# Control of Stochastic Systems
# Lecture 3
# Stochastic Systems

Jan H. van Schuppen

27 February 2025

Delft University of Technology

# Outline

# Outline

# Example. Paper Machine

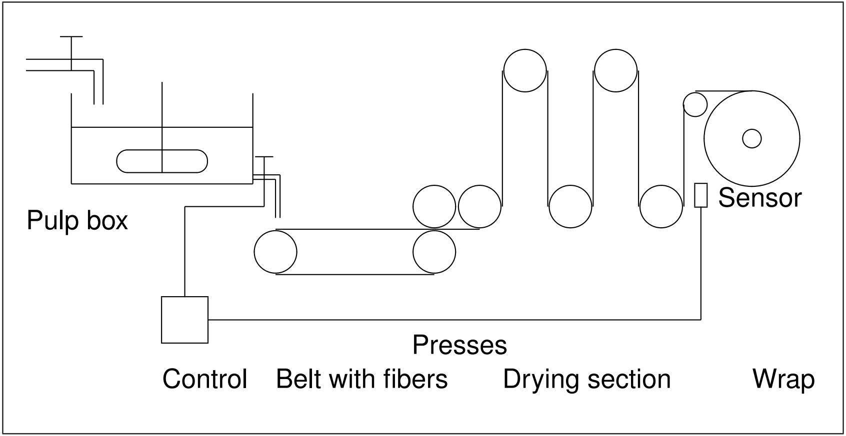
### Project. Control of a paper machine

Billerud Kraft Paper Mill, at Gruvön, Sweden.
Aim of project:
Establish that computers can carry out control of a paper mill and achieve the control objectives.
K.J. Aström carried out the project for the IBM company.

### Control objectives

Reduce the standard deviation of dry paper weight,
from 1.3 $g/m^2$ to 0.7 $g/m^2$.
The control objective of reduction of the standard deviation
(or reduction of the variance)
often occurs in the research area called process control.

# Example. Paper Machine - A diagram of a paper machine



# Example. Paper Machine

## Operation of paper machine

- Wood goes into the machine chest, with chemicals and water.
  After stirring this becomes a homogeneous solution called pulp.
- From machine chest flows the thick stock flow, onto a belt with fibers.
  After presses, which remove water, a paper web is formed.
- Steam-heated cylinders dry the paper.
  Paper is wrapped on another cylinder.

## Inputs and outputs

Inputs:
(1) Thick stock flow amount and consistency.
(2) Drying activity.
Outputs:
(1) Wet paper weight by beta-ray gauge at end of drying section.
(2) Water contents of paper.
(3) Estimate of dry paper weight.

# Example. Paper Machine

## Control system

- Variables: output $y$ dry basis weight and input $u$ thick stock flow.
- Dynamics: only the transporation delay
  from the thick stock flow to the sensor for wet paper weight.
- Control system specified as an ARMAX representation,

$$y(t) = \sum_{i=1}^{n_y} a_i \, y(t-i) + \sum_{j=0}^{n_u} b_j \, u(t-j-k) + \sum_{m=0}^{n_v} c_m \, v(t-m),$$

$k \in \mathbb{Z}_+$ models transportation delay between input and sensor.

- $v$ is Gaussian white noise.
- Gaussian stochastic control system (defined in Lecture 7)
  includes the above defined ARMAX representation.

# Example. Paper Machine

## Investigation phases

1. Modeling, described above.
2. System identification. Maximum likelihood method.
   Estimate the parameteres of the ARMAX representation,
   $\{a_i,\ b_j,\ c_m,\ \forall\ i,\ j,\ m\}$.
3. Control synthesis and control design.
   Design a control law $g$.
   Then the input is $u(t) = g(x(t))$.
4. Test the control law on the actual paper machine;
   possibly adjust the control law.

# Example. Paper Machine

### Investigation outcome

1.3 $g/m^2$ dry paper weight before use of control,
0.7 $g/m^2$ dry paper weight control objective,
0.3 $g/m^2$ dry paper weight with control.

### Comments

- Process control can benefit from control engineering.
- Minimum variance control is much used in control engineering.
- Modeling of the control system and feedback control can be quite effective.
  Modeling takes about 2/3 to 3/4 of the duration of the project.
- There may be a saving of material used in case of minimum variance control.

# Outline

# Concept Stochastic System

### Def. Stochastic system

Define a (discrete-time) stochastic system as a collection satisfying,

$$(F_t^{y+} \vee F_t^{x+},\ F_{t-1}^{y-} \vee F_t^{x-} \mid F^{x(t)}) \in CI,\ \forall\, t \in T;$$

where,

$(\Omega,\ F,\ P),$ complete probability space,

$T \subseteq \mathbb{Z},$ time index set,

$(Y,\ B_Y),$ output space, $(X,\ B_X),$ state space,

$y : \Omega \times T \to Y,$ output process, is observed,

$x : \Omega \times T \to X,$ state process, in general not observed,

$$F_t^{x-} = \sigma(\{x(s),\ \forall\, s \leq t\}),\quad F_{t-1}^{y-} = \sigma(\{y(s),\ \forall\, s \leq t-1\}),$$

$$F_t^{x+} = \sigma(\{x(s),\ \forall\, s \geq t\}),\quad F_t^{y+} = \sigma(\{y(s),\ \forall\, s \geq t\}),$$

$$F_{-1}^{y-} = \{\Omega,\ \emptyset\};$$

$\{\Omega,\ F,\ P,\ T,\ Y,\ B_Y,\ X,\ B_X,\ y,\ x\},$ notation.

# Concept Stochastic System

## Comments on definition of a stochastic system

- Definition in terms of conditional independence relation.
- Definition specifies that: the future and the past of the combination of the output process and of the state process are conditionally independent conditioned on the current state.
- Definition implies that the state process is a Markov process.
- Definition does not impose a restriction on the probability distributions of the state and of the output processes.
- With respect to the future and the past,
  - symmetry in $F_t^{x+}$ and $F_t^{x-}$
  because $F^{x(t)} \subset F_t^{x+}$ and $F^{x(t)} \subset F_t^{x-}$, and
  - asymmetry in $F_t^{y+}$ and $F_{t-1}^{y-}$
  because $F^{y(t)} \subset F_t^{y+}$ and $F^{y(t)} \not\subset F_{t-1}^{y-}$.
  This is a choice motivated by the representation to be presented shortly.
- Definition of a stochastic control system in Lecture 7.

# Concept Stochastic System

## Alternative conditions

$$
\begin{aligned}
& (F_t^{y+} \vee F_t^{x+},\ F_{t-1}^{y-} \vee F_t^{x-} \mid F^{x(t)}) \in CI,\ \forall\ t \in T; \\
\Leftrightarrow\quad & \text{the future conditioned on the past} \\
& \text{equals the future conditioned on the present state;} \\
\Leftrightarrow\quad & (F^{y(t)} \vee F^{x(t+1)},\ F_{t-1}^{y-} \vee F_t^{x-} \mid F^{x(t)}) \in CI,\ \forall\ t \in T; \\
& \text{equivalently,} \\
\Leftrightarrow\ & E[z^+ \mid F_{t-1}^{y-} \vee F_t^{x-}] = E[z^+ \mid F^{x(t)}], \\
& \forall\ z^+ \in L(\Omega,\ F_t^{y+} \vee F_t^{x+},\ \mathbb{R}^+), \\
\Leftrightarrow\ & E[a^+ \mid F_{t-1}^{y-} \vee F_t^{x-}] = E[a^+ \mid F^{x(t)}], \\
& \forall\ a^+ \in L(\Omega,\ F^{x(t+1)} \vee F^{y(t)},\ \mathbb{R}^+).
\end{aligned}
$$

See Proposition 5.10.1 of the lecture notes and of the book.

# Concept Stochastic System

**Def. Stochastic system (continued)**

Call a stochastic system:

**(a)** a stationary system
if the joint state and output process $(x,\ y)$
is a jointly stationary process;

**(b)** a Gaussian system
if the state and output processes, $(x,\ y)$
is a jointly Gaussian process;

**(c)** a finite stochastic system
if both the state process and the output process
are finite-valued processes.

# Outline

# Gaussian Systems Representations

**Def. Gaussian system representation (1)**

Define a time-varying Gaussian system representation
(in discrete-time and in a forward representation)
if the output and the state process
are specified by the system equations,

$$\begin{aligned} x(t+1) &= A(t)x(t) + M(t)v(t),\ x(0) = x_0, \\ y(t) &= C(t)x(t) + N(t)v(t); \end{aligned}$$

where the objects are described by the collection,

$$\left\{ \begin{array}{l} \Omega,\ F,\ P,\ T,\ \mathbb{R}^{n_y},\ B(\mathbb{R}^{n_y}),\ \mathbb{R}^{n_x},\ B(\mathbb{R}^{n_x}), \\ y,\ x,\ v,\ x_0,\ A,\ C,\ M,\ N \end{array} \right\}.$$

(continued on next slide)

# Gaussian Systems Representations

## Def. Gaussian system representation (2)

where the objects are described by,

$$
\begin{aligned}
&(\Omega,\ F,\ P),\ \text{a complete probability space,}\\
T = \mathbb{N} = \{0,\ 1,\ 2,\ \ldots\},\ &\text{time index set,}\\
&x_0 : \Omega \rightarrow X = \mathbb{R}^{n_x},\ x_0 \in G(m_{x_0},\ Q_{x_0}),\\
&v : \Omega \times T \rightarrow \mathbb{R}^{n_v},\ \text{standard Gaussian white noise,}\\
&v(t) \in G(0,\ I_{n_v}),\ \forall\ t \in T,\\
&F^{x_0},\ F^{v}_{\infty}\ \text{independent}\ \sigma\text{-algebras,}\\
&A : T \rightarrow \mathbb{R}^{n_x \times n_x},\ C : T \rightarrow \mathbb{R}^{n_y \times n_x},\ M : T \rightarrow \mathbb{R}^{n_x \times n_v},\ N : T \rightarrow \mathbb{R}^{n_y \times n_v},\\
&(\mathbb{R}^{n_y},\ B(\mathbb{R}^{n_y})),\ \text{output space,}\ (\mathbb{R}^{n_x},\ B(\mathbb{R}^{n_x})),\ \text{state space,}\\
&y : \Omega \times T \rightarrow \mathbb{R}^{n_y},\ \text{output process,}\\
&x : \Omega \times T \rightarrow \mathbb{R}^{n_x},\ \text{state process.}
\end{aligned}
$$

(continued on next slide)

# Gaussian Systems Representations

**Def. Gaussian system representation (3)**

Define a
time-invariant Gaussian system representation,

$$
\begin{aligned}
& \text{if } \forall\ t \in T, \\
A(t) &= A(0),\ C(t) = C(0),\ M(t) = M(0),\ N(t) = N(0);\ \text{define,} \\
& A \in \mathbb{R}^{n_x \times n_x},\ C \in \mathbb{R}^{n_y \times n_x},\ M \in \mathbb{R}^{n_x \times n_v},\ N \in \mathbb{R}^{n_y \times n_v}, \\
A &= A(0),\ \text{etc.}; \\
x(t+1) &= Ax(t) + Mv(t),\ x(0) = x_0, \\
y(t) &= Cx(t) + Nv(t); \\
& \text{where } v(t) \in G(0,\ I_{n_v}),\ x_0 \in G(m_{x_0},\ Q_{x_0}); \\
& \text{notation } (n_y,\ n_x,\ n_v,\ A,\ C,\ M,\ N).
\end{aligned}
$$

See Lecture 4 for
(1) properties of time-invariant Gaussian systems,
(2) specific conditions to be used.

# Gaussian Systems Representations

## Comments. Gaussian system representation

In the literature one finds a Gaussian system representation of the form,

$$\begin{aligned} x(t+1) &= A(t)x(t) + M_1(t)r(t),\ x(0) = x_0, \\ y(t) &= C(t)x(t) + N_2(t)w(t), \\ & r : \Omega \times T \to \mathbb{R}^{n_r},\ w : \Omega \times T \to \mathbb{R}^{n_w}, \end{aligned}$$

where $r$ and $w$ are each a standard Gaussian white noise process, and $r$ and $w$ are independent ($\Leftrightarrow$ $F^r_\infty$ and $F^w_\infty$ are independent). Apply a transformation, $v : \Omega \times T \to \mathbb{R}^{n_r+n_w}$,

$$\begin{aligned} v(t) &= \begin{pmatrix} r(t) \\ w(t) \end{pmatrix},\ Q_v(t) = \begin{pmatrix} I_{n_r} & 0 \\ 0 & I_{n_w} \end{pmatrix} = I_{n_r+n_w}, \\ M(t) &= \begin{pmatrix} M_1(t) & 0 \end{pmatrix},\ N(t) = \begin{pmatrix} 0 & N_2(t) \end{pmatrix}, \\ &\Rightarrow\ M(t)v(t) = M_1(t)r(t),\ N(t)v(t) = N_2(t)w(t). \end{aligned}$$

Hence one obtains a Gaussian system representation.

# Gaussian Systems Representations

## Comments. Gaussian system representation

In the literature one finds a Gaussian system representation of the form,

$$\begin{aligned} x(t+1) &= A(t)x(t) + M(t)v(t),\ x(0) = x_0, \\ y(t+1) &= C(t)x(t) + N(t)v(t). \end{aligned}$$

In the literature,
there are results for the system representation with $y(t+1)$.

The above system representation differs
from the system representation of the Def. stated on a previous slide,

$$\begin{aligned} x(t+1) &= A(t)x(t) + M(t)v(t),\ x(0) = x_0, \\ y(t) &= C(t)x(t) + N(t)v(t). \end{aligned}$$

These two system representations cannot be converted into each other.
Choice in this lecture based on the convention of system theory.

# Gaussian Systems Representations

### Def. State-output conditional independence

Consider a Gaussian system representation.
This system representation is called
state-output conditionally independent if,

$$(F^{x(t+1)},\ F^{y(t)} |\ F_t^x \vee F_{t-1}^y) \in CI,\ \forall\ t \in T.$$

### Proposition 4.3.3. State-output conditional independence

A Gaussian system representation is state-output conditionally independent if and only if,

$$0 = M(t)N(t)^T = E[M(t)v(t)\ (N(t)v(t))^T],\ \forall\ t \in T.$$

In the literature, attention often restricted to
the state-output conditional independent case
by assuming that the state and the output noise terms are independent.

# Gaussian Systems Representations

**Def. State-transition function**
Consider a Gaussian system representation.
Define the state-transition function,

$$\Phi(t+1,s) = \begin{cases} A(t)\Phi(t,s), & s < t+1, \\ I_{n_x}, & s = t+1, \\ 0, & s > t+1; \end{cases}$$

$$\Phi : T \times T \to \mathbb{R}^{n_x \times n_x};\ \text{then,}$$

$$\Phi(t,s) = A(t-1)A(t-2)\ldots A(s),\ s < t.$$

# Gaussian Systems Representations

## Theorem. Properties of a Gaussian system representation (1)

Consider a Gaussian system representation,

$$\begin{aligned} x(t+1) &= A(t)x(t) + M(t)v(t),\ x(0) = x_0, \\ y(t) &= C(t)x(t) + N(t)v(t). \end{aligned}$$

(a) Independence of the $\sigma$-algebras,

$$F_t^{v+} = \sigma(\{v(s),\ \forall\, s \geq t\}),\ F_t^x \vee F_{t-1}^y,\ \forall\, t \in T.$$

Proof.

$$\begin{aligned} & F_t^{v+},\ F_{t-1}^{v-},\ \text{independent}\ \forall\, t \in T,\ \text{by } v \text{ Gaussian white noise;} \\ & F^{x_0},\ F_\infty^v,\ \text{independent by assumption,} \\ \Rightarrow\quad & F_t^{v+},\ F^{x_0} \vee F_{t-1}^{v-}, \text{independent,} \\ & (F_t^x \vee F_{t-1}^y) \subseteq (F^{x_0} \vee F_{t-1}^v), \forall\, t \in T,\ \text{by induction,} \\ \Rightarrow\quad & F_t^{v+},\ F_t^x \vee F_{t-1}^y,\ \text{independent}\ \forall\, t \in T. \end{aligned}$$

# Gaussian Systems Representations

**Theorem. Properties of a Gaussian system representation (2)**

(b) Explicit expressions for $s,\ t \in T,\ s < t$,

$$x(t) = \Phi(t,s)x(s) + \sum_{r=s}^{t-1} \Phi(t-1,r)M(r)v(r),$$

$$y(t) = C(t)\Phi(t,s)x(s) + \sum_{r=s}^{t-1} C(t)\Phi(t-1,r)M(r)v(r) + N(t)v(t).$$

Proof. By induction.

$$x(t) = A(t-1)x(t-1) + M(t-1)v(t-1),$$
$$y(t) = C(t)x(t) + N(t)v(t),\ \text{etc.}$$

# Gaussian Systems Representations

**Theorem. Properties of a Gaussian system representation (3)**

(c) The process $(x,\ y)$ is a jointly Gaussian process.
Proof.

$$\begin{array}{ll}
(1) & x_0 \in G,\ v \text{ Gaussian white noise process}, \\
& F^{x_0},\ F^v_\infty,\ \text{independent}, \\
\Rightarrow & x_0,\ v(0),\ v(1), \ldots, \\
& \text{are independent jointly Gaussian random variables;} \\
(2) & \forall\ \ t_0 = 0,\ t_1,\ t_2,\ \ldots,\ t_k \in T, \\
& (x(0),\ x(t_1),\ x(t_2),\ \ldots,\ x(t_k),\ y(0),\ y(t_1),\ \ldots,\ y(t_k)), \\
& \text{are jointly Gaussian random variables,} \\
& \text{because by (b) every element is a linear function} \\
& \text{of the same jointly Gaussian random variables,} \\
& x_0,\ v(0),\ v(1), \ldots,\ v(t_k).
\end{array}$$

# Gaussian Systems Representations

### Theorem. Properties of a Gaussian system representation (4)

(d) A Gaussian system representation defines a stochastic system.
Proof.

$$
\begin{aligned}
& E\left[\exp\left(i\begin{bmatrix} w_x \\ w_y \end{bmatrix}^T \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}\right) \mid F_t^x \vee F_{t-1}^y\right] \\
&= \exp\left(i\begin{bmatrix} w_x \\ w_y \end{bmatrix}^T \begin{bmatrix} A(t) \\ C(t) \end{bmatrix} x(t)\right) \; E\left[\exp\left(i\begin{bmatrix} w_x \\ w_y \end{bmatrix}^T \begin{bmatrix} M(t) \\ N(t) \end{bmatrix} v(t)\right) \mid F_t^x \vee F_{t-1}^y\right] \\
&= \exp\left(i\begin{bmatrix} w_x \\ w_y \end{bmatrix}^T \begin{bmatrix} A(t) \\ C(t) \end{bmatrix} x(t) - \frac{1}{2}\begin{bmatrix} w_x \\ w_y \end{bmatrix}^T \begin{bmatrix} M(t)M(t)^T & M(t)N(t)^T \\ N(t)M(t)^T & N(t)N(t)^T \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix}\right) \\
&= E\left[\exp\left(i\begin{bmatrix} w_x \\ w_y \end{bmatrix}^T \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}\right) \mid F^{x(t)}\right], \; \forall\, (w_x,\ w_y) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_y}.
\end{aligned}
$$

Last step by a property of conditional expectation.

# Gaussian Systems Representations

## Theorem. Properties of a Gaussian system representation (5)

(e) The state process is a Gauss-Markov process with,

$$
\begin{aligned}
x(t) &\in G(m_x(t),\ Q_x(t)), \\
m_x(t) &= E[x(t)],\ m_x : T \to \mathbb{R}^{n_x}, \\
Q_x(t) &= E[(x(t) - m_x(t))(x(t) - m_x(t))^T],\ Q_x : T \to \mathbb{R}^{n_x \times n_x}, \\
m_x(t+1) &= A(t)m_x(t),\ m_x(0) = m_{x_0}, \\
Q_x(t+1) &= A(t)Q_x(t)A(t)^T + M(t)M(t)^T,\ Q_x(0) = Q_{x_0}, \\
W_x(t,s) &= E[(x(t) - m_x(t))(x(s) - m_x(s))^T] \\
&= \begin{cases} Q_x(t), & s = t, \\ \Phi(t,s)Q_x(s), & s < t, \\ Q_x(t)\Phi(s,t)^T, & s > t. \end{cases}
\end{aligned}
$$

$m_x$ called the mean value function of $x$,

$Q_x$ called the variance function of $x$,

$W_x$ called the covariance function of $x$.

# Gaussian Systems Representations

**Theorem. Properties Gaussian system representation (6)**

(e) Proof.

$$\begin{aligned}
m_x(t+1) &= E[x(t+1)] = E[A(t)x(t) + M(t)v(t)] = A(t)m_x(t), \\
Q_x(t+1) &= E[(x(t+1) - m_x(t+1))\,(x(t+1) - m_x(t+1))^T] \\
&= E[(A(t)(x(t) - m_x(t)) + M(t)v(t))\,(\ldots)^T] \\
&= A(t)Q_x(t)A(t)^T + M(t)M(t)^T; \\
x(t) - m_x(t) &= \Phi(t,s)(x(s) - m_x(s)) + \sum_{r=s}^{t-1} \Phi(t-1,r)M(r)v(r), \\
W_x(t,s) &= E[(x(t) - m_x(t))(x(s) - m_x(s))^T] \\
&= \Phi(t,s)W_x(s,s) = \Phi(t,s)Q_x(s), \text{ if } s < t;
\end{aligned}$$

continued;

# Gaussian Systems Representations

**Theorem. Properties Gaussian system representation (7)**

(e) Proof.

$$
\begin{aligned}
& E[\exp(iw^T x(t+1))|\ F_t^x] \\
& = \exp(iw^T A(t)x(t)))E[E[\exp(iw^T M(t)v(t))|\ F_t^x \vee F_{t-1}^v]|\ F_t^x] \\
& = \exp(iw^T A(t)x(t) - \frac{1}{2} w^T M(t)M(t)^T w) \\
& = E[\exp(iw^T x(t+1))|\ F^{x(t)}];
\end{aligned}
$$

hence $x$ is a Markov process.

# Gaussian Systems Representations

## Theorem. Properties Gaussian system representation (8)

(f) Output process is a Gaussian process,

$$
\begin{aligned}
y(t) &\in G(m_y(t),\ Q_y(t)), \\
m_y(t) &= E[y(t)] = C(t)m_x(t), \\
Q_y(t) &= C(t)Q_x(t)C(t)^T + N(t)N(t)^T, \\
W_y(t,s) &= E[(y(t)-m_y(t))(y(s)-m_y(s))^T] \\
&= \begin{cases} Q_y(t), \text{ if } s = t, \\ C(t)\Phi(t,s)Q_x(s)C(s)^T + C(t)\Phi(t-1,s)M(s)N(s)^T, \\ \qquad \text{if } s < t, \end{cases} \\
Q_{x+,y}(t) &= E[(x(t+1)-m_x(t+1))(y(t)-m_y(t))^T] \\
&= A(t)Q_x(t)C(t)^T + M(t)N(t)^T; \\
y(t)-m_y(t) &= C(t)\phi(t,s)(x(s)-m_x(s)) + \sum \ldots + N(t)v(t), \\
y(s)-m_y(s) &= C(s)(x(s)-m_x(s)) + N(s)v(s).
\end{aligned}
$$

Proof. See book, similar to proof of (e).

# Outline

# Gaussian Systems Representations

**Def. Backward Gaussian system representation**

Define a backward Gaussian system representation by the formulas,

$$\begin{aligned} x(t-1) &= A_b(t)x(t) + M_b(t)v_b(t),\ x(0) = x_0, \\ y(t-1) &= C_b(t)x(t) + N_b(t)v_b(t), \\ T &= \{0,\ -1,\ -2,\ \ldots\}. \end{aligned}$$

See book for corresponding result on probability distributions of $(x,\ y)$.
Backward representations needed in stochastic realization theory.

# Forward and Backward Representations

## Theorem. Forward and backward representations (1)

Consider a Gaussian stochastic system with $(x,\ y)$. Assume,

$$\begin{aligned} 0 &= E[x(t)],\ 0 = E[y(t)],\ \forall\ t \in T, \\ Q_x(t) &= E[x(t)x(t)^T] \succ 0,\ \forall\ t \in T. \end{aligned}$$

(a) There exists a forward Gaussian system representation of the form,

$$\begin{aligned} x(t+1) &= A_f(t)x(t) + M_f v_f(t),\ x(0) = x_0, \\ y(t) &= C_f(t)x(t) + N_f v_f(t), \\ A_f(t) &= E[x(t+1)x(t)^T]Q_x(t)^{-1},\ C_f(t) = E[y(t)x(t)^T]Q_x(t)^{-1}, \\ v_f(t) &= \begin{pmatrix} x(t+1) - A_f(t)x(t) \\ y(t) - C_f(t)x(t) \end{pmatrix},\ v_f : \Omega \times T \to \mathbb{R}^{n_x+n_y}, \\ & v_f \text{ Gaussian white noise, } v_f(t) \in G(0, Q_{v_f}(t)), \\ & F^{x_0},\ F_t^{v_f},\ \text{independent}, \forall\ t \in T, \\ M_f &= \begin{pmatrix} I_{n_x} & 0 \end{pmatrix},\ N_f = \begin{pmatrix} 0 & I_{n_y} \end{pmatrix}. \end{aligned}$$

# Forward and Backward Representations

### Theorem. Forward and backward representations (2)

(b) There exists a backward Gaussian representation of the form,

$$
\begin{aligned}
x(t-1) &= A_b(t)x(t) + M_b v_b(t), \; x(0) = x_0, \\
y(t-1) &= C_b(t)x(t) + N_b v_b(t), \\
A_b(t) &= E[x(t-1)x(t)^T]Q_x(t)^{-1}, \\
C_b(t) &= E[y(t-1)x(t)^T]Q_x(t)^{-1}, \\
v_b(t) &= \begin{pmatrix} x(t-1) - A_b(t)x(t) \\ y(t-1) - C_b(t)x(t) \end{pmatrix}, \; v_b : \Omega \times T \to \mathbb{R}^{n_x+n_y}, \\
& v_b \text{ Gaussian white noise, } v_b(t) \in G(0, Q_{v_f}(t)), \\
& F^{x_0}, \; F^{v_b}_t, \text{ independent, } \forall \; t \in T, \\
M_b &= \begin{pmatrix} I_{n_x} & 0 \end{pmatrix}, \; N_b = \begin{pmatrix} 0 & I_{n_y} \end{pmatrix}.
\end{aligned}
$$

# Forward and Backward Representations

**Theorem. Forward and backward representations (3)**

(c) Relations between forward and backward Gaussian representations.

$$
\begin{aligned}
A_f(t)Q_x(t) &= Q_x(t+1)A_b(t+1)^T, \\
C_b(t)Q_x(t) &= C_f(t-1)Q_x(t-1)A_f(t-1)^T + N_b Q_{v_f}(t-1)M_b^T, \\
C_f(t)Q_x(t) &= C_b(t+1)Q_x(t+1)A_b(t+1)^T + N_b Q_{v_f}(t+1)M_b^T.
\end{aligned}
$$

# Forward and Backward Representations

### Proof. Forward and backward representations

(a) Because by assumption there is a Gaussian system with the (state, output) processes $(x,\ y)$, one obtains,

$$E\left[\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} \mid F_t^x \vee F_{t-1}^y\right] = E\left[\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} \mid F^{x(t)}\right]$$

$$= \begin{bmatrix} A_f(t) \\ C_f(t) \end{bmatrix} x(t), \quad \text{by } (x(t+1),\ y(t),\ x(t)) \in G,$$

by conditional expectation of Gaussian rvs (Theorem 2.8.3.(a)),

$$E\left[\exp\left(iw^T \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}\right) \mid F_t^x \vee F_{t-1}^y\right]$$

$$= \exp\left(iw^T \begin{bmatrix} A_f(t) \\ C_f(t) \end{bmatrix} x(t) - \frac{1}{2} w^T Q_{v_f}(t) w\right),$$

by a property of conditional expectation (Theorem 2.8.3.(c)).

# Forward and Backward Representations

### Proof. Forward and backward representations (continued)

(a)

$$
\begin{aligned}
& E[\exp(iw^T v_f(t)) |\ F_{t-1}^{v_f}] \\
&= E\left[ E\left[ \exp\left( iw^T \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} \right) \mid F_t^x \vee F_{t-1}^y \right] \times \right. \\
&\qquad \left. \times \exp\left( iw^T \begin{bmatrix} -A_f(t) \\ -C_f(t) \end{bmatrix} x(t) \right) \mid F_{t-1}^{v_f} \right] \\
&= \exp\left( -\frac{1}{2} w^T Q_{v_f}(t) w \right) = E[\exp(iw^T v_f(t))],
\end{aligned}
$$

hence $v_f(t)$ is independent of $F_{t-1}^{v_f}$ for all $t \in T$,
$v_f(t) \in G(0, Q_{v_f}(t))$ for all $t \in T$,
hence $v_f$ is a Gaussian white noise process.
(b) This proof is similar to that of (a).
(c) This follows directly from (a) and (b).

# Outline

# Observability

## Motivation of observability

- Observability is a major concept
  of system theory and of control theory.
- It is used as a condition for observers and
  for control with partial observations.
- Observability of a linear deterministic system and
  of a nonlinear deterministic system have been defined.
  These are used often in control theory.
- Equivalent conditions for observability have been proven
  which can be checked by computations.
- Needed for a stochastic system:
  What is a definition of stochastic observability?
- Below discussed first,
  observability of a linear deterministic system.

# Observability

### Def. Injectivity of a function

Call a function $h$ injective if,

$$h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y},\ n_x,\ n_y \in \mathbb{Z}_+;$$
$$\forall\ x_a,\ x_b \in \mathbb{R}^{n_x},\ h(x_a) = h(x_b)\ \Rightarrow\ x_a = x_b.$$

### Remark

- Think of $y = h(x)$ with $x$ a state and $y$ an observation.
- A direct consequence of this definition is that:
  if $h$ is injective
  then from $y_a = h(x_a)$ one can uniquely determine $x_a$.
- Concept of an injective function
  is from the area algebra of sets.

# Observability

**Proposition. Characterization**

Consider the linear observation function

$$h(x) = Cx, \quad C \in \mathbb{R}^{n_y \times n_x}; \quad \text{define},$$
$$\ker(C) = \{x_a \in \mathbb{R}^{n_x} |\ Cx_a = 0\} = \mathrm{N}(C),$$
$$\mathrm{Im}(C) = \{Cx_b \in \mathbb{R}^{n_y} |\ \forall\ x_b \in \mathbb{R}^{n_x}\} = \mathrm{Range}(C).$$

Call $\ker(C)$ the kernel of $C$ or the null space of $C$, and call $\mathrm{Im}(C)$ the image of $C$ or the range space of $C$.

Equivalence:

**(a)** The function $h(x) = Cx$ is injective.

**(b)** $\forall\ x_a \in \mathbb{R}^{n_x}$, $Cx_a = 0 \ \Rightarrow\ x_a = 0$.

**(c)** $\ker(C) \subseteq \{0\} \Leftrightarrow \ker(C) = \{0\}$.

**(d)** $\mathrm{rank}(C) = n_x$.

# Observability

## Proof of proposition

$$
\begin{aligned}
h(x) = Cx \text{ is injective} \\
&\Leftrightarrow\ \forall\, x_a,\ x_b \in X,\ Cx_a = Cx_b\ \Rightarrow x_a = x_b, \\
&\Leftrightarrow\ \forall\, x_a,\ x_b \in X,\ C(x_a - x_b) = 0\ \Rightarrow x_a - x_b = 0, \\
&\Leftrightarrow\ \forall\, x_a \in X,\ Cx_a = 0\ \Rightarrow x_a = 0 \\
&\Leftrightarrow\ \ker(C) \subseteq \{0\}, \\
&\Leftrightarrow\ \ker(C) = \{0\},\ \text{converse inclusion always true,} \\
&\Leftrightarrow\ \mathrm{rank}(C) = n_x;\ \text{by linear algebra, because}
\end{aligned}
$$

$$
\mathrm{rank}(C) + 0 = \dim(\mathrm{Range}(C)) + \dim(\ker(C)) = \dim(\mathrm{Domain}(C)) = n_x.
$$

For formula of linear algebra, see:
Chapter 17, [1], Section 8.7, Th. 11, p. 242;
Chapter 17, [8], paragraph 50, Th. 1, p. 90.

# Observability

## Example

$$C = \begin{bmatrix} I_{n_x} \\ 0 \end{bmatrix} \in \mathbb{R}^{n_y \times n_x}, \ \mathrm{rank}(C) = n_x;$$

$$0 = Cx_a = \begin{bmatrix} I_{n_x} \\ 0 \end{bmatrix} x_a \ \Rightarrow \ \ x_a = 0.$$

How to determine $x_a$ from $y$ if map is injective?

$$y = Cx_a$$

$$\Leftrightarrow$$

$$\begin{bmatrix} y_1 \\ 0 \end{bmatrix} = \begin{bmatrix} I_{n_x} \\ 0 \end{bmatrix} x_a \ \Rightarrow \ x_a = y_1.$$

# Observability

### Def. Observability of a linear deterministic system

Consider a time-varying linear deterministic system,

$$
\begin{aligned}
x(t+1) &= A(t)x(t),\ x(0) = x_0,\\
y(t) &= C(t)x(t), \quad T = \mathbb{N},\ x_0 \in \mathbb{R}^{n_x},\\
&\quad x : T \to \mathbb{R}^{n_x},\ y : T \to \mathbb{R}^{n_y},\ \text{defined above;}\\
&\quad \text{define } \forall\ t_0 \in T,\ t_1 \in \mathbb{Z}_+ \text{ such that } t_0 + t_1 - 1 \in T,\\
&\quad \text{the state } x(t_0) \in \mathbb{R}^{n_x} \text{ to be}\\
&\quad \text{observable from the future outputs on the interval}\\
&\quad \{t_0,\ t_0+1,\ \ldots,\ t_0+t_1-1\} \subseteq T,\\
&\quad \text{if the following state-to-output map is injective}\\
&\quad x(t_0) \mapsto \{y(t_0),\ y(t_0+1),\ \ldots,\ y(t_0+t_1-1)\}.
\end{aligned}
$$

Call this system an observable system if the above condition holds for all $t_0 \in T$, $t_1 \in \mathbb{Z}_+$ sufficiently large, and for all $x(t_0) \in \mathbb{R}^{n_x}$.

# Observability

## Comments on observability of a linear system

- Observability means that from the future outputs
  $\{y(t_0),\ y(t_0+1),\ \ldots,\ y(t_0+t_1-1)\}$,
  one can uniquely determine
  the state $x(t_0)$ at the initial time $t_0 \in T$.
- In state-to-output map, one understands output
  as the output on the corresponding interval of time.
- See Section 21.3 of the book
  for observability of a deterministic linear system.

# Observability

## Proposition

Consider a time-varying linear system and $t_0 \in T$ and $t_1 \in \mathbb{Z}_+$ etc.
The state $x(t_0) \in \mathbb{R}^{n_x}$
is observable from the future outputs on the interval
if and only if,

$$n_x = \mathrm{rank}(O(A,\ C,\ t_0 : t_0 + t_1 - 1));$$

define,

$$O(A,\ C,\ t_0 : t_0 + t_1 - 1) = \begin{bmatrix} C(t_0) \\ C(t_0+1)\Phi(t_0+1, t_0) \\ \vdots \\ C(t_0+t_1-1)\Phi(t_0+t_1-1, t_0) \end{bmatrix}.$$

Call $O(.)$ the observability matrix of the corresponding interval.

# Observability

### Proof of proposition

The state-to-output map for this linear system has the form,

$$\bar{y} = \begin{bmatrix} y(t_0) \\ y(t_0+1) \\ \vdots \\ y(t_0+t_1-1) \end{bmatrix} = O(A,\ C,\ t_0 : t_0+t_1-1)\ x(t_0).$$

$$\begin{aligned} & x(t_0) \mapsto \bar{y} \text{ is injective,} \\ \Leftrightarrow \quad & \text{(by an earlier proposition),} \\ n_x = {} & \mathrm{rank}(O(A,\ C,\ t_0 : t_0+t_1-1)). \end{aligned}$$

# Observability

### Def. Observable pair

Consider a time-invariant linear system representation,

$$\begin{aligned} x(t+1) &= Ax(t),\ x(0) = x_0, \\ y(t) &= Cx(t). \end{aligned}$$

Call the matrix tuple $(A,\ C)$ an observable pair if

$$n_x = \operatorname{rank}(O_f(A,\ C)); \quad \text{define}$$

$$O_f(A,\ C) = \text{obsmat}(A,\ C) = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n_x-1} \end{bmatrix}.$$

Call $O_f(A,\ C)$ the observability matrix of the time-invariant linear system.

# Observability

**Proposition**

Consider a time-invariant linear system and $t_0 \in T$.
Equivalence of:

**(a)** The state $x(t_0) \in \mathbb{R}^{n_x}$
is observable from the future outputs on a sufficient large interval;

**(b)** $(A,\ C)$ is an observable pair;

**(c)** $n_x = \mathrm{rank}(O_f(A,\ C))$.

# Outline

# Stochastic Observability and Co-Observability

**Def. Stochastic observability**
Consider a stochastic system,

$$\{\Omega,\ F,\ P,\ T,\ \mathbb{R}^{n_y},\ B(\mathbb{R}^{n_y}),\ \mathbb{R}^{n_x},\ B(\mathbb{R}^{n_x}),\ y,\ x\}.$$

Call this system
stochastically observable on the interval
$\{t_0,\ t_0+1,\ \ldots,t_0+t_1-1\} \subseteq T$
if the stochastic state-to-output map
is injective on the support of $x(t_0)$,

$$x(t_0) \mapsto \text{cpdf}(\{y(t_0),\ y(t_0+1),\ \ldots,\ y(t_0+t_1-1)\}|\ F^{x(t_0)}).$$

The stochastic state-to-output map
goes from a random variable to a conditional probability measure.

# Stochastic Observability and Co-Observability

## Comments on stochastic observability

- Terms used:
  the stochastic state-to-output map
  the stochastic state-to-future-output map.
- By measurements one can in principle
  approximate the conditional measure used above.
- Support of the Gaussian random variable $x(t_0)$ is $\mathrm{Range}(Q_x(t_0))$.

$$\mathrm{Range}(Q_x(t_0)) = \{Q_x(t_0)x_a \in \mathbb{R}^{n_x} | \, \forall \, x_a \in \mathbb{R}^{n_x}\}.$$

- Stochastic observability of a time-invariant stochastic system
  is to be presented in Lecture 4.

# Stochastic Observability and Co-Observability

### Def. Stochastic co-observability

Consider a stochastic system,

$$\{\Omega,\ F,\ P,\ T,\ \mathbb{R}^{n_y},\ B(\mathbb{R}^{n_y}),\ \mathbb{R}^{n_x},\ B(\mathbb{R}^{n_x}),\ y,\ x\}.$$

Call this system
stochastically co-observable on the interval
$\{t_0 - 1,\ t_0 - 2,\ \ldots, t_0 - t_1\} \subseteq T$
if the stochastic state-to-past-output map
is injective on the support of $x(t_0)$,

$$x(t_0) \mapsto \mathrm{cpdf}(\{y(t_0 - 1),\ y(t_0 - 2),\ \ldots,\ y(t_0 - t_1)\} |\ F^{x(t_0)}).$$

### Remark

Stochastic observability and stochastic co-observability
are different concepts.

# Stochastic Observability and Co-Observability

## Theorem. Stochastic observability of a Gaussian system

Consider a time-varying forward Gaussian system representation,

$$\begin{aligned} x(t+1) &= A_f(t)x(t) + M_f(t)v_f(t), \; x(0) = x_0, \\ y(t) &= C_f(t)x(t) + N_f(t)v_f(t), \\ &\quad v_f(t) \in G(0, I_{n_v}), \; x(t) \in G(0, Q_x(t)). \end{aligned}$$

This system is stochastically observable on the interval,

$$\begin{aligned} &\{t_0, \; t_0+1, \; \ldots, \; t_0+t_1-1\}, \\ &\Leftrightarrow \; \ker(O_f(t_0 : t_0+t_1-1)Q_x(t_0)) = \ker(Q_x(t_0)); \\ &\Leftrightarrow \; \mathrm{rank}(O_f(t_0 : t_0+t_1-1)) = n_x, \;\; (\text{if } 0 \prec Q_x(t_0)); \\ &O_f(t_0 : t_0+t_1-1) \\ &= \begin{pmatrix} C_f(t_0) \\ C_f(t_0+1)\Phi_f(t_0+1, \; t) \\ \vdots \\ C_f(t_0+t_1-1)\Phi_f(t_0+t_1-1, \; t) \end{pmatrix}. \end{aligned}$$

# Stochastic Observability and Co-Observability

## Proof of theorem (1)

$$\overline{y}(t_0) = \begin{bmatrix} y(t_0) \\ y(t_0+1) \\ \vdots \\ y(t_0+t_1-1) \end{bmatrix} \in \mathbb{R}^{t_1 n_y},\ \overline{v}_f = \begin{bmatrix} v_f(t_0) \\ \vdots \\ v_f(t_0+t_1) \end{bmatrix} \in \mathbb{R}^{(t_1+1)n_{v_f}},$$

$$\begin{aligned} y(t_0+s) = {} & C_f(t_0+s)\Phi(t_0+s,t_0)x(t_0) + \\ & + [\sum_{r=t_0}^{t_0+s-1} C_f(t_0+s)\Phi(t_0+s-1,r)M_f(r)v_f(r)] + \\ & + N_f(t_0+s)v_f(t_0+s), \\ \overline{y}(t_0) = {} & O_f(t_0 : t_0+t_1-1)x(t_0) + \overline{M}_f(t_0)\overline{v}_f; \end{aligned}$$

Continued.

# Stochastic Observability and Co-Observability

## Proof of theorem (2)

Note the stochastic state-to-output map,

$$\begin{aligned}
& x(t_0) \mapsto E[\exp(iw^T \overline{y}(t_0)) |\ F^{x(t_0)}] \\
= & \exp(iw^T O_f(t_0 : t_0 + t_1 - 1)x(t_0) - w^T\ Q\ w/2),\ \forall\ w \in \mathbb{R}^{t_1 n_y}, \\
& \text{injective on the support of } x(t_0), \\
\Leftrightarrow\ & x(t_0) \mapsto O_f(t_0 : t_0 + t_1 - 1)x(t_0),\ \text{injective on the support of } x(t_0), \\
\Leftrightarrow\ & \forall\ x_a \in \mathbb{R}^{n_x},\ O_f(t_0 : t_0 + t_1 - 1)Q_x(t_0)x_a = 0\ \Rightarrow Q_x(t_0)x_a = 0, \\
& \text{injectivity of a linear map}, \\
\Leftrightarrow\ & \ker(O_f(t_0 : t_0 + t_1 - 1)\ Q_x(t_0)) \subseteq \ker(Q_x(t_0)), \\
\Leftrightarrow\ & \ker(O_f(t_0 : t_0 + t_1 - 1)\ Q_x(t_0)) = \ker(Q_x(t_0)), \\
& \text{because the converse inclusion always holds;} \\
\Leftrightarrow\ & \\
& n_x = \mathrm{rank}(O_f(t_0 : t_0 + t_1 - 1)),\ (\text{if } 0 \prec Q_x(t_0)).
\end{aligned}$$

# Stochastic Observability and Co-Observability

## Theorem. Stochastic co-observability of a Gaussian system

Consider a time-varying backward Gaussian system representation,

$$\begin{aligned} x(t-1) &= A_b(t)x(t) + M_b(t)v_b(t),\ x(0) = x_0, \\ y(t-1) &= C_b(t)x(t) + N_b(t)v_b(t),\ v_b(t) \in G(0, I_{n_v}), \\ & x(t) \in G(0, Q_x(t)). \end{aligned}$$

This system is stochastically co-observable on the interval,

$$\begin{aligned} & \{t_0-1,\ t_0-2,\ \ldots,\ t_0-t_1\}, \\ \Leftrightarrow\ & \ker(O_b(t_0-1:t_0-t_1)Q_x(t_0)) = \ker(Q_x(t_0)); \\ \Leftrightarrow\ & \mathrm{rank}(O_b(t_0-1:t_0-t_1)) = n_x,\ (\text{if } 0 \prec Q_x(t_0)); \end{aligned}$$

$$O_b(t_0-1:t_0-t_1) = \begin{pmatrix} C_b(t_0) \\ C_b(t_0-1)\Phi_b(t_0-1,\ t_0) \\ \vdots \\ C_b(t_0-t_1+1)\Phi_b(t_0-t_1+1,\ t_0) \end{pmatrix}.$$

# Stochastic Observability and Co-Observability

**Remark**

- Stochastic observability and stochastic co-observability are not identical.
  A Gaussian system representation
  can be stochastically observable and
  simultaneously not be stochastically co-observable.
  See an exercise of Homework Set 4.
- Stochastic observability and stochastic co-observability of a finite stochastic system is not fully characterized.
  More research to be done.

# Outline

# Concluding Remarks

## Overview of Lecture 03

- Concept of a stochastic system.
  Symmetric in time (forward and backward representations), and
  for any probability distributions.
  Concept of a state of a stochastic system.
  The state process is a Markov process.
- Gaussian system representations.
- Any Gaussian system has
  both a forward and a backward Gaussian system representation.
  The relation of these two representations is specified.
- Concepts of
  time-varying stochastic observability and of
  time-varying stochastic co-observability,
  and their characterizations.