 10 \end{pmatrix}$$

which is non-singular (full rank). Thus, from the finite rank theorem $\mathcal{R}(L_r) = \mathcal{R}(L_r L_r^T)$, the particle transfer problem is solvable using piecewise constant control.

If we make the number of pieces in the piecewise constant control larger and larger, then the multiplication of L_r by L_r^T becomes an integral:

$$L_r L_r^T \rightarrow \int_{t_0}^{t_1} \Phi(t_1, t) B(t) B(t)^T \Phi(t_1, t)^T dt.$$

Therefore for the *continuous time* linear system is controllable over the interval $[t_0, t_1]$ if and only if the matrix,

$$\int_{t_0}^{t_1} \Phi(t_1, t) B(t) B(t)^T \Phi(t_1, t)^T dt \quad (3.3)$$

is full rank.

3.4 Reachability Grammian

The matrix $W_r = L_r L_r^T \in \mathfrak{R}^{n \times n}$ is called the *Reachability Grammian* for the interval that L_r is defined. For the continuous time system

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

the Reachability Grammian on the time interval $[t_0, t_1]$ is defined to be:

$$W_{r,[t_0,t_1]} = \int_{t_0}^{t_1} \Phi(t_1, t) B(t) B(t)^T \Phi(t_1, t)^T dt$$

For time invariant systems,

$$\begin{aligned} W_r([t_1, t_0]) &= L_r L_r^* = \int_{t_0}^{t_1} e^{A(t_1-t)} B B^T e^{A^T(t_1-t)} dt \\ &= \int_{t_1-t_0}^0 e^{A\tau} B B^T e^{A^T \tau} d\tau. \end{aligned}$$

For the discrete time system,

$$x(k+1) = A(k)x(k) + B(k)u(k),$$

the Reachability Grammian on the interval $[k_0, k_1]$ is

$$\sum_{k=k_0}^{k_1-1} \Phi(k_1, k+1) B(k) B(k)^T \Phi(k_1, k+1)^T.$$

From the finite rank theorem, we know that the controllability of a system on $[t_0, t_1]$ (or on $[k_0, k_1]$) is equivalent to Reachability Grammian on the same time interval being full rank.

3.4.1 Reduction theorem: Controllability in terms of Controllability to Zero

Thus far, we have been deciding controllability by considering the reachability map L_r which concerns what states we are able to steer to from $x(t_0) = 0$. The **reduction theorem** allows us to formulate the question in terms of whether a state can be steered to $x(t_f) = 0$ in finite time.

Theorem 3.4.1 *The followings are equivalent for a linear time varying differential system:*

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

1. *It is controllable on $[t_0, t_1]$.*
2. *It is controllable to 0 on $[t_0, t_1]$, i.e. $\forall x(t_0) = x_0$, there exists $u(\tau), \tau \in [t_0, t_1)$ such that final state is $x(t_1) = 0$.*
3. *It is reachable from $x(t_0) = 0$, i.e. $\forall x(t_1) = x_f$, there exists $u(\tau), \tau \in [t_0, t_1)$ such that for $x(t_0) = 0$, the final state is $x(t_1) = x_f$.*

Proof:

- Clearly, (1) \Rightarrow (2) and (3).
- To prove (2) \Rightarrow (1) and (3) \Rightarrow (1), use

$$x(t_1) = \Phi(t_1, t_0)x(t_0) + \int_{t_0}^{t_1} \Phi(t_1, \tau)B(\tau)u(\tau)d\tau.$$

and $\Phi(t_1, t_0)$ is invertible, to construct the appropriate $x(t_0)$ and $x(t_f)$.

Controllability map: $L_{c,[t_0,t_1]}$

The controllability (to zero) map on $[t_0, t_1]$ is the map between $u(\cdot)$ to the *initial state* x_0 such that $x(t_1) = 0$.

$$L_{c,[t_0,t_1]}(u(\cdot)) = - \int_{t_0}^{t_1} \Phi(t_0, t)B(t)u(t)dt$$

Proof:

$$\begin{aligned} 0 &= x(t_1) = \Phi(t_1, t_0)x_0 + \int_{t_0}^{t_1} \Phi(t_1, \tau)B(\tau)u(\tau)d\tau \\ x_0 &= -\Phi(t_1, t_0)^{-1} \int_{t_0}^{t_1} \Phi(t_1, \tau)B(\tau)u(\tau)d\tau \\ x_0 &= - \int_{t_0}^{t_1} \Phi(t_0, t_1)\Phi(t_1, \tau)B(\tau)u(\tau)d\tau \\ x_0 &= - \int_{t_0}^{t_1} \Phi(t_0, \tau)B(\tau)u(\tau)d\tau \end{aligned}$$

Notice that

$$\mathcal{R}(L_{c,[t_0,t_1]}) = \Phi(t_1, t_0)\mathcal{R}(L_{r,[t_0,t_1]})$$

These two ranges are typically not the same for time varying systems. However, one is full rank if and only if the other is, since $\Phi(t_1, t_0)$ is invertible.

For time invariant continuous time systems, $\mathcal{R}(L_{c,[t_0,t_1]}) = \mathcal{R}(L_{r,[t_0,t_1]})$.

The above two statements are not true for time invariant discrete time systems. Can you find counter-examples?

Because of the reduction theorem, the linear continuous time system is controllable on the interval $[t_0, t_f]$ if $\mathcal{R}(L_{c,[t_0,t_1]}) = \mathfrak{R}^n$. From the finite rank theorem, $\mathcal{R}(L_{c,[t_0,t_1]}) = \mathcal{R}(L_{c,[t_0,t_1]}L_{c,[t_0,t_1]}^*) = \mathfrak{R}^n$. The controllability-to-zero grammian on the time interval $[t_0, t_1]$ is defined to be:

$$W_c = L_c L_c^* = \int_{t_0}^{t_1} \Phi(t_0, t) B(t) B(t)^T \Phi(t_0, t)^T dt$$

For time invariant system,

$$\begin{aligned} W_c = L_c L_c^* &= \int_{t_0}^{t_1} e^{A(t_0-t)} B B^T e^{A^T(t_0-t)} dt \\ &= - \int_{t_1-t_0}^0 e^{-A\tau} B B^T e^{-A^T\tau} d\tau. \end{aligned}$$

Since controllability of a system on $[t_0, t_1]$ is equivalent to controllability to 0 which is determined by the controllability map being surjective, it is also equivalent to controllability grammian on the same time interval $[t_0, t_1]$ being full rank.

3.5 Minimum norm control

3.5.1 Formulation

The Controllability Grammian is related to the cost of control.

Given $x(t_0) = x_0$, find a control function or sequence $u(\cdot)$ so that $x(t_1) = x_1$. Let $x_d = x_1 - \Phi(t_1, t_0)x_0$. Then we must have

$$x_d = L_{r,[t_0,t_1]}(u)$$

where $L_{r,[t_0,t_1]} : \{u(\cdot)\} \rightarrow \mathfrak{R}$ is the reachability map which, for continuous time system is:

$$L_{r,[t_0,t_1]}(u) = \int_{t_0}^{t_1} \Phi(t_1, t) B(t) u(t) dt.$$

and for discrete time system:

$$L_{r,[k_0,k_1]}(u(\cdot)) = \sum_{k=k_0}^{k_1-1} \Phi(k_1, k+1) B(k) u(k) = \mathbf{C}(k_0, k_1) \mathbf{U}$$

where $\mathbf{C} \in \mathfrak{R}^{n \times (k_1-k_0)m}$, $\mathbf{U} = \begin{pmatrix} u(k_0) \\ \vdots \\ u(k_1-1) \end{pmatrix}$.

Let us focus on the discrete time system.

Since $x_d \in \text{Range}(L_{r,[k_0,k_1]})$ for solutions to exist, we assume $\text{Range}(L_{r,[k_0,k_1]}) = \mathfrak{R}^n$. This implies that $W_{r,[k_0,k_1]}$ is invertible.

Generally there are multiple solutions. To resolve the non-uniqueness, we find the solution so that

$$J(U) = \frac{1}{2} \sum_{k=k_0}^{k_1-1} u(k)^T u(k) = U^T U$$

is minimized. Here $U = [u(k_0); u(k_0+1); \dots; u(k_1-1)]$.

3.5.2 Least norm control solution

This is a constrained optimization problem (with $J(U)$ as the cost to be minimized, and $L_r U - x_d = 0$ as the constraint). It can be solved using the Lagrange Multiplier method of converting a constrained optimization problem into an unconstrained optimization.

Define an augmented cost function, with the Lagrange multipliers $\lambda \in \mathfrak{R}^n$

$$J'(U, \lambda) = J(U) + \lambda^T (L_r U - x_d)$$

The optimal given by (U^*, λ^*) must satisfy:

$$\frac{\partial J'}{\partial U}(U^*, \lambda^*) = 0; \quad \frac{\partial J'}{\partial \lambda}(U^*, \lambda^*) = 0$$

The second condition is just the constraint: $L_r U = x_d$.

This means that

$$L_r^T \lambda^* + U^* = 0; \quad L_r U^* = x_d$$

Solving, we get:

$$\begin{aligned} \lambda^* &= -(L_r L_r^T)^{-1} x_d = -W_r^{-1} x_d \\ U^* &= -L_r^T \lambda^* = L_r^T W_r^{-1} x_d. \end{aligned}$$

The **optimal cost** of control is:

$$J(U^*) = x_d^T W_r^{-1} L_r L_r^T W_r^{-1} x_d = x_d^T W_r^{-1} x_d$$

Thus, the inverse of the Reachability Grammian tells us how difficult it is to perform a state transfer from $x = 0$ to x_d . In particular, if W_r is not invertible, for some x_d , the cost is infinite.

3.5.3 Geometric interpretation of least norm solution

Geometrically, we can think of the cost as $J = U^T U$, i.e. the inner product of U with itself. In notation of inner product,

$$J = \langle U, U \rangle_R$$

The advantage of the notation is that we can change the definition of inner product, e.g. $\langle U, V \rangle_R = U^T R V$ where R is a positive definite matrix. The usual inner (dot) product has $R = I$.

We say that U and V are normal to each other if $\langle U, V \rangle_R = 0$.

Any solution that satisfies the constraint must be of the form

$$(U - U^p) \in \text{Null}(L_r)$$

where $L_r U^p = x_f$ is any *particular* solution.

Claim: Let U^* be the optimal solution, and U is any solution that satisfies the constraint. Then, $(U - U^*) \perp U^*$, i.e.

$$\langle (U - U^*), U^* \rangle_R = 0$$

which is the normal equation for the least norm solution problem.

Proof: Direct substitution.

$$(U - U^*)^T R_U R_U^{-1} L_{pc}^T (L_r R_U^{-1} L_r^T)^{-1} x_f = (L_r (U - U^*))^T (L_r R_U^{-1} L_r^T)^{-1} x_f = 0.$$

◇

3.5.4 Open-loop least norm control

Applying the above result to linear continuous time system using standard norms, i.e.

$$J = \int_{t_0}^{t_1} u^T(\tau)u(\tau)d\tau.$$

$$\begin{aligned} (L_{r,[t_0,t_1]})^*(t) &= B^T(t)\Phi(t,t_1)^T \\ W_{r,[t_0,t_1]} &= L_r L_r^* = \left[\int_{t_0}^{t_1} \Phi(t_1,\tau)B(\tau)B^T(\tau)\Phi(t_1,\tau)^T d\tau \right] \\ u_{opt}(t) &= B(t)^T \Phi(t,t_1)^T W_{r,[t_0,t_1]}^{-1} (x_1 - \Phi(t_1,t_0)x_0) \end{aligned} \quad (3.4)$$

Eq, (3.4) solves $u_{opt}(\cdot)$ in a batch form given x_0 , i.e. it is openloop.

3.5.5 Recursive formulation

- Openloop control does not make use of feedback. It is not robust to disturbances or model uncertainty. Suppose now that $x(t)$ is measured. Conceptually, we can solve (3.4) for $u_{opt}(t)$ and let $t_0 = t$.
- Computing $W_{r,[t,t_1]}$ is expensive. Try recursion:

$$\begin{aligned} u_{opt}(t) &= B(t)^T \Phi(t,t_1)^T W_{r,[t,t_1]}^{-1} (x_1 - \Phi(t_1,t)x(t)) \\ &= \left(B(t)^T \Phi(t,t_1)^T W_{r,[t,t_1]}^{-1} \right) x_1 - \left(B(t)^T \Phi(t,t_1)^T W_{r,[t,t_1]}^{-1} \Phi(t_1,t) \right) x(t) \end{aligned}$$

For any invertible matrix $Q(t) \in \mathfrak{R}^{n \times n}$, $Q(t)Q^{-1}(t) = I$, therefore

$$\frac{d}{dt}Q^{-1}(t) = -Q^{-1}(t)\dot{Q}(t)Q^{-1}(t)$$

Since

$$\begin{aligned} W_{r,[t,t_1]} &= \int_t^{t_1} \Phi(t_1,\tau)B(\tau)B(\tau)^*\Phi(t_1,\tau)^* d\tau \\ \frac{d}{dt}W_{r,[t,t_1]} &= -\Phi(t_1,t)B(t)B(t)^*\Phi(t_1,t)^* \end{aligned}$$

- Thus we can solve for $W_{r,[t,t_1]}^{-1}$ dynamically by integrating on-line

$$\frac{d}{dt}W_{r,[t,t_1]}^{-1} = -W_{r,[t,t_1]}^{-1} \Phi(t_1,t)B(t)B(t)^*\Phi(t_1,t)^* W_{r,[t,t_1]}^{-1}$$

- **Issue:** Does not work well when $t \rightarrow t_1$ because $W_{r,[t,t_1]}$ can be close to singular.

3.5.6 Optimal cost

Optimal cost as a function of initial time t_0

Since

$$W_{r,[t_0,t_f]} = \int_{t_0}^{t_f} \Phi(t_f, \tau) B(\tau) B(\tau)^* \Phi(t_f, \tau)^* d\tau$$

and the integrand is a positive semi-definite term, for each $v \in \mathfrak{R}^n$, $v^T W_{r,[t_0,t_f]}^{-1} v$ decreases as t_0 decreases, i.e. given more time, the cost of control decreases.

To see this, fix the final time t_f , and consider the minimum cost function,

$$E_{min}(x_f, t_0) = x_f^T W_{r,[t_0,t_f]}^{-1} x_f$$

and consider

$$\frac{d}{dt_0} E_{min}(x_f, t_0) = v^T \frac{d}{dt_0} W_{r,[t_0,t_f]}^{-1} x_f$$

Since

$$\begin{aligned} \frac{d}{dt_0} W_{r,[t_0,t_f]}^{-1} &= -W_{r,[t_0,t_f]}^{-1} \frac{d}{dt_0} W_{r,[t_0,t_f]} W_{r,[t_0,t_f]}^{-1} \\ &= - \left\{ W_{r,[t_0,t_f]}^{-1} \Phi(t_f, t_0) B(t_0) B(t_0)^* \Phi(t_f, t_0)^* W_{r,[t_0,t_f]}^{-1} \right\} \end{aligned}$$

is negative semi-definite, $\frac{d}{dt_0} E_{min}(v, t_0) \leq 0$. Hence, the optimal control increases as the initial time decreases, increasing the allowable time interval.

In other words, the set of states reachable with a cost less than a given value grows as t_0 decreases.

If the system is unstable, in that, $\Phi(t_f, t_0)$ becomes unbounded as $t_0 \rightarrow -\infty$, then one can see that $W_{r,[t_0,t_f]}$ also becomes unbounded. This means that for some x_f , $E_{min}(x_f, t_0) \rightarrow 0$ and $t_f \rightarrow \infty$. This means that some directions do not require any cost of control. Which direction do these correspond to?

For time invariant system, decreasing t_0 is equivalent to increasing t_f . Thus for time-invariant system, increasing the time interval increases W_r , thus decreasing cost.

Optimal cost as a function final state x_f

Since $W_{r,[t_0,t_f]}$ is symmetric and positive semi-definite, has n orthogonal eigenvectors, $V = [v_1, \dots, v_n]$, and associated eigenvalues $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, such that $VV^T = I$. Then

$$W_{r,[t_0,t_f]} = V\Lambda V^T \Rightarrow W_{r,[t_0,t_f]}^{-1} = V \text{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1}) V^T$$

This shows that most expensive direction to transfer to is the eigenvector associated with the minimum λ_i , and the least expensive direction is the eigenvector associated with the maximum λ_i

3.6 Linear Time Invariant (LTI) Continuous Time Systems

Consider first the LTI continuous time system:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned} \tag{3.5}$$

Note: Because of time invariance, the system is controllable (observable) at t_0 means that it is controllable (observable) at any time.

We will develop tests for controllability and observability by using directly the matrices A and B .

The three tests for controllability of system (3.5) is given by the following theorem.

Theorem For the continuous time system (3.5), the followings are equivalent:

1. The system is controllable over the interval $[0, T]$, for some $T > 0$.
2. The system is controllable over any interval $[t_0, t_1]$ with $t_1 > t_0$.
3. The reachability grammian,

$$\begin{aligned} W_{r,T} &:= \int_0^T \Phi(T,t)B(t)B(t)^T\Phi(T,t)^T dt \\ &= \int_0^T e^{At}BB^*e^{A^*t} dt \end{aligned}$$

is full rank n . (This test works for time varying systems also)

4. The controllability matrix

$$\mathcal{C} := (B \quad AB \quad A^2B \quad \dots \quad A^{n-1}B)$$

has rank n . Notice that the controllability matrix has dimension $n \times nm$ where m is the dimension of the control vector u .

5. (**Belovich-Popov-Hautus test**) For each $s \in \mathbb{C}$, the matrix,

$$(sI - A \quad B)$$

has rank n .

Note that rank of $(sI - A, B)$ is less than n only if s is an eigenvalue of A .

We need the Cayley-Hamilton Theorem to prove this result:

Theorem 3.6.1 (Cayley Hamilton) Let $A \in \mathfrak{R}^{n \times n}$. Consider the polynomial

$$\Delta(s) = \det(sI - A)$$

Then $\Delta(A) = 0$.

Notice that if λ_i is an eigenvalue of A , then $\Delta(\lambda_i) = 0$.

Proof: This is true for any A , but is easy to see using the eigen decomposition of A when A is semi-simple.

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

where Λ is diagonal with eigenvalues on its diagonal. Since $\psi(\lambda_i) = 0$ by definition,

$$\Delta(A) = T\psi(\Lambda)T^{-1} = 0.$$

Proof: Controllability test

(1) \Leftrightarrow (3): We have already seen before that controllability over the interval $[0, T]$ means that the range space of the reachability map,

$$L_r(u) = \int_0^T e^{A(T-\tau)} B u(\tau) d\tau$$

must be the whole state space, \mathfrak{R}^n . Notice that

$$W_{r,T} := \int_0^T e^{At} B B^* e^{A^*t} dt = L_r L_r^*$$

where L_r^* is the adjoint (think transpose) of L_r .

From the finite rank linear map theorem, $\mathcal{R}(L_r) = \mathcal{R}(L_r L_r^*)$ but, $L_r L_r^*$ is nothing but $W_{r,T}$.

(3) \Rightarrow (4): We will show this by showing “not (4) \Rightarrow not (3)”. Suppose that (4) is not true. Then, there exists a $1 \times n$ vector v^T so that,

$$v^T B = v^T A B = v^T A^2 B \dots = v^T A^{n-1} B = 0.$$

Consider now,

$$v^T W_{r,T} v = \int_0^T \|v^T e^{At} B\|_2^2 dt.$$

Since

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

and by the Cayley Hamilton Theorem, A^k is a linear combination of I, A, \dots, A^{n-1} , therefore,

$$v^T e^{At} B = 0$$

for all t . Hence, $v^T W_{r,T} v = 0$ or $W_{r,T}$ is not full rank.

(3) \Rightarrow (2): We will show this by showing “not (2) \Rightarrow not (3)”. If (2) is not true, then there exists $1 \times n$ vector v^T so that

$$v^T W_{r,T} v = \int_0^T \|v^T e^{At} B\|_2^2 dt = 0.$$

Because e^{At} is continuous in t , this implies that $v^T e^{At} B = 0$ for all $t \in [0, T]$.

Hence, the all time derivatives of $v^T e^{At} B = 0$ for all $t \in [0, T]$. In particular, at $t = 0$,

$$\begin{aligned} v^T e^{At} B |_{t=0} &= v^T B = 0 \\ v^T \frac{d}{dt} e^{At} B |_{t=0} &= v^T A B = 0 \\ v^T \frac{d^2}{dt^2} e^{At} B |_{t=0} &= v^T A^2 B = 0 \\ &\vdots \\ v^T \frac{d^{n-1}}{dt^{n-1}} e^{At} B |_{t=0} &= v^T A^{n-1} B = 0 \end{aligned}$$

Hence,

$$v^T (B \ AB \ A^2B \ \cdots \ A^{n-1}B) = (0 \ 0 \ \cdots 0)$$

i.e. the controllability matrix of (A, B) is not full rank.

(4) \Rightarrow (5): We will show “not (5) implies not (4)”. Suppose (5) is not true so that there exists a $1 \times n$ vector v , and $\lambda \in \mathcal{C}$,

$$v^T (\lambda I - A) = 0, \quad v^T B = 0.$$

for some λ . Then $v^T A = \lambda v^T$, $v^T A^2 = \lambda^2 v^T$ etc. Because $v^T B = 0$ by assumption, we have $v^T AB = \lambda v^T B = 0$, $v^T A^2 B = \lambda^2 v^T B = 0$ etc. Hence,

$$v^T B = v^T AB = v^T A^2 B = \cdots v^T A^{n-1} B = 0.$$

Therefore the controllability is not full rank.

The proof of (4) \Rightarrow (3) requires the 2nd representation theorem, and we shall leave it afterward.

Note that (2) \Rightarrow (1) is obvious. To see that (1) \Rightarrow (2), notice that the controllability matrix \mathcal{C} does not depend on T , so controllability does not depend on duration of time interval. Since the system is time invariant, it does not depend on the initial time (t_0) . \diamond

Proof of (4) \Rightarrow (3) - Optional

To prove (4) \Rightarrow (3), we need the so called 2nd Representation Theorem. It is included here for completeness. We will not go through this part.

Definition Let $A : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ and $W \subset \mathfrak{R}^n$ a subspace with the property that for each $w \in W$, $Aw \in W$. We say that W is A -invariant.

Theorem (2nd Representation theorem) Let

- $A : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ is a linear map (i.e. a matrix)
- $W \subset \mathfrak{R}^n$ be an A -invariant k - dimensional subspace of \mathfrak{R}^n .

There exists a nonsingular $T = (e_1 \ e_2 \ \cdots \ e_n)$ s.t.

- e_1, \dots, e_k form a basis of W ,
- in the basis of e_1, e_2, \dots, e_n , A is represented by:

$$T^{-1}AT = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix}$$

$\tilde{A}_{11} \in \mathfrak{R}^{k \times k}$ where $k = \dim W$.

- For any vector $B \in W$, the representation of B is given by

$$B = T \begin{pmatrix} \tilde{B} \\ 0 \end{pmatrix}.$$

The point of the theorem is that A can be put in a form with 0 in strategic locations (\tilde{A}_{21} are zeros), similarly for any vector $B \in W$ (\tilde{B}_2 are zeros).

Proof: Take e_1, \dots, e_k to be a basis of W , and choose e_{k+1}, \dots, e_n to complete the basis of \mathfrak{R}^n . Because $Ae_i \in W$ for each $i = 1, \dots, k$,

$$Ae_i = \sum_{l=1}^k \alpha_{li} e_l.$$

Therefore, α_{li} are the coefficients of \tilde{A}_{11} and the coefficients of \tilde{A}_{21} are all zero.

Let $B \in W$, then

$$B = \sum_{l=1}^k \beta_l e_l$$

so that β_l are simply the coefficients in \tilde{B}_1 and the coefficients of \tilde{B}_2 are zeros. \diamond

Proof of (4) \Rightarrow (3) in controllability test: Notice that

$$\begin{aligned} & A(B \ AB \ A^2B \ \dots \ A^{n-1}B) \\ &= (AB \ A^2B \ \dots \ A^{n-1}B \ A^nB) \end{aligned}$$

therefore, by the Cayley Hamilton Theorem, the range space of the controllability matrix is A -invariant.

Suppose that (3) is not true, so that the rank of

$$(B \ AB \ A^2B \ \dots \ A^{n-1}B)$$

is less than n . Since the controllability matrix is A -invariant, by the 2nd representation theorem, there is an invertible matrix T so that in the new coordinates, $z = T^{-1}x$,

$$\dot{z} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} z + \begin{pmatrix} \tilde{B} \\ 0 \end{pmatrix} u \quad (3.6)$$

where $B = T \begin{pmatrix} \tilde{B} \\ 0 \end{pmatrix}$, and

$$A = T \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} T^{-1}.$$

and the dim of \tilde{A}_{22} is non-zero (it is $n - \text{rank}(\mathcal{C})$).

Let v_2^T be a left eigenvector of \tilde{A}_{22} so that

$$v_2^T(\lambda I - \tilde{A}_{22}) = 0$$

for some $\lambda \in \mathcal{C}$. Define $v^T := (0 \ v_2^T)T^{-1}$. Evaluating $v^T B$, $v^T(\lambda I - A)$ gives

$$\begin{aligned} v^T B &= (0 \ v_2^T)T^{-1}T \begin{pmatrix} \tilde{B} \\ 0 \end{pmatrix} = 0 \\ v^T(\lambda I - A) &= \lambda(0 \ v_2^T)T^{-1}(\lambda T T^{-1} - T \tilde{A} T^{-1}) \\ &= (0 \ v_2^T)(\lambda I - \tilde{A})T^{-1} = 0. \end{aligned}$$

This shows that $[\lambda I - A, B]$ has rank less than n .

Remarks:

1. Notice that the controllability matrix is not a function of the time interval. Hence, if a LTI system is controllable over some interval, it is controllable over any (non-zero) interval. *c.f.* with result of linear time varying system.
2. Because of the above fact, we often say that the pair (A, B) is controllable.
3. Controllability test can be done by just examining A and B without computing the grammian. The test in (4) is attractive in that it enumerates the vectors in the controllability subspace. However, numerically, since it involves power of A , numerical stability needs to be considered.
4. The PBH Test in (5), or the Hautus test for short, involves simply checking the condition at the eigenvalues. It is because for $(sI - A, B)$ to have rank less than n , s must be an eigenvalue.
5. The range space of the controllability matrix is of special interests. It is called the **controllable subspace** and is the set of all states that can be reached from zero-initial condition. **This is A -invariant.**
6. Using the basis for the **controllable subspace** as part of the basis for \mathfrak{R}^n , the controllability property can be easily seen in (3.6).

3.7 Linear Time Invariant (LTI) discrete time system

For the discrete time system (3.7),

$$\begin{aligned}x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k)\end{aligned}\tag{3.7}$$

where $x \in \mathfrak{R}^n$, $u \in \mathfrak{R}^m$, the conditions for controllability is fairly similar except that we need to consider controllability over period of length greater than n .

First, recall that the reachability map for a linear discrete time system $L_{r,[k,k_1]} : u(\cdot) \mapsto s(k, k_0, x=0, u(\cdot))$ is given by:

$$L_{r,[k_0,k_1]}[u(\cdot)] = \sum_{i=k_0}^{k-1} \prod_{j=i+1}^{k-1} A(j)B(i)u(i)\tag{3.8}$$

Theorem For the discrete time system (3.7), the followings are equivalent:

1. The system is controllable over the interval $[0, T]$, with $T \geq n$ (i.e. it can transfer from any arbitrary state at $k=0$ to any other arbitrary state at $k=T$).
2. The system is controllable over the interval $[k_0, k_1]$, with $k_1 - k_0 \geq n$
3. The controllability grammian,

$$W_{r,T} := \sum_{l=0}^{T-1} A^l B B^* A^{*l}$$

is full rank n .

4. The controllability matrix

$$(B \quad AB \quad A^2B \quad \cdots \quad A^{n-1}B)$$

has rank n . Notice that the controllability matrix has dimension $n \times nm$ where m is the dimension of the control vector u .

5. For each $z \in \mathcal{C}$, the matrix,

$$(zI - A \quad B)$$

has rank n .

Proof: The proof of this theorem is exactly analogous to the continuous time case. However, in the case of the equivalence of (2) and (3), we can notice that

$$x(k) = A^k x(0) + [B, AB, A^2 B, \dots, A^{n-1} B] \begin{pmatrix} u(k-1) \\ u(k-2) \\ \vdots \\ u(0) \end{pmatrix}$$

so clearly one can transfer to arbitrary states at $k = n$ iff the controllability matrix is full rank. Controllability does not improve for $T > n$ is again the consequence of Cayley-Hamilton Theorem. \diamond

3.8 Observability

The observability question deals with the question of while not knowing the initial state, whether one can determine the state from the input and output. This is equivalent to the question of whether one can determine the initial state $x(0)$ when given the input $u(t), t \in [0, T]$ and the output $y(t), t \in [0, T]$.

Definition 3.8.1 A state determined dynamical system is called observable on $[t_0, t_1]$ if for all inputs, $u(\tau)$, and outputs $y(\tau)$, $\tau \in [t_0, t_1]$, the state $x(t_0) = x_0$ can be uniquely determined.

For observability, the null space of the observability map on the interval $[t_0, t_1]$, given by:

$$L_o : x_0 \mapsto y(\cdot) = C(\cdot)\Phi(\cdot, t_0)x_0, \quad y(\tau) = C(\tau)\Phi(\tau, t_0)x_0, \quad \forall \tau \in [t_0, t_1]$$

must be trivial (i.e. only contains 0). Otherwise, if $L_o(x_n)(t) = y(t) = 0$, for all $t \in [0, T]$, then for any $\alpha \in \mathfrak{R}$,

$$y(t) = C(t)\Phi(t, 0)x = \Phi(t, 0)(x + \alpha x_n).$$

Recall from the finite rank theorem that if the observability map L_o is a matrix, then the null space of L_o , $\mathcal{N}(L_o) = \mathcal{N}(L_o^T L_o)$. For continuous time system, L_o can be thought of as an infinitely long matrix, then the matrix product operation in $L_o^T L_o$ becomes an integral, and is written as $L_o^* L_o$ where L_o^* denotes the adjoint of the L_o .

The observability Grammian is given by:

$$W_{o, [t_0, t_1]} = L_o^* L_o = \int_{t_0}^{t_1} \Phi(t, t_0)^T C^T(t) C(t) \Phi(t, t_0) dt \in \mathfrak{R}^{n \times n}$$

where L_o^* is the adjoint (think transpose) of the observability map.

3.9 Least square estimation

3.9.1 Formulation and solution

Consider the continuous time varying system:

$$\dot{x}(t) = A(t)x(t) \quad y(t) = C(t)x(t).$$

The observability map on $[0, t_1]$ is given by: $\forall \tau \in [0, t_1]$

$$L_o : x(0) \mapsto y(\tau) = C(\tau)\Phi(\tau, 0)x(0) \quad (3.9)$$

The so-called Least squares estimate of $x(0)$ given the output $y(\tau)$, $\tau \in [0, t_1]$ is:

$$\hat{x}(0|t_1) = \arg \min_{x_0} \int_0^{t_1} e^T(\tau)e(\tau)d\tau.$$

The least squares estimation problem can be cast into a set of linear equations. For example, we can stack all the outputs $y(t)$, $t \in [0, t_1]$ into a vector $Y = [Y(0), \dots, Y(0.1), \dots, Y(0.2), \dots, Y(t_1)]^T \in \mathfrak{R}^{pp}$, and represent L_o as a matrix $\in \mathfrak{R}^{pp \times n}$, then,

$$Y = L_o x_0 + E;$$

$$x_o^* = \operatorname{argmin}_{x_0} E^T E = (L_o^T L_o)^{-1} L_o^T Y$$

Here, $E = [E(0), \dots, E(t_1)]^T \in \mathfrak{R}^{pp}$ is the modeling error or noise. The optimal estimate x_0^* uses the minimum amount of noise E to explain the data Y .

Notice that the resulting E satisfies $E^T L_o = 0$, i.e. E is normal to $\operatorname{Range}(L_o)$.

Using the observability map for the LTI continuous time system (3.9) as L_o ,

$$\hat{x}(0|t_1) = W_o^{-1}(t_1) \int_0^{t_1} \Phi^T(\tau, 0) C^T(\tau) y(\tau) d\tau$$

$$W_o(t_1) = L_o^* L_o = \int_0^{t_1} \Phi^T(\tau, 0) C^T(\tau) C(\tau) \Phi(\tau, 0) d\tau.$$

One can also find estimates of the states at time t_2 given data up to time t_1 :

$$\hat{x}(t_2|t_1) = \Phi(t_2, 0)\hat{x}(0|t_1)$$

This refers to three classes of estimation problems:

- $t_2 < t_1$: Filtering problem
- $t_2 > t_1$: Prediction problem
- $t_2 = t_1$: Observer problem

The estimate $\hat{x}(t|t)$ is especially useful for control systems as it allows us to effectively do state feedback control.

3.9.2 Recursive Least Squares Observer

The least squares solution computed above is expensive to compute:

$$\hat{x}(t|t) = \Phi(t, 0)W_o^{-1}(t) \int_0^t \Phi^T(\tau, 0)C^T(\tau)y(\tau)d\tau$$

$$W_o(t) = L_o^*L_o = \int_0^t \Phi^T(\tau, 0)C^T(\tau)C(\tau)\Phi(\tau, 0)d\tau.$$

It is desirable if the estimate $\hat{x}(t|t)$ can be obtained recursively via a differential equation.

$$\begin{aligned} \frac{d}{dt}\hat{x}(t|t) &= A(t) \underbrace{\Phi(t, 0)W_o^{-1}(t) \int_0^t \Phi^T(\tau, 0)C^T(\tau)y(\tau)d\tau}_{\hat{x}(t|t)} \\ &\quad + \Phi(t, 0)\frac{d}{dt}W_o^{-1}(t) \int_0^t \Phi^T(\tau, 0)C^T(\tau)y(\tau)d\tau \\ &\quad + \Phi(t, 0)W_o^{-1}(t)\Phi^T(t, 0)C^T(t)y(t) \end{aligned}$$

Now by differentiating $W_o(t)W_o^{-1}(t) = I$, we have:

$$\frac{d}{dt}W_o^{-1}(t) = -W_o^{-1}(t)\frac{dW_o}{dt}(t)W_o^{-1}(t)$$

where

$$\frac{d}{dt}W_o(t) = \Phi^T(t, 0)C^T(t)C(t)\Phi(t, 0)$$

$$\begin{aligned} \frac{d}{dt}\hat{x}(t|t) &= A(t)\hat{x}(t|t) - \Phi(t, 0)W_o^{-1}(t)\Phi^T(t, 0)C^T(t)C(t)\hat{x}(t|t) + \Phi(t, 0)W_o^{-1}(t)\Phi^T(t, 0)C^T(t)y(t) \\ &= A(t)\hat{x}(t|t) - \underbrace{\Phi(t, 0)W_o^{-1}(t)\Phi^T(t, 0)}_{P(t)}C^T(t)\underbrace{[C(t)\hat{x}(t|t) - y(t)]}_{\hat{y}(t|t)} \end{aligned} \quad (3.10)$$

$P(t)$ satisfies:

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)C^T(t)C(t)P(t)$$

Typically, we would initialize $P(0)$ to be a large non-singular matrix, e.g. $P(0) = \alpha \cdot I$, where α is large.

In the recursive least squares observer in (3.10), the first term is an openloop prediction term, and the second term is called an output injection where an output prediction error is used to correct the estimation error.

We can think of the observer feedback gain as:

$$L(t) = P(t)C^T(t)$$

3.9.3 Covariance Windup

Let us rewrite $P(t)$ as follows:

$$\begin{aligned} P(t) &= \Phi(t, 0) \left[\int_0^t \Phi^T(\tau, 0)C^T(\tau)C(\tau)\Phi(\tau, 0)d\tau \right]^{-1} \Phi^T(t, 0) \\ &= \left[\int_0^t \Phi^T(\tau, t)C^T(\tau)C(\tau)\Phi(\tau, t)d\tau \right]^{-1} =: K^{-1}(t) \end{aligned}$$

The matrix $K(t) := P^{-1}(t)$ is known as the *Information Matrix* and satisfies:

$$\dot{K} = C^T(t)C(t) - A^T(t)K(t) - K(t)A(t)$$

Notice that $W_o(t)$ increases with time. Thus, $W_o^{-1}(t)$ decreases with time, as would $K^{-1}(t)$. This means that the observer gain $L(t)$ will decrease. The observer will asymptotically rely more and more on open loop estimation:

$$\frac{d}{dt}\hat{x}(t|t) = A(t)\hat{x}(t|t)$$

Forgetting factor

Forget old information (τ far away from current time): Choose $\lambda > 0$:

$$K(t) = \int_0^t \Phi^T(\tau, t)C^T(\tau)C(\tau)\Phi(\tau, t)e^{-\lambda(t-\tau)}d\tau;$$

$$\dot{K} = C^T(t)C(t) - A^T(t)K(t) - K(t)A(t) - \lambda K$$

Covariance reset

If $K(t)$ becomes large, reset it to small number.

- Check maximum singular value of $K(t)$, $\bar{\sigma}K(t)$.
- If $\bar{\sigma}K(t) \geq \lambda_{max}$, $K(t) = \epsilon I$, i.e. $P(t) = I/\epsilon$.

3.10 Observability Tests

The observability question deals with the question of whether one can determine what the initial state $x(0)$ is, given the input $u(t)$, $t \in [0, T]$ and the output $y(t)$, $t \in [0, T]$.

For observability, the null space of the observability map,

$$L_o : x \mapsto y(\cdot) = \Phi(\cdot, 0)x, \quad \tau \in [0, T],$$

must be trivial (i.e. only contains 0). Otherwise, if $L_o(x_n)(t) = y(t) = 0$, for all $t \in [0, T]$, then for any $\alpha \in \mathfrak{R}$,

$$y(t) = \Phi(t, 0)x = \Phi(t, 0)(x + \alpha x_n).$$

Hence, one cannot distinguish the various initial conditions from the output.

Just like for controllability, it is inconvenient to check the rank (and the null space) of the L_o which is tall and thin. We can instead check the rank and the null space of the observability gramian given by:

$$W_{o,T} = L_o^*L_o$$

where L_o^* is the adjoint (think transpose) of the observability map.

Proposition 3.10.1 $\text{Null}(L_o) = \text{Null}(L_o^*L_o)$.

Proof:

- Let $x \in \text{Nul}(L_o)$ so, $L_o x = 0$. Then, clearly, $L_o^T L_o x = L^T 0$. Therefore, $\text{Nul}(L_o) \subset \text{Nul}(L_o^* L_o)$.
- Let $x \in \text{Nul}(L_o^* L_o)$. Then, $x^T L_o^* L_o x = 0$. Or, $(L_o x)^T (L_o x) = 0$. This can only be true if $L_o x = 0$. Therefore, $\text{Nul}(L_o^* L_o) \subset \text{Nul}(L_o)$.

◇

3.11 Observability Tests for Continuous time LTI systems

Theorem For the LTI continuous time system (3.5), the followings are equivalent:

1. The system is observable over the interval $[0, T]$
2. The observability grammian,

$$W_{o,T} := \int_0^T e^{A^* t} C^* C e^{A t} dt$$

is full rank n .

3. The observability matrix $\begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{pmatrix}$ has rank n . Notice that the observability matrix has dimension $np \times n$ where p is the dimension of the output vector y .

4. For each $s \in \mathcal{C}$, the matrix,

$$\begin{pmatrix} sI - A \\ C \end{pmatrix}$$

has rank n .

Proof: The proof is similar to the ones as in the controllable case and will not be repeated here. Some differences are that instead of considering $1 \times n$ vector v^T multiplying on the left hand sides of the controllability matrix and the grammians, we consider $n \times 1$ vector multiplying on the RHS of the observability matrix and the grammian etc.

Also instead of considering the range space of the controllability matrix, we consider the NULL space of the observability matrix. Its null space is also A-invariant. Hence if the observability matrix is not full rank, then using basis for its null space as the last k basis vectors of \mathfrak{R}^n , the system can be represented as:

$$\begin{aligned} \dot{z} &= \begin{pmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix} z + \begin{pmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{pmatrix} u \\ y &= (\tilde{C} \quad 0) z \end{aligned}$$

where $C = (\tilde{C} \quad 0)T^{-1}$, and

$$A = T \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} T^{-1}.$$

and the dim of \tilde{A}_{22} is non-zero (and is the dim of the null space of the observability matrix). \diamond

Remark

1. Again, observability of a LTI system does not depend on the time interval. So, theoretically speaking, if observing the output and input for an arbitrary amount of time will be sufficient to figure out $x(0)$. In reality, when more data is available, one can do more averaging to eliminate effects of noise (e.g. using the Least square ie. Kalman Filter approach).
2. The subspace of particular interest is the null space of the controllability matrix. An initial state lying in this set will generate identically 0 zero-input response. This subspace is called the **unobservable subspace**.
3. Using the basis of the **unobservable subspace** as part of the basis of \mathfrak{R}^n , the observability property can be easily seen.

3.12 Observability Tests for Discrete time LTI systems

The tests for observability of the discrete time system (3.7) is given similarly by the following theorem.

Theorem 3.12.1 *For the discrete time system (3.7), the followings are equivalent:*

1. *The system is observable over the interval $[0, T]$, for some $T \geq n$.*
2. *The observability grammian,*

$$W_{o,T} := \sum_{k=0}^{T-1} A^{*k} C^* C A^k$$

is full rank n .

3. *The observability matrix $\begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{pmatrix}$ has rank n . Notice that the observability matrix has dimension $np \times n$ where p is the dimension of the output vector y .*

4. *For each $z \in \mathcal{C}$, the matrix,*

$$\begin{pmatrix} zI - A \\ C \end{pmatrix}$$

has rank n .

The controllability matrix can have the following interpretation: zero-input response is given by:

$$\begin{pmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(n-1) \end{pmatrix} = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{pmatrix} x(0).$$

Thus, clearly if the observability matrix is full rank, one can reconstruct $x(0)$ from measurements of $y(0), \dots, y(n-1)$. One might think that by increasing the number of output measurement times (e.g. $y(n)$) the system can become observable. However, because of Cayley-Hamilton theorem, $y(n) = CA^n x(0)$ can be expressed as $\sum_{i=0}^{n-1} a_i CA^i x(0)$ consisting of rows already in the observability matrix.

3.13 PHB test and Eigen/Jordan Forms

Consider

$$\dot{x} = Ax + Bu, \quad y = Cx$$

Suppose first that A is semi-simple, then $A = TDT^{-1}$ and $D = \text{diag } \lambda_1, \dots, \lambda_n$ where λ_i might be repeating.

The PBH test in the transformed eigen-coordinates is to check

$$\text{rank}(\lambda_i D, T^{-1}B)$$

for each i if it is n . This shows that $T^{-1}B$ must be non-zero in each row. Also, if $\lambda_i = \lambda_j$ for some $i \neq j$, then a single input system cannot be controllable.

Now suppose that $A = TJT^{-1}$ and J is in Jordan form,

$$J = \begin{pmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{pmatrix}$$

This Jordan form has 3 generalized eigenvector chains.

Controllability

- From homework, we saw that if the entries of $B \in \mathfrak{R}^4$ at the **beginning** of some *chain* is 0, i.e..

$$T^{-1}B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ 0 \end{pmatrix}, \quad T^{-1}B = \begin{pmatrix} b_1 \\ b_2 \\ 0 \\ b_4 \end{pmatrix}, \quad T^{-1}B = \begin{pmatrix} b_1 \\ 0 \\ b_3 \\ b_4 \end{pmatrix}$$

then the system is uncontrollable.

- Thus, $T^{-1}B$ being nonzero at the beginning of each chain is a necessary condition for controllability.
- **PBH test shows that it is in fact a sufficient condition.**

Observability

- Similarly, if $C \cdot T \in \mathfrak{R}^{1 \times n}$ is a zero entry at the **end** of any *chain* (these correspond to the coordinates for the eigenvector), i.e.

$$CT = (0 \quad c_2 \quad c_3 \quad c_4), \quad CT = (c_1 \quad c_2 \quad 0 \quad c_4), \\ \text{or, } CT = (c_1 \quad c_2 \quad c_3 \quad 0)$$

then we saw from the homework that the system is not observable.

- Hence, CT having non-zero entries at the **end** of each *chain* is necessary for observability.
- PHB test shows that it is also sufficient.

Note If the system is either uncontrollable or unobservable, then the order of the transfer function will be less than the order of A .

However, for a multi-input-multi-output system, the fact that the order of each transfer function in the transfer function matrix has order less than the order of A , does NOT imply that the system is uncontrollable or unobservable. e.g.

$$A = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix} \quad B = C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

gives,

$$Y(s) = \begin{pmatrix} \frac{1}{s+1} & 0 \\ 0 & \frac{1}{s+2} \end{pmatrix} U(s).$$

3.14 Kalman Decomposition

3.14.1 Controllable / uncontrollable decomposition

Suppose that the controllability matrix $\mathcal{C} \in \mathfrak{R}^{n \times n}$ of a system has rank $n_1 < n$. Then there exists an invertible transformation, $T \in \mathfrak{R}^{n \times n}$ such that:

$$\begin{aligned} z &= T^{-1}x, \\ \dot{z} &= \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} z + \begin{pmatrix} \tilde{B} \\ 0 \end{pmatrix} u \end{aligned} \quad (3.11)$$

where $B = T \begin{pmatrix} \tilde{B} \\ 0 \end{pmatrix}$, and

$$A = T \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} T^{-1}.$$

and the dim of \tilde{A}_{22} is $n - n_1$.

3.14.2 Observable / unobservable decomposition

Hence if the observability matrix is not full rank, then using basis for its null space as the last k basis vectors of \mathfrak{R}^n , the system can be represented as:

$$\begin{aligned} \dot{z} &= \begin{pmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} \end{pmatrix} z + \begin{pmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{pmatrix} u \\ y &= (\tilde{C} \quad 0)z \end{aligned}$$

where $C = (\tilde{C} \quad 0)T^{-1}$, and

$$A = T \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{pmatrix} T^{-1}.$$

and the dim of \tilde{A}_{22} is non-zero (and is the dim of the null space of the observability matrix).

3.14.3 Kalman decomposition

The Kalman decomposition is just the combination of the observable/unobservable, and the controllable/uncontrollable decomposition.

Theorem 3.14.1 *There exists a coordinate transformation $z = T^{-1}x \in \mathbb{R}^n$ such that*

$$\begin{aligned} \dot{z} &= \begin{pmatrix} \tilde{A}_{11} & 0 & \tilde{A}_{13} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} & \tilde{A}_{24} \\ 0 & 0 & \tilde{A}_{33} & 0 \\ 0 & 0 & \tilde{A}_{43} & \tilde{A}_{44} \end{pmatrix} z + \begin{pmatrix} \tilde{B}_1 \\ \tilde{B}_2 \\ 0 \\ 0 \end{pmatrix} u \\ y &= (\tilde{C}_1 \quad 0 \quad \tilde{C}_2 \quad 0)z \end{aligned}$$

Let $T = (T_1 \ T_2 \ T_3 \ T_4)$ (with compatible block sizes), then

- $T_2 \rightarrow (C - \bar{O})$: $T_2 =$ basis for $\text{Range}(C) \cap \text{Null}(\bar{O})$.
- $T_1 \rightarrow (C - O)$: $T_1 \cup T_2 =$ basis for $\text{Range}(C)$
- $T_4 \rightarrow (\bar{C} - \bar{O})$: $T_2 \cup T_4 =$ basis for $\text{Null}(\bar{O})$.
- $T_3 \rightarrow (\bar{C} - O)$: $T_1 \cup T_2 \cup T_3 \cup T_4 =$ basis for \mathbb{R}^n .

Note: Only T_2 is uniquely defined.

Proof: This is just combination of the obs/unobs and cont/uncont decomposition. \diamond

3.14.4 Stabilizability and Detectability

If the uncontrollable modes \bar{A}_{33} , \bar{A}_{44} are stable (have eigenvalues on the Left Half Plane), then, system is called *stabilizable*.

- One can use feedback to make the system stable;
- Uncontrollable modes decay, so not an issue.

If the unobservable modes \bar{A}_{22} , \bar{A}_{44} are stable (have eigenvalues on the Left Half Plane), then, system is called *detectable*.

- The states z_2, z_4 decay to 0
- Eventually, they will have no effect on the output
- State can be reconstructed by ignoring the unobservable states (eventually).

The PBH tests can be modified to check if the system is stabilizable or detectable, namely, check

$$\text{rank}(sI - A \quad B) \quad \text{rank} \begin{pmatrix} sI - A \\ C \end{pmatrix}$$

for all s with non-negative real parts if they lose rank. Specifically, one needs to check only s which are the unstable eigenvalues of A .

3.15 Realization

3.15.1 Degree of controllability / observability

Formal definitions for controllability and observability are black and white. In reality, some states are very costly to control to, or have little effect on the outputs. They are therefore, effectively uncontrollable or unobservable.

Degree of controllability and observability can be evaluated by the sizes of W_r and W_o with infinite time horizon:

$$\begin{aligned} W_r(0, \infty) &= \lim_{T \rightarrow \infty} \int_0^T e^{A(T-\tau)} B B^T e^{A^T(T-\tau)} d\tau \\ &= \int_0^\infty e^{At} B B^T e^{A^T t} dt \\ W_o(0, \infty) &= \lim_{T \rightarrow \infty} \int_0^T e^{A^T \tau} C^T C e^{A \tau} d\tau \end{aligned}$$

Notice that W_r satisfies differential equation:

$$\frac{d}{dt} W_r(0, t) = A W_r(0, t) + W_r(0, t) A^T + B B^T$$

Thus, if all the eigenvalues of A have strictly negative real parts (i.e. A is stable), then, as $T \rightarrow \infty$,

$$0 = A W_r + W_r A^T + B B^T$$

which is a set of linear equations in the entries of W_r , and can be solved effectively. The equation is called a Lyapunov equation. Matlab can solve for W_r using **gram**.

Similarly, if A is stable, as $T \rightarrow \infty$, W_o satisfies the Lyapunov equation

$$0 = A^T W_o + W_o A + C^T C$$

which can be solved using linear method (or using Matlab via **gram**).

W_r is related to the minimum norm control. To reach $x(0) = x_0$ from $x(-\infty) = 0$ in infinite time,

$$\begin{aligned} u &= L_r^T W_c^{-1} x_0 \\ \min_{u(\cdot)} J(u) &= \min_{u(\cdot)} \int_{-\infty}^0 u^T(\tau) u(\tau) d\tau = x_0^T W_c^{-1} x_0. \end{aligned}$$

If state x_0 is difficult to reach, then $x_0^T W_c^{-1} x_0$ is large meaning that x_0 is practically uncontrollable.

Similarly, if the initial state is $x(0) = x_0$, then, with $u(\tau) \equiv 0$

$$\int_0^\infty y^T(\tau) y(\tau) d\tau = x_0^T W_o x_0.$$

Thus, $x_0^T W_o x_0$ measures the effect of the initial condition on the output. If it is small, the effect of x_0 in the output $y(\cdot)$ is small, thus, it is hard to observe, or practically unobservable.

Generally, one can look at the smallest eigenvalues W_c and W_o , the span of the associated eigenvectors will be difficult to control, or difficult to observe.

3.15.2 Relation to Transfer Function

For the system

$$\begin{aligned}\dot{x} &= Ax + Bu \\ y &= Cx + Du\end{aligned}$$

The transfer function from $U(s) \rightarrow Y(s)$ is given by:

$$\begin{aligned}G(s) &= C(sI - A)^{-1}B + D \\ &= \frac{C \operatorname{adj}(sI - A)B}{\det(sI - A)} + D.\end{aligned}$$

Note that transfer function is not affected by similarity transform.

From Kalman decomposition, it is obvious that

$$G(s) = \tilde{C}_1(sI - \tilde{A}_{11})^{-1}\tilde{B}_1 + D$$

where \tilde{A}_{11} , \tilde{B}_1 , \tilde{C}_1 correspond to the controllable and observable component.

Poles

- they are values of $s \in \mathcal{C}$ s.t. $G(s)$ becomes infinite.
- poles of $G(s)$ are clearly the eigenvalues of \tilde{A}_{11} .

Zeros

- For the SISO case, $z \in \mathcal{C}$ (complex numbers) is a zero if it makes $G(z) = 0$.
- Thus, zeros satisfy

$$C \operatorname{adj}(zI - A)B + D \det(zI - A) = 0$$

assuming system is controllable and observable, otherwise, need to apply Kalman decomposition first.

- In the general MIMO case, a zero implies that there can be non-zero inputs $U(s)$ that produce output $Y(s)$ that is identically zero. Therefore, there exists $X(z)$, $U(z)$ such that:

$$\begin{aligned}zI - A X(z) &= B U(z) \\ 0 &= C X(z) + D U(z)\end{aligned}$$

This will be true if and only if at $s = z$

$$\operatorname{rank} \begin{pmatrix} sI - A & -B \\ -C & -D \end{pmatrix}$$

is less than the normal rank of the matrix at other s .

We shall return to the multi-variable zeros when we discuss the effect of state feedback on the zero of the system.

3.15.3 Pole-zero cancellation

Cascade system with input/output (u_1, y_2) :

- System 1 with input/output (u_1, y_1) has a zero at α and a pole at β .

$$\begin{aligned}\dot{x}_1 &= A_1x_1 + B_1u_1 \\ y_1 &= C_1x_1\end{aligned}$$

- System 2 with input/output (u_2, y_2) has a pole at α and a zero at β .

$$\begin{aligned}\dot{x}_2 &= A_2x_2 + B_2u_2 \\ y_2 &= C_2x_2\end{aligned}$$

- Cascade interconnection: $u_2 = y_1$.

Then

1. The system pole at β is not observable from y_2
2. The system pole at α is not controllable from u_1 .

Combined system:

$$A = \begin{pmatrix} A_1 & 0 \\ B_2C_1 & A_2 \end{pmatrix}; \quad B = \begin{pmatrix} B_1 \\ 0 \end{pmatrix}; \quad C = (0 \quad C_2)$$

Consider $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

- Use PBH test with $\lambda = \beta$.

$$\begin{pmatrix} \beta I - A_1 & 0 \\ -B_2C_1 & \beta I - A_2 \\ 0 & C_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0???$$

- Let x_1 be eigenvector associated with β for A_1 .
- Let $x_2 = (\beta I - A_2)^{-1}B_2C_2x_1$ (i.e. solve second row in PBH test).
- Then since

$$Cx = C_2x_2 = (C_2(\beta I - A_2)^{-1}B_2)C_2x_1$$

and β is a zero of system 2, PHB test shows that the mode β is not observable.

Similar for the case when α mode in system 2 not controllable by u .

These example shows that when cascading systems, it is important not to do unstable pole/zero cancellation.

3.16 Balanced Realization

The objective is to reduce the number of states while minimizing effect on I/O response (transfer function). This is especially useful when the state equations are obtained from distributed parameters system, P.D.E. via finite elements methods, etc. which generate many states. If states are unobservable or uncontrollable, they can be removed. The issue at hand is if some states are lightly controllable while others are lightly observable. Do we cancel them?

In particular,

- Some states can be lightly controllable, but heavily observable.
- Some states can be lightly observable, but heavily controllable.
- Both contribute to significant component to input/output (transfer function) response.

Idea: Exhibit states so that they are simultaneously lightly (heavily) controllable and observable.

Consider a stable system:

$$\begin{aligned}\dot{x} &= Ax + Bu \\ y &= Cx + Du\end{aligned}$$

If A is stable, reachability and observability grammians can be computed by solving the Lyapunov equations:

$$\begin{aligned}0 &= AW_r + W_r A^T + BB^T \\ 0 &= A^T W_o + W_o A + C^T C\end{aligned}$$

- If state x_0 is difficult to reach, then $x_0^T W_r^{-1} x_0$ is large \Rightarrow practically uncontrollable.
- If $x_0^T W_o x_0$ is small, the signal of x_0 in the output $y(\cdot)$ is small, thus, it is hard to observe \Rightarrow practically unobservable.
- Generally, one can look at the smallest eigenvalues W_r and W_o , the span of the associated eigenvectors will be difficult to control, or difficult to observe.

Transformation of grammians: (beware of which side T goes !):

$$z = Tx$$

Minimum norm control should be invariant to coordinate transformation:

$$x_0^T W_r^{-1} x_0 = z^T W_{r,z}^{-1} z = x^T T^T W_{r,z}^{-1} T x$$

Therefore

$$W_r^{-1} = T^T W_{r,z}^{-1} T \Rightarrow W_{r,z} = T \cdot W_r T^T.$$

Similarly, the energy in a state transmitted to output should be invariant:

$$x^T W_o x = z W_{o,z} z \Rightarrow W_{o,z} = T^{-T} W_o T^{-1}.$$

Theorem 3.16.1 (*Balanced realization*) *If the linear time invariant system is controllable and observable, then, there exists an invertible transformation $T \in \mathbb{R}^{n \times n}$ s.t. $W_{o,z} = W_{r,z}$.*

The balanced realization algorithm:

1. Compute controllability and observability grammians:

```
>> Wr = gram(sys, 'c');
>> Wo = gram(sys, 'o');
```

2. Write $W_o = R^T R$ [One can use SVD]

```
>> [U,S,V]=svd(Wo);
>> R = sqrtm(S)*V';
```

3. Diagonalize RW_rR^T [via SVD again]:

$$RW_rR^T = U\Sigma^2U^T,$$

where $UU^T = I$ and Σ is diagonal.

```
>> [U1,S1,V1]=svd(R*Wr*R');
>> Σ = sqrtm(S1');
```

4. Take $T = \Sigma^{-\frac{1}{2}}U^T R$

```
>> T = inv(sqrtm(Σ))*U1'*R;
```

5. This gives $W_{r,z} = W_{o,z} = \Sigma$.

```
>> Wr,z = T * Wr * T',
>> Wo,z = inv(T)' * Wo * inv(T)
```

Proof: Direct substitution!

- Once $W_{r,z}$ and $W_{o,z}$ are diagonal, and equal, one can decide to eliminate states that correspond to the smallest few entries.
- Do not remove unstable states from the realization since these states need to be controlled (stabilized)
- It is also possible to maintain some D.C. performance.

Consider the system in the balanced realization form:

$$\begin{aligned} \frac{d}{dt} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} u \\ y &= C_1 z_1 + C_2 z_2 + Du \end{aligned}$$

where $z_1 \in \mathbb{R}^{n_1}$ and $z_2 \in \mathbb{R}^{n_2}$, and $n_1 + n_2 = n$. Suppose that the z_1 are associated with the simultaneously lightly controllable and light observable mode.

The naive truncation approach is to remove z_2 to form the reduced system:

$$\begin{aligned} \dot{z}_1 &= A_{11}z_1 + B_1u \\ y &= C_1z_1 + Du \end{aligned}$$

However, this does not retain the D.C. (steady state) gain from u to y (which is important from a regulation application point of view).

To maintain the steady state response, instead of eliminating z_2 completely, replace z_2 by its steady state value:

In steady state:

$$\begin{aligned}\dot{z}_2 &= 0 = A_{21}z_1 + A_{22}z_2^* + B_2u \\ \Rightarrow z_2^* &= -A_{22}^{-1}(A_{21}z_1 + B_2u)\end{aligned}$$

This is feasible if A_{22} is invertible (0 is not an eigenvalue).

A truncation approach that maintains the steady state gain is to replace z_2 by its steady state value:

$$\begin{aligned}\dot{z}_1 &= A_{11}z_1 + A_{12}z_2^* + B_1u \\ y &= C_1z_1 + C_2z_2^* + Du\end{aligned}$$

so that,

$$\begin{aligned}z_1 &= \underbrace{[A_{11} - A_{12}A_{22}^{-1}A_{21}]}_{A_r} + \underbrace{[B_1 - A_{12}A_{22}^{-1}B_2]}_{B_r} u \\ y &= \underbrace{[C_1 - C_2A_{22}^{-1}A_{21}]}_{C_r} z_1 + \underbrace{[D - C_2A_{22}^{-1}B_2]}_{D_r} u\end{aligned}$$

The truncated system with state $z_1 \in \mathfrak{R}^{n_1}$ is

$$\begin{aligned}\dot{z}_1 &= A_r z_1 + B_r u \\ y &= C_r z_1 + D_r u\end{aligned}$$

which will have the same steady state response as the original system.