Module 3F2: Systems and Control

LECTURE NOTES 3: OBSERVERS

1 Sampled Data Control System

2 Solving Linear Equations

3 Observability
   3.1 Effect of Initial Condition on Output
   3.2 Change of State Coordinates when System is not Observable

4 Observers
   4.1 Differentiating signals is a bad idea
   4.2 Observer structure
   4.3 Application in Ball and Beam experiment
   4.4 Tracking disturbances, ignoring noise
   4.5 Special case: Kalman Filter
   4.6 Application to sensor fusion
   4.7 Application to sensor bias estimation

K. Glover
Jan 2002

Revised by J.M. Maciejowski
February 2008
1 Sampled Data Control System

The sampled data system satisfies:

\[ \dot{x}(t) = Ax(t) + Bu(t), \quad \text{with } u(t) = u(kT), \quad \text{for } kT \leq t < (k+1)T. \]

Apply result from Handout 1, section 4.5 (Convolution integral):

\[
\begin{align*}
\hat{x}(k+1) & = e^{AT} \hat{x} + \int_{kT}^{(k+1)T} e^{A(T-\tau)} B d\tau u(kT) \\
& = \Phi\hat{x}(kT) + \Gamma u(kT)
\end{align*}
\]

where

\[ \Gamma = \int_0^T e^{A\tau} d\tau B = A^{-1}(e^{AT} - I) B, \quad \text{if } \det(A) \neq 0. \]

This gives the standard state-space model in discrete time.

Entirely analogous results can be obtained for the discrete time case as in the continuous time case:

- Solution of vector difference equations,
- Discrete-time convolution,
- \(z\)-transform for frequency response calculations etc,
- Notions of controllability and observability — coming next.
2 Solving Linear Equations

For convenience we will repeat some results and definitions from linear algebra.

Given an \( m \times n \) matrix \( A \) and an \( m \times 1 \) vector \( b \), consider the equation:

\[
Ax = b.
\]

in the unknown \( x \) in \( \mathbb{R}^n \). Two natural questions are:

(a) Does there exist a solution, \( x \)?

(b) If so, is it unique?

**Fact 2.1** For the case \( m = n \):

(a) If \( \det(A) \neq 0 \) then for any \( b \) there exists a solution, \( x \), such that \( Ax = b \), and this solution is unique (Indeed it is given by \( x = A^{-1}b \)).

(b) If \( \det(A) = 0 \) then there exists \( x \neq 0 \) such that \( Ax = 0 \).

**Fact 2.2** For any \( m \times n \) matrix, \( M \),

\[
M^T M x = 0 \iff M x = 0.
\]

**Fact 2.3** For the case \( m \leq n \),

(a) If \( \det(AA^T) \neq 0 \) then \( x = A^T (AA^T)^{-1} b \) solves \( Ax = b \) for any \( b \).

(b) If \( \det(AA^T) = 0 \) then there exists a \( z \neq 0 \) such that \( z \perp A x \) (i.e. \( z^T A x = 0 \)) for all \( x \).

For the case \( m \geq n \),

(c) If \( \det(A^T A) \neq 0 \) then if there is a solution to \( Ax = b \) then it is unique.

(d) If \( \det(A^T A) = 0 \) then there exists \( x \neq 0 \) such that \( Ax = 0 \).
Definition 2.4 Let $A$ be an $m \times n$ matrix then,

(a) the set of all $x$ such that $Ax = 0$ is called the Null Space of $A$;

(b) the set of all $y$ such that $y = Ax$ for some $x$ is called the Range Space of $A$ (or the range of $A$);

(c) $A$ is said to have full row rank if $z^T A \neq 0$ for all $z \neq 0$;

(d) $A$ is said to have full column rank if $Ax \neq 0$ for all $x \neq 0$.

For hand calculations it is generally easiest to use the following observations:

(a) If you can find a set of $n$ rows of $A$ such that the determinant of the $n \times n$ submatrix given by these rows is nonzero, then $A$ has full column rank.

(b) If you can find a nonzero vector, $x$, such that $Ax = 0$ then clearly $A$ does not have full column rank.

(c) If you can find a set of $m$ columns of $A$ such that the determinant of the $m \times m$ submatrix given by these columns is nonzero, then $A$ has full row rank.

(d) If you can find a nonzero vector, $z$, such that $z^T A = 0$ then clearly $A$ does not have full row rank.
3 Observability

A system:

\[ \dot{x} = Ax + Bu \]
\[ y = Cx \]

is called **observable** if we can deduce the state, \( x(t) \), from measurements of \( u(\tau) \) and \( y(\tau) \) over some time interval.

If two initial states \( x_0 \neq x_1 \) give the same outputs then 
\[ 0 = Ce^{At}(x_1 - x_0) \]

A state \( x_0 \) such that \( Ce^{At}x_0 = 0 \) for all \( t \) is called an **unobservable state**.

Define the **observability matrix** \( Q = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \).

\[ Qx_0 = 0 \Rightarrow CA^k x_0 = 0 \] for \( k = 0, \ldots, n-1 \)
\[ \Rightarrow CA^n x_0 = C(-\alpha_1 A^{n-1} - \alpha_{n-1} A - \alpha_n I) x_0 \] by Cayley-Hamilton Theorem
\[ = 0 \]
\[ \Rightarrow CA^k x_0 = 0 \] for all \( k \geq 0 \) by repeated use of Cayley-Hamilton theorem.
\[ \Rightarrow Ce^{At}x_0 = 0 \] for all \( t \) by the power series expansion of \( e^{At} \).

Hence \( x_0 \) is an unobservable state if and only if \( Qx_0 = 0 \).
Consider differentiating $y(t)$ to give

$$\begin{bmatrix}
y(t) \\
y'(t) \\
y''(t) \\
\vdots \\
y^{(n-1)}(t)
\end{bmatrix} = \begin{bmatrix}
C \\
CA \\
CA^2 \\
\vdots \\
CA^{n-1}
\end{bmatrix} \begin{bmatrix}
x(t) + \int_0^t \dot{y}(\tau) \, d\tau \\
\dot{x}(t) + \int_0^t \ddot{y}(\tau) \, d\tau \\
\ddot{x}(t) + \int_0^t \dddot{y}(\tau) \, d\tau \\
\vdots \\
\dddot{x}(t) + \int_0^t \dddddot{y}(\tau) \, d\tau
\end{bmatrix} \begin{bmatrix}
0 \\
CBu(t) \\
CABu(t) + CBu(t) \\
\ddots \\
CA^{n-2}Bu + \cdots + CBu^{(n-2)}
\end{bmatrix}$$

known

We can solve the above equation uniquely for $x(t)$ if and only if $\text{rank } Q = n$, hence

Observability test:

The system is observable if and only if $\text{rank } Q = n$

### 3.1 Effect of Initial Condition on Output

Now consider the difference between two initial condition responses:

$$y_o(t) = Ce^{At}x_o$$

and

$$y(t) = Ce^{At}(x_o + d)$$

so

$$y(t) - y_o(t) = Ce^{At}d$$

Can $\left(y(t) - y_o(t)\right)$ be small in spite of $d$ being large? Measure the size of $\left(y(t) - y_o(t)\right)$ over the time interval $0 < t < t_1$ by

$$\int_0^{t_1} \left\| y(t) - y_o(t) \right\|^2 \, dt = \int_0^{t_1} \left( y(t) - y_o(t) \right)^T \left( y(t) - y_o(t) \right) \, dt$$

$$= \int_0^{t_1} d^T e^{A^T \tau + CT} Ce^{A \tau} d \, d\tau = d^T W_o(t_1) d$$

where $W_o(t_1) = \int_0^{t_1} e^{A^T \tau + CT} Ce^{A \tau} \, d\tau$

Clearly this difference must be $\geq 0$ so $W_o(t_1)$ is a positive semi-definite matrix. The system will be observable if $d^T W_o(t_1) d > 0$ for all $d \neq 0$, i.e. if $W_o(t_1)$ is a positive definite matrix. Also,

$\Rightarrow$ Null Space of $W_o(t_1)$ = Null Space of $Q$. 

$d$ in Null Space of $W_o(t_1)$ $\Leftrightarrow$ $W_o(t_1)d = 0$ $\Leftrightarrow$ $d^T W_o(t_1) d = 0$ $\Leftrightarrow$ $Ce^{At} d = 0$ for all $t < t_1$ $\Leftrightarrow$ $d$ is an unobservable state.
Example

\[ \dot{x} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} x, \quad y = \begin{bmatrix} 1 & 1 \end{bmatrix} x \Rightarrow Ce^{At} = \begin{bmatrix} e^{-t} & e^{-2t} \end{bmatrix} \]

\[ W_o(t_1) = \int_0^{t_1} \begin{bmatrix} e^{-2t} & e^{-3t} \\ e^{-3t} & e^{-4t} \end{bmatrix} dt = \begin{bmatrix} \frac{1}{2} (1 - e^{-2t_1}) & \frac{1}{3} (1 - e^{-3t_1}) \\ \frac{1}{3} (1 - e^{-3t_1}) & \frac{1}{4} (1 - e^{-4t_1}) \end{bmatrix} \xrightarrow{t_1 \to \infty} \begin{bmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{4} \end{bmatrix} \]

3.2 Change of State Coordinates when System is not Observable

If \((A, C)\) is not observable then we can make a change of state coordinates to isolate the unobservable states as follows.

If the rank \(Q = r < n\) then there exists a nonsingular \(n \times n\) matrix \(T\) and a \(pn \times r\) matrix \(\tilde{Q}_1\) of rank \(r\), such that \((\text{Recall QR factorization})\)

\[ Q = \begin{bmatrix} \tilde{Q}_1 & 0 \end{bmatrix} T \]

Now change the state coordinates to \(\tilde{x} = T x\):

\[ \dot{\tilde{x}} = \frac{TAT^{-1}}{\tilde{A}} \dot{\tilde{x}} + \frac{T}{\tilde{B}} u, \quad y = \frac{CT^{-1}}{\tilde{C}} \tilde{x} \]

**Theorem 3.1** In these coordinates if we partition the state, \(\tilde{x} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix}\) with \(\tilde{x}_1\) of dimension \(r\), and compatibly partition:

\[ \tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}; \quad \tilde{B} = \begin{bmatrix} \tilde{B}_1 \\ \tilde{B}_2 \end{bmatrix}; \quad \tilde{C} = \begin{bmatrix} \tilde{C}_1 & \tilde{C}_2 \end{bmatrix} \]

then

\[ \tilde{C}_2 = 0, \quad \tilde{A}_{12} = 0, \quad \text{and} \ (\tilde{A}_{11}, \tilde{C}_1) \text{ is observable} \]
**Proof:** Firstly $CA^{k+1} = CT^{-1}TA^{k+1} = \tilde{C}\tilde{A}^{k}$ so

$$[\tilde{Q}_1 \ 0] = QT^{-1} = \begin{bmatrix} \tilde{C} \\
\tilde{C}\tilde{A} \\
\vdots \\
\tilde{C}\tilde{A}^{n-1} \end{bmatrix}$$

and hence $\tilde{C}\tilde{A}^{k} \begin{bmatrix} 0 \\
n \\
\vdots \\
n \end{bmatrix} = 0$ for all $k$.

Hence $\tilde{C}_2 = 0$. Furthermore

$$\tilde{Q}_1 \tilde{A}_{12} = [\tilde{Q}_1 \ 0] \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\
\tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} \begin{bmatrix} 0 \\
n \end{bmatrix} = \begin{bmatrix} \tilde{C}\tilde{A} \\
\tilde{C}\tilde{A}^2 \\
\vdots \\
\tilde{C}\tilde{A}^{n-1} \end{bmatrix} \begin{bmatrix} 0 \\
n \end{bmatrix} = 0$$

which implies that $\tilde{A}_{12} = 0$ since $\tilde{Q}_1$ is full column rank.

Hence in these state coordinates we have,

$$\dot{\tilde{x}}_1 = \tilde{A}_{11} \tilde{x}_1 + \tilde{B}_1 u, \\ y = \tilde{C}_1 \tilde{x}_1$$

and the input/output response (i.e. the transfer function) depends only on $\tilde{x}_1$ and the states $\tilde{x}_2$ are all unobservable.

## 4 Observers

### 4.1 Differentiating signals is a bad idea

Typically the state is not available for measurement, but we can estimate $\tilde{z}(t)$ from $\tilde{y}$ and $\tilde{u}$

In the section on observability we saw how to exactly deduce $\tilde{z}(t)$ from $y, \dot{y}, \ldots, y^{(n-1)}, u, \dot{u}, \ldots u^{(n-2)}$

but differentiating signals has bad noise amplification problems:

- $y(t) = \sin \omega t + \epsilon \sin \omega_n t$ \quad S/N ratio $= 1/\epsilon$
- $\dot{y}(t) = \omega \cos \omega t + \epsilon \omega_n \cos \omega_n t$ \quad S/N ratio $= (\omega/\omega_n)$
- $\ddot{y}(t) = -\omega^2 \sin \omega t - \epsilon \omega_n^2 \sin \omega_n t$ \quad S/N ratio $= 1/\epsilon \left(\frac{\omega}{\omega_n}\right)^2$
4.2 Observer structure

Instead we will use a state observer (Luenberger Observer) which contains a dynamic model of the system and whose state, $\hat{x}(t)$, approaches $x(t)$ as $t \to \infty$.

\[
\begin{align*}
\dot{\hat{x}} &= A\hat{x} + Bu + L(y - \hat{y}) \\
\hat{y} &= C\hat{x}
\end{align*}
\]
Consider the error \( e(t) = x(t) - \hat{x}(t) \)

\[
\dot{\hat{e}} = \dot{x} - \dot{\hat{x}} = (Ax + Bu) - (A\hat{x} + B\hat{u} + L(y - \hat{y}))
\]

\[
= A(x - \hat{x}) - LC(x - \hat{x}) = (A - LC)e
\]

We want \( e^{(A-LC)t} \to 0 \) quickly as \( t \) increases.

This is achieved if the eigenvalues of \( (A - LC) \) are large and negative, for example.

Can we assign the eigenvalues of \( (A - LC) \) by choice of \( L \)?

Suppose \( (A, C) \) is not observable then in section 3.2 we found a change of coordinates, \( \tilde{x} = T x \) such that,

\[
\begin{bmatrix}
\dot{\tilde{x}}_1 \\
\dot{\tilde{x}}_2
\end{bmatrix} = \begin{bmatrix}
\tilde{A}_{11} & 0 \\
\tilde{A}_{21} & \tilde{A}_{22}
\end{bmatrix} \begin{bmatrix}
\tilde{x}_1 \\
\tilde{x}_2
\end{bmatrix} + \begin{bmatrix}
\tilde{B}_1 \\
\tilde{B}_2
\end{bmatrix} u, \quad y = \begin{bmatrix}
\tilde{C}_1 & 0
\end{bmatrix} x + D u
\]

Hence

\[
T(A - LC)T^{-1} = \tilde{A} - \tilde{L}\tilde{C} = \begin{bmatrix}
\tilde{A}_{11} & 0 \\
\tilde{A}_{21} & \tilde{A}_{22}
\end{bmatrix} - \begin{bmatrix}
\tilde{L}_1 \\
\tilde{L}_2
\end{bmatrix} \begin{bmatrix}
\tilde{C}_1 & 0
\end{bmatrix} = \begin{bmatrix}
(\tilde{A}_{11} - \tilde{L}_1\tilde{C}_1) & 0 \\
(\tilde{A}_{21} - \tilde{L}_2\tilde{C}_1) & \tilde{A}_{22}
\end{bmatrix},
\]

and the eigenvalues of the observer,

\[
\lambda_i(A - LC) = \lambda_i(\tilde{A} - \tilde{L}\tilde{C}) = \lambda_i(\tilde{A}_{11} - \tilde{L}_1\tilde{C}_1) \cup \lambda_i(\tilde{A}_{22}),
\]

and \( \lambda_i(\tilde{A}_{22}) \) are not changed by \( \tilde{L} \).
However it can be shown that

We can arbitrarily assign the eigenvalues of \((A - LC)\) by choice of \(L\) if and only if the system is observable.

- We can thus make the error, \(e(t) \to 0\) arbitrarily quickly.
- But high gains might imply very large transient errors and noisy estimates.

4.3 Application in Ball and Beam experiment

Ball Position (BP) is measured by discrete sensors. We need Ball Velocity (BV). Differentiate? BP signal absent sometimes. Use observer.

Take \(\mathbf{x} = [BP, BV]^T\) and \(u = PP\) (Plank Position):

**PLANT:** \(\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ \gamma \end{bmatrix} u, \quad y = [1, 0]^T \mathbf{x}\)

**OBSERVER:** \(\dot{\hat{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \hat{x} + \begin{bmatrix} 0 \\ \gamma \end{bmatrix} u + \begin{bmatrix} \ell_1 \\ \ell_2 \end{bmatrix} (y - \hat{y})\)

**ERROR:** \(\dot{e} = (A - LC)e = \begin{bmatrix} -\ell_1 & 1 \\ -\ell_2 & 0 \end{bmatrix} e \quad \text{Eigenvalues: } s^2 + \ell_1 s + \ell_2 = 0\)

Observable? \[\begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\] rank = 2 ⇒ Yes.
Op. amps 1 and 2 implement $\dot{x} = A\dot{x} + Bu + L(y - \hat{y})$. (2 integrators in series.)

Op. amp 3, together with the 33k resistor on its output, implements gain $\ell_2$.

33k resistor on input of op.amp 2 implements $\ell_1$.

$\ell_1$ and $\ell_2$ chosen to give eigenvalues $-15 \pm 8.6j$

Op. amp 4 gives output = 0 if ball goes open-circuit — observer then predicts $BP$ and $BV$ without any data. (Clever circuit devised by Prof. Glover.)
4.4 Tracking disturbances, ignoring noise

Imagine tracking aircraft by radar (1-D). Aircraft position $z$ is affected by random turbulence.

Take $x = [z, \dot{z}]^T$:

$$\dot{x}(t) = Ax(t) + Bd(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} d(t)$$

The radar measurement is corrupted by noise:

$$y(t) = C\dot{x}(t) + n(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) + n(t)$$

Observer: $\ddot{x}(t) = A\dot{x}(t) + L[y(t) - C\dot{x}(t)]$  

NB: $d(t)$ not known, so not used.

$d$ large, $n$ small: Believe the measurements. Use large $L$. React quickly.

d small, $n$ large: Don’t trust measurements, believe model. Use small $L$.

— Smooth the measurements.

4.5 Special case: Kalman Filter

Suppose that $d$ and $n$ are both ‘white noise’.

Suppose we know their relative ‘sizes’: $\text{var}(d)$, $\text{var}(n)$.

How to design the observer gain $L$ optimally — minimise variance of tracking error $E\{\|x - \dot{x}\|^2\}$?

The solution is given by Kalman Filter theory — optimal trade-off between tracking $d$ and rejecting $n$. Guarantees stable observer.

Generalises to arbitrary dimension state vectors, multi-input, multi-output, and to arbitrary disturbance/noise spectra. Matlab: \texttt{kalman}, \texttt{dkalman}, \texttt{estim} etc.

Very widely used Navigation & guidance, Telecomms, Control, Finance, …

Especially in discrete time — software implementation.
4.6 Application to sensor fusion

Satellite, 1 axis of rotation: $J\ddot{\theta} = u + d$ (u = control torque, d = disturbance torque).

Two noisy sensors: Star sensor: $y_1 = \theta + n_\theta$, Rate gyro: $y_2 = \dot{\theta} + n_\omega$

Let $\mathbf{x} = [\theta, \dot{\theta}]^T$. State-space model:

\[
\dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1/J & 1/J \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix}
\]

\[
y = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} n_\theta \\ n_\omega \end{bmatrix} = I \mathbf{x} + \begin{bmatrix} n_\theta \\ n_\omega \end{bmatrix}
\]

Observable? Yes. ($C = I$, so rank $C = 2$, so rank $[C \ CA] = 2$.)

Observer:

\[
\dot{\hat{x}} = A\hat{x} + B \begin{bmatrix} u \\ 0 \end{bmatrix} + L(y - C\hat{x}) \quad (d \text{ not known})
\]

\[
= (A - LC)\hat{x} + \begin{bmatrix} 0 \\ 1/J \end{bmatrix} u + Ly \quad \text{but } C = I \text{ so:}
\]

\[
= \begin{bmatrix} -\ell_{11} & 1 - \ell_{12} \\ -\ell_{21} & -\ell_{22} \end{bmatrix} \hat{x} + \begin{bmatrix} 0 \\ 1/J \end{bmatrix} u + \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \end{bmatrix} y
\]

Place both eigenvalues at $-10$ (say): Using $\text{trace}(A - LC) = \sum \lambda_i$ and $\text{det}(A - LC) = \prod \lambda_i$:

$-\ell_{11} - \ell_{22} = -20$ and $\ell_{11}\ell_{22} + \ell_{21}(1 - \ell_{12}) = 100$. This leaves some design freedom.

$n_\theta \ll n_\omega$: Make $\ell_{11} \gg \ell_{12}$ and $\ell_{21} \gg \ell_{22}$.

$n_\theta \gg n_\omega$: Make $\ell_{11} \ll \ell_{12}$ and $\ell_{21} \ll \ell_{22}$.

Optimal trade-off: Kalman Filter again.
4.7 Application to sensor bias estimation

Satellite, as before: \( J\ddot{\theta} = u \)

Sensors: Star tracker measures angular position: \( y_1 = \theta \)
Rate gyro measures angular velocity with bias: \( y_2 = \dot{\theta} + b_\omega \).

Augment state vector: \( \mathbf{x} = [\theta, \dot{\theta}, b_\omega]^T \), and assume bias is constant: \( \dot{b}_\omega = 0 \).

State-space model:

\[
\begin{align*}
\dot{\mathbf{x}} &= \begin{bmatrix} 0 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\
1/J \\
0 \end{bmatrix} u \\
y &= \begin{bmatrix} 1 & 0 & 0 \\
0 & 1 & 1 \end{bmatrix} \mathbf{x} \\
\end{align*}
\]

Is the state observable?

\[
\begin{bmatrix} C \\
CA \\
CA^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\
0 & 1 & 1 \\
0 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \end{bmatrix}
\]

First 3 rows are linearly independent (Or: All three columns are linearly independent).

So rank = 3. Hence: **Observable**. So can use observer to estimate \( \hat{\mathbf{x}} \):

\[
\dot{\hat{\mathbf{x}}} = A\hat{\mathbf{x}} + Bu + L(y - C\hat{\mathbf{x}})
\]

\( A - LC \) stable \( \Rightarrow \dot{x}_3 \to b_\omega \) as \( t \to \infty \). Rate of convergence depends on eigenvalues of \( A - LC \).