Lecture 15
Observability
reconstructibility
estimability
controllability
in a stochastic context

Work done with Andrew Liu, PhD
What do we mean by observability?

Deterministic linear systems

\[
\begin{align*}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k
\end{align*}
\]
What do we mean by observability?

Deterministic linear systems

\[
\begin{align*}
x_{k+1} &= Ax_k + Bu_k \\
y_k    &= Cx_k
\end{align*}
\]

(Complete) Observability: measurements \( \{u_k, y_k: k=0,\ldots,n-1\} \) allow us to determine \( x_0 \) uniquely
What do we mean by observability?

**Deterministic linear systems**

\[
\begin{align*}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k
\end{align*}
\]

(Complete) Observability: measurements \(\{u_k,y_k: k=0,...,n-1\}\) allow us to determine \(x_0\) uniquely.
What do we mean by observability?

Deterministic linear systems

\[
\begin{align*}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k
\end{align*}
\]

(Complete) Observability: measurements \( \{u_k, y_k: k=0,\ldots,n-1\} \)
allow us to determine \( x_0 \) uniquely

(Complete) Reconstructibility: measurements \( \{u_k, y_k: k=0,\ldots,n-1\} \)
allow us to uniquely determine \( x_{n-1} \)
What do we mean by observability?

Deterministic linear systems

\[ \begin{align*}
x_{k+1} &= Ax_k + Bu_k \\
y_k &= Cx_k
\end{align*} \]

(Complete) Observability: measurements \( \{u_k, y_k: k=0, \ldots, n-1\} \) allow us to determine \( x_0 \) uniquely.

(Complete) Reconstructibility: measurements \( \{u_k, y_k: k=0, \ldots, n-1\} \) allow us to uniquely determine \( x_{n-1} \).

Property: the system above is completely observable iff \( O \) is full rank

\[ O = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix} \]
What do we mean by observability?

Deterministic linear systems

\[ x_{k+1} = Ax_k + Bu_k \]
\[ y_k = Cx_k \]

(Complete) Observability: measurements \( \{u_k,y_k: k=0,...,n-1\} \) allow us to determine \( x_0 \) uniquely

(Complete) Reconstructibility: measurements \( \{u_k,y_k: k=0,...,n-1\} \) allow us to uniquely determine \( x_{n-1} \)

Property: the system above is completely observable iff \( O \) is full rank

\[
O = \begin{pmatrix}
C \\
CA \\
\vdots \\
CA^{n-1}
\end{pmatrix}
\]

Property: the system above is completely reconstructible if \( \text{Range}(O_n) \supseteq \text{Range}(A^n) \)
The unique initial state is recoverable iff the rank of the observability matrix is full.

The effect of the control signal is additive of a known term.

Hence the control does not enter the observability condition.

Discrete-time and continuous-time have the same condition by Cayley-Hamilton.
Reconstructibility

More important for control than is observability

Since it involves knowing the current state

In LTI continuous time observability $\Leftrightarrow$ reconstructibility

In LTI discrete time observability $\Rightarrow$ reconstructibility

The deterministic LTI system

$$
x_{k+1} = Ax_k + Bu_k \\
y_k = Cx_k
$$

is reconstructible iff (PBH version)

If $\{Cw = 0 \text{ and } Aw = \lambda w\}$ then $\{\lambda = 0 \text{ or } w = 0\}$

In discrete time, singular $A$ matrices are the only special cases
Linear Stochastic Observability

Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]
Linear Stochastic Observability

Stochastic Linear Systems

\[ \begin{align*}
    x_{k+1} &= Ax_k + Bu_k + w_k \\
    y_k &= Cx_k + v_k
\end{align*} \]

Han-Fu Chen

If when the initial state covariance \( \Sigma_0 \rightarrow \infty I \) we have \( \Sigma_0|_{n-1} < \infty I \) then the system is (Chen) completely observable

Property: the stochastic linear system is completely observable if the observability matrix \( \mathcal{O} \) is full rank
Linear Stochastic Observability

Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]

Han-Fu Chen

If when the initial state covariance \( \Sigma_0 \rightarrow \infty I \) we have \( \Sigma_0|_{n-1} < \infty I \)
then the system is (Chen) completely observable

Property: the stochastic linear system is completely observable if the observability matrix \( \mathcal{O} \) is full rank

Baram & Kailath

The system is completely estimable if \( \Sigma_{n-1}|_{n-1} < \Sigma_{n-1}(\Sigma_0) \)

Compares effects of measurements on state covariance

Depends on the initial state density

Different from observability
Linear Stochastic Observability

Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]
Linear Stochastic Observability

Stochastic Linear Systems

\[
\begin{align*}
    x_{k+1} &= Ax_k + Bu_k + w_k \\
    y_k &= Cx_k + v_k
\end{align*}
\]

Definition: Consider the system above with zero-mean, gaussian noises \( w_k \) and \( v_k \). Assume the initial state \( x_0 \) has mean zero and covariance \( \Sigma_0 \). If for every vector \( \xi \) we have either \( \xi^T \Sigma_0^{-1} \xi = 0 \) or \( \xi^T \Sigma_0^{-1} \xi > 0 \) and \( \xi^T \Sigma_n^{-1} \xi < \xi^T \Sigma_0^{-1} \xi \), then the system is Linear Stochastically Completely Observable.
Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]

Definition: Consider the system above with zero-mean, gaussian noises \( w_k \) and \( v_k \). Assume the initial state \( x_0 \) has mean zero and covariance \( \Sigma_0|_{-1} \). If for every vector \( \xi \) we have either \( \xi^T \Sigma_0|_{-1} \xi = 0 \) or \( \xi^T \Sigma_0|_{-1} \xi > 0 \) and \( \xi^T \Sigma_0|_{n-1} \xi < \xi^T \Sigma_0|_{-1} \xi \), then the system is Linear Stochastically Completely Observable.

Theorem: Consider the system above with \( \text{cov}(w_k)=Q \) and \( \text{cov}(v_k)=R \). Then we have LSCO if \( R \) is finite, \( Q \) is finite if \( n>1 \) and \( \text{rank}(\mathcal{O})=n \).
Linear Stochastic Observability

Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]

Definition: Consider the system above with zero-mean, gaussian noises \( w_k \) and \( v_k \). Assume the initial state \( x_0 \) has mean zero and covariance \( \Sigma_0 \). If for every vector \( \xi \) we have either \( \xi^T \Sigma_0^{-1} \xi = 0 \) or \( \xi^T \Sigma_0^{-1} \xi > 0 \) and \( \xi^T \Sigma_0^{-1} \xi < \xi^T \Sigma_0^{-1} \xi \) then the system is Linear Stochastically Completely Observable.

Theorem: Consider the system above with \( \text{cov}(w_k) = Q \) and \( \text{cov}(v_k) = R \). Then we have LSCO if \( R \) is finite, \( Q \) is finite if \( n > 1 \) and \( \text{rank}(\mathcal{O}) = n \).

Complete observability focuses on estimation benefit from data

Distribution of quantity being estimated matters
Linear Stochastic Observability

Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]

Definition: Consider the system above with zero-mean, gaussian noises \( w_k \) and \( v_k \). Assume the initial state \( x_0 \) has mean zero and covariance \( \Sigma_0|_{-1} \). If for every vector \( \xi \) we have either \( \xi^T \Sigma_0|_{-1} \xi = 0 \) or \( \xi^T \Sigma_0|_{-1} \xi > 0 \) and \( \xi^T \Sigma_0|_{n-1} \xi < \xi^T \Sigma_0|_{-1} \xi \) then the system is Linear Stochasticely Completely Observable.

Theorem: Consider the system above with \( \text{cov}(w_k) = Q \) and \( \text{cov}(v_k) = R \). Then we have LSCO if \( R \) is finite, \( Q \) is finite if \( n > 1 \) and \( \text{rank}(O) = n \).

Complete observability focuses on estimation benefit from data.

Distribution of quantity being estimated matters.

\[ \Sigma_0|_{n-1} < \Sigma_0|_{-1} \]
Gaussian Systems - proving things

\[
\begin{pmatrix}
  y_0 \\
  y_1 \\
  y_2 \\
  \vdots \\
  y_{n-1}
\end{pmatrix} = \begin{pmatrix} C & CA & CA^2 & \cdots & CA^{n-1} \\
\end{pmatrix} x_0 + \begin{pmatrix}
  0 & 0 & \cdots & 0 \\
  CB & 0 & \cdots & 0 \\
  CAB & CB & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  CA^{n-1}B & CA^{n-3}B & \cdots & CB
\end{pmatrix} \begin{pmatrix}
  w_0 \\
  w_1 \\
  w_2 \\
  \vdots \\
  w_{n-1}
\end{pmatrix} + \begin{pmatrix}
  v_0 \\
  v_1 \\
  v_2 \\
  \vdots \\
  v_{n-1}
\end{pmatrix}
\]

Measurement vector and initial state are jointly gaussian
Simple formula for conditional mean and conditional variance for \( x_0 \)

\[
\Sigma_{0|n-1} = \Sigma_0 - \Sigma_0 \mathcal{O}^T (\mathcal{O} \Sigma_0 \mathcal{O}^T + \mathcal{H} \mathcal{Q} \mathcal{H}^T + \mathcal{R})^{-1} \mathcal{O} \Sigma_0
\]

Strict inequality requires

\[
\Sigma_{0|n-1} > 0 \quad \text{rank}(\mathcal{O}) = n \quad R < \infty \quad Q < \infty \quad (n > 1)
\]
Linear Stochastic Reconstructibility

Stochastic Linear Systems

\[ x_{k+1} = Ax_k + Bu_k + w_k \]
\[ y_k = Cx_k + v_k \]

Effectively: Reconstructible/estimable if

\[ \sum_{n-1}^{n-1} < \sum_{n-1}^{n-1} - 1 \]

Stochastic observability \( \not\Rightarrow \) stochastic reconstructibility

Strict inequality can be hard to achieve
Linear Stochastic Reconstructibility

Stochastic Linear Systems

\[
\begin{align*}
    x_{k+1} &= Ax_k + Bu_k + w_k \\
    y_k &= Cx_k + v_k
\end{align*}
\]

Effectively: Reconstructible/estimable if

\[
\Sigma_{n-1|n-1} < \Sigma_{n-1|-1}
\]

Stochastic observability \(\n\not\Rightarrow\) stochastic reconstructibility

Strict inequality can be hard to achieve

Non-gaussian Linear Stochastic Systems

Extension of observability and reconstructibility to not-necessarily gaussian linear systems is straightforward using properties of the Best Linear Unbiased Estimator
Extension to Nonlinear Stochastic Systems

\[ y = |x| \]

Is \( x \) observable from \( y \)?

Depends on the distribution of \( x \)

If the distribution is symmetric then \( \text{sgn}(x) \) is unobservable

Example 1: \( x = \begin{cases} -1, \text{ with probability } 1/2 \\ 1, \text{ with probability } 1/2 \end{cases} \) \( y \) unobservable

Example 2: \( x \sim \mathcal{N}(1, \sigma^2) \) \( y \) observable

Ideas of complete observability need rethinking

There is a need to consider the observability of functions of \( x \)

Variance might not be the correct quantity to consider

Especially for HMMs where we are reconstructing an estimate of the probability density

Use entropy \( H(x) \) and conditional entropy \( H(x|y) \)
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in Y} P(y = y_i) \sum_{x_j \in X} P(x = x_j | y = y_i) \ln P(x = x_j | y = y_i) \]
\[ = H(y, x) - H(x) \]
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in \mathcal{Y}} P(y = y_i) \sum_{x_j \in \mathcal{X}} P(x = x_j|y = y_i) \ln P(x = x_j|y = y_i) \]

\[ = H(y, x) - H(x) \]

Properties of conditional entropy
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in \mathcal{Y}} P(y = y_i) \sum_{x_j \in \mathcal{X}} P(x = x_j | y = y_i) \ln P(x = x_j | y = y_i) \]

\[ = H(y, x) - H(x) \]

Properties of conditional entropy

\[ H(x) \geq 0 \]
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in Y} P(y = y_i) \sum_{x_j \in X} P(x = x_j | y = y_i) \ln P(x = x_j | y = y_i) \]
\[ = H(y, x) - H(x) \]

Properties of conditional entropy

\[ H(x) \geq 0 \]
\[ H(g(x)) \leq H(x) \quad \text{with equality if } g(.) \text{ is injective} \]
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in Y} P(y = y_i) \sum_{x_j \in X} P(x = x_j | y = y_i) \ln P(x = x_j | y = y_i) \]
\[ = H(y, x) - H(x) \]

Properties of conditional entropy

\[ H(x) \geq 0 \]

\[ H(g(x)) \leq H(x) \quad \text{with equality if } g(.) \text{ is injective} \]

\[ H(x|y) \leq H(x) \quad \text{with equality iff } x \text{ and } y \text{ are independent} \]
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in Y} P(y = y_i) \sum_{x_j \in X} P(x = x_j|y = y_i) \ln P(x = x_j|y = y_i) \]

\[ = H(y, x) - H(x) \]

Properties of conditional entropy

\[ H(x) \geq 0 \]

\[ H(g(x)) \leq H(x) \text{ with equality if } g(.) \text{ is injective} \]

\[ H(x|y) \leq H(x) \text{ with equality iff } x \text{ and } y \text{ are independent} \]

\[ H(x) = \text{tr}(\Sigma) \text{ if } x \text{ is gaussian } N(\bar{x}, \Sigma) \]
Conditional Entropy

\[ H(x|y) = \sum_{y_i \in Y} P(y = y_i) \sum_{x_j \in X} P(x = x_j|y = y_i) \ln P(x = x_j|y = y_i) \]
\[ = H(y, x) - H(x) \]

Properties of conditional entropy

\[ H(x) \geq 0 \]
\[ H(g(x)) \leq H(x) \quad \text{with equality if } g(.) \text{ is injective} \]
\[ H(x|y) \leq H(x) \quad \text{with equality iff } x \text{ and } y \text{ are independent} \]
\[ H(x) = \text{tr}(\Sigma) \quad \text{if } x \text{ is gaussian } N(\bar{x}, \Sigma) \]

**Definition:** Random quantity \( x \) is completely observable from random quantity \( y \) if, for every, measurable function \( g(.) : \mathcal{X} \rightarrow \mathbb{R} \), either \( H(g(x)) = 0 \) or \( H(g(x)|y) < H(g(x)) \)
Observability via entropy

Random vector $x$ is unobservable from random vector $y$ if $x$ is independent from $y$. It is observable if it is not unobservable.

Equivalently, $x$ is observable from $y$ if either $H(x) = 0$ or $H(x|y) < H(x)$.

Random vector $x$ is completely observable from random vector $y$ if, for every scalar measurable function $g: X \rightarrow \mathbb{R}$, $g(x)$ is observable from $y$.

That is, either $H(g(x)) = 0$ or $H(g(x)|y) < H(g(x))$.

System with initial state, state distribution, state entropy

$x_{k+1} = f(x_k, u_k, w_k)$
$y_k = h(x_k, u_k, v_k)$

Compute state entropy $H(x_m)$ at time $m$ given control $[u_0, u_1, \ldots, u_{m-1}]$.

If $x_m$ is completely observable from $[y_0, y_1, \ldots, y_m]$ then the system is completely stochastically reconstructible from this initial condition and with this control sequence.
Observability and reconstructibility tests

If the values taken by the state and by the output of a stochastic systems take on only finitely many values then the tests for complete observability and complete reconstructibility are finite.

The observability and reconstructibility test involve evaluating the entropy for any measurable function $g(.)$.

For finite state and output sets the number of sets of equivalent functions is finite because $H(g(x)) = H(x)$ if $g$ is injective.

While finite, the number of tests could be huge.
Reconstructibility and optimal control

Suppose the state of a markov system is not reconstructible from the set of preceding inputs and outputs no matter which sequence of inputs is applied. Suppose we seek to solve a stochastic optimal control problem with incremental reward function $\ell(x_k, u_k)$.

Then the optimal open-loop cost function is the same as the optimal closed-loop cost function

Furthermore the optimal control sequences are identical for open or closed loop

Suppose the incremental reward function value $\ell(x_k, u_k) = \ell(x_k)$ is not reconstructible i.e. $H(\ell(x_k) | \{r_t\}_0^k, \{y_t\}_0^k) = H(\ell(x_k) | \{r_t\}_0^k)$

Then the same conclusion holds
Conclusion

We have moved from a deterministic definition of observability and reconstructibility to a stochastic one.

This includes nonlinear systems, which themselves are still an issue in deterministic systems theory.

The core idea has been to use the entropy properties of measurable functions of the state.

The stochastic definition differs subtly from the deterministic one.

If we know the initial state, then it is observable regardless of the dynamics and readout map.

The linkage between reconstructibility and optimal feedback control making an improvement over open-loop is a new twist.

The proof relies on stochastic dynamic programming ideas.