WI4221: Control of Discrete-Time Stochastic Systems Lecture Notes (Q3-Q4, 24/25)

Gerb

Compilation Date: April 14, 2025

Contents

Lecture 01: Introduction and Probability

Probability Distributions and Measures

Probability Distribution Function

The definition of pdfs over $\mathbb{R}$ is defined below. We can define them over $\mathbb{R}^n$, however this is particularly difficult compared to $\mathbb{R}$ (see Ash, 1972). THe class of pds is a convex set:

$$\forall f_1, f_2 \text{ pdf and } \forall c \in [0,1] \implies cf_1 + (1-c)f_2 \text{ is a pdf}$$

Definition (Probability Distribution Function on $\mathbb{R}$)**.** *A pdf is a function $f : \mathbb{R} \to \mathbb{R}_+$ such that*

1. *f is increasing.* $u \leq v \implies f(u) \leq f(v)$
2. *f satisfies the limits* $\lim_{u\to-\infty} f(u) = 0$ *and* $\lim_{u\to+\infty} f(u) = 1$
3. *f is right continuous, satisfying the limit* $\lim_{v\downarrow u} f(v) = f(u)$

There are many subclasses of pdfs. They can be discrete:

$$f(u) = \sum_{k\in\mathbb{Z}} p(k) I_{[u_k,\infty)}(u)$$

With

$$I_{[u_k,\infty)}(u) = \begin{cases} 0, & u \in (-\infty, u_k) \\ 1, & u \in [u_k, \infty) \end{cases}$$

Where $p : \mathbb{Z} \to \mathbb{R}_+$ is a frequency function satisfying $\sum_{k\in\mathbb{Z}} p(k) = 1$.

We can also have an absolute continuous pdf:

$$f(u) = \int_{-\infty}^{u} p(v)\, dv$$

Note that the integral here is a Lebesgue-Stieltjes integral. We have $p : \mathbb{R} \to \mathbb{R}_+$, satisfying $\int_{-\infty}^{\infty} p(v)\, dv = 1$.

σ-Algebras

σ-Algebras formalize probability distributions

Definition (σ-Algebras of subsets of Ω)**.** *Let $F \subseteq Pwrset(\Omega)$ such that*

1. $\Omega \in F$
2. $A \in F \implies A^C \in F$
3. $\{A_k \in F, k \in \mathbb{Z}_+\} \implies \cup_{k\in\mathbb{Z}_+} A_k \in F$

(Ω, F) is called a measurable space. $G \subseteq F$ is called a sub-σ-algebra of F if G is a σ-algebra and $G \subseteq F$.

Some examples of σ-Algebras are the trivial σ-Algebra $F_0 = \{\emptyset, \Omega\}$. And $\{\emptyset, A, A^c, \Omega\}$, $\forall A \in \Omega$.

We can generate a σ-Algebra, using the following proposition: Consider Ω and a familiy $\{A_i \subseteq \Omega, i \in I\}$. There exists a smallest σ-Algebra $F(\{A_i, i \in I\})$ such that $\forall i \in I$, $A_i \in F(\{A_i, i \in I\})$.

Using σ-Algebras, we can define probability measures. We can define a measure if it is σ-additive. This should hold for any countable subset.

Definition (Probability Measures)**.** *Consider measurable space (Ω, F). The function $P : F \to \mathbb{R}_+$ called a measure if it is σ-additive:*

$$\forall \{A_k \in F, k \in \mathbb{Z}_+\} \text{ disjoint}$$

then this implies that

$$P(\cup_{k\in\mathbb{Z}_+ A_k} = \sum_{k\in\mathbb{Z}_+} P(A_k)$$

$P : F \to \mathbb{R}_+$ is called a probability measure if (1) it is measureable and (2) $P(\Omega) = 1$.

We can now use this to call (Ω, F, P) a probability space if (Ω, F) is a measurable space and P is a probability measure of (Ω, F).

On the real numbers we can construct probability measure as a function that assigns a probability to each element from a sample space Ω.

Theorem (Properties of Probability Measures)**.** *Properties of a probability measure (Ω, F, P).*

1. $P(\emptyset) = 0$
2. *Monoticity.* $A_1 \subseteq A_2$ *implies that* $P(A_1) \leq P(A_2)$
3. *Subadditivity.* $\{A_k \in F, k \in \mathbb{Z}_+\}$ *not neccesarily disjoint, implies that* $P(\cup_{k\in\mathbb{Z}_+} P(A_k)$
4. $0 \leq P(A) \leq 1$ *for all* $A \in F$

A useful result is the independence of σ-algebras. This mathematically formalizes the notion of statistical independence.

Random Variables

Definition (Random Variable)**.** *Consider a probability space $k(\Omega, F)$ and a measureable space (X, G). Define a random variable (rb) as a function $x : \Omega \to X$ such that , if $\forall A \in G, x^{-1}(A) = \{\omega \in \Omega | x(\omega) \in A\} \in F$*

To define whether this exists or not we use indicatior function $I_A : \Omega \to \mathbb{R}$ of a subset $A\ subset \Omega$, which is defined as

$$I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \iff \omega \in A^c \end{cases}$$

This tells us that I_A is a random variable off $A \in F$, where F is a σ-Algebra.

Examples of modeling with random variables are

- Binary random variables, used in e.g. information theory, networked control
- Uniform random variables, used to represent many outcomes with equal probability such as a fair die.
- Continuous random variables can be used to measure e.g. lifetimes of devices such as lamps.

Random Variables and σ-algebras

Definition (σ-Algebras generated by random variables)**.** *Consider* (Ω, F), (X, G) *and* $x : \Omega \to X$. *We define the sets*

$$x^{-1}(A) = \{\omega \in \Omega | X(\omega) \in A\}$$
$$x^{-1}(G) = \{x^{-1}(A) \in F | \forall A \in G\}$$

Then $x^{-1}(G)$ *is a* σ*-algebra. We denote* $F^x = F(x) = x^{-1}(G)$

Random variables generate σ-algebras. Additionally, random variables induce probability measures on the range space.

Definition (Probability Measure Induced by RV)**.** *The random variable* $x : \Omega \to \mathbb{R}$ *induces probability measure* P_x *on the range space according to*

$$P_x : B(\mathbb{R}) \to [0, 1]$$
$$P_x(A) = P(x^{-1}(A)) = P(\{\omega \in \Omega | x(\omega) \in A\})$$
$$(\Omega, F, P) \mapsto^x (\mathbb{R}, B(\mathbb{R}), P_x)$$
$$f_x(w) = P_x((-\infty, w]) \quad (\textit{note this is cdf})$$

Definition (Borel Measurable Function)**.** *Call a function* $h : \mathbb{R}^m \to \mathbb{R}^n$ *a Borel measurable function if*

$$h^{-1}(A) = \{x \in \mathbb{R}^{m|h(x) \in A}\}, \ \forall A \in B(\mathbb{R}^n)$$

A σ-algebra is a representation of any (nonlinear) Borel measure $y = h(x)$ such that x is measurable.

Characteristic Function

$$\mathbb{E}\left[\exp(jw^T x)\right] = \int_{\mathbb{R}^n} \exp(iw^T v) p_x(v) \ dv \quad \forall w \in \mathbb{R}^n$$

Note that the characterstic function is the **Fourier transform** of the probability density function.

Gaussian Random Variables

A Gaussian random variable with parameters m_x, Q_x is defined to be

$$x : \Omega \to \mathbb{R}^{n_x}, \ (m_x, Q_x) \in (\mathbb{R}^{n_x} \times \mathbb{R}^{n_x \times n_x}_{pds}$$

if

$$\mathbb{E}\left[\exp(jw^T x)\right] = \exp(jw^T m_x - \frac{1}{2} w^T Q_x w) \quad \forall w \in \mathbb{R}^{n_x}$$

Note that $x \in G(m_x, Q_x)$. Notation: $(x_1, \cdots, x_n)$ is jointly Gaussian if

$$x = (x_1, \cdots, x_n)^T \in G(m_x, Q_x)$$

Where $Q_x \succ 0$. An important result is the following: Affine functions of Gaussian random variables are themselves Gaussian. That is, if x is a random Gaussian variable, then

$$y = Ax + b$$

.is also Gaussian.

We can also have tuples of Gaussians:

$$x : \Omega \to \mathbb{R}^{n_x}, \ y : \Omega \to \mathbb{R}^{n_y}$$

Then we can write

$$(x, y) \in G(m_{(x,y)}, Q_{(x,y)})$$

With $Q_{(x,y)}$ the covariance matrix of x and y. They are independent if and only if there sigma algebras F^x and F^y are independent, or more simply if their covariance is 0 ($Q_{x,y} = 0$). In engineering, we often represent observations as signals with noise:

$$y = Cx + w$$

Where y is an observation, x the signal and w the noise. We write (y, x, w) are jointly Gaussian. We usually assume that y and w, and x and w are independent. (x, w) are also jointly Gaussian.

Conditional Expectation

Theorem (Conditional Expectation of a random variable given a σ-algebra)**.** *Consider a positive random variable*

$$(\Omega, F), G \subseteq F \ \textit{sub-}\sigma\textit{-algebra of } F$$
$$x : \Omega \to \mathbb{R}_+, \mathbb{E}[x] < \infty$$

There exists a random variable

$$\mathbb{E}[x|G] : \Omega \to \mathbb{R}_+$$

such that $\mathbb{E}[x|G]$ *is* G*-measurable and* $\mathbb{E}[x \ I_A] = \mathbb{E}[\mathbb{E}[x|G] \ \ I_A]$, $\forall A \in G$, *hence* $\mathbb{E}[\mathbb{E}[x|G]] = \mathbb{E}[x] < \infty$

The definition above is unique up to an almost sure modification. That is, the difference goes to 0 with expectation 1.

Theorem (Properties of Conditional Epectation)**.** *Consider* $(\Omega, F, P), G, G_1, G_2 \subseteq F$. $x, y : \Omega \to \mathbb{R}$, $\mathbb{E}[x] < \infty$, $\mathbb{E}[y] < \infty$

1. *Linearity,* $\mathbb{E}[ax + by|G] = a\mathbb{E}[x|G] + b\mathbb{E}[y|G]$
2. *Order Preservation,* $x \le y \implies \mathbb{E}[x|G] \le \mathbb{E}[y|G]$
3. *Measurability,* y *is* G *measurable, then* $\mathbb{E}[xy|G] = y\mathbb{E}[x|G]$, *in particular* $\mathbb{E}[y|G] = y$
4. *Reconditioning,* $G_1 \subseteq G_2 \implies \mathbb{E}[x|G_1] = \mathbb{E}[\mathbb{E}[x|G_2]|G_1]$ *and in particular* $\mathbb{E}[\mathbb{E}[x|G]] = \mathbb{E}[x]$
5. *Independence,* F^x *and* G *independent implies that* $\mathbb{E}[x|G] = \mathbb{E}[x]$

For simple random variable, we define variables in terms of an indicator function representation:

$$y = C_y i_y = \sum_{k=1}^{n_{i_y}} C_{y,k} i_{y,k}$$

Lecture 02: Stochastic Processes

Concepts of Stochastic Processes

Stochastic processes are time-series of phenomena. A stochastic system is, fora signal with fluctuations, a more realistic and of lower complexity than a deterministic system.

Definition (Stochastic Processes)**.** *Let* $(\Omega, F), (X, G), T \subseteq \mathbb{Z}$*.* $x : \Omega \times T \to X$ *is called a stochastic process if,* $\forall t \in T$*,* $x(\cdot, t) : \Omega \to X$ *is a random variable (a measureable function),*

$$\iff \forall t \in T, \forall A \in G, \underbrace{\{\omega \in \Omega | x(\omega, t) \in\}}_{x^{-1}(A)} \in F$$

Notation commonly used

$$x(t) = x_t = x_t(\omega) = x(\omega, t)$$
$$x = \{x(\omega, t) \in X, \forall t \in T, \forall \omega \in \Omega\}$$

Where T is a discrete-time index set which can have a finite horizon, half-infinite forward horizon or infinite horizon. Commonly we call

$$\forall \omega \in \Omega, x(\omega, \cdot) : T \to X$$

the **sample path** of the process.

Families of Distributions

An important concept for stochastic processes is a family of finite-dimensional probability distributions. This is because every time step t of the process corresponds to a distribution.

Definition (Family of finite-dimensional pdfs)**.** *Let* $x : \Omega \times T \to \mathbb{R}^n$*,* $T = \mathbb{N}$*, then*

In principle we can approximate such distributions based on observations using statistical tools.

Kolmogorov (1950) formalized the notation of existence of stochastic processes. Kolmogorov's theorem proved for $T = [0, 1]$, then also true for $T = \mathbb{R}_+$ and for $T = \mathbb{R}$. Similarly true for $T = \mathbb{N}$

Theorem (Existence of Stochastic Process (Kolmogorov, 1950))**.**

Consider T and (X, G) two stochastic process on these spaces are considered *equivalent* if their family of distributions is the same.

Defining Stochastic Process

Gaussian Process

> **Definition** (Gaussian Process). *A stochastic process $x : \Omega \times T \to \mathbb{R}^n$ is called a Gaussian process is each member of its family of FDPDFs is a Gaussian pdf. In terms of notation we write*
>
> $$\forall m \in \mathbb{Z}_+, \forall t_i \in T, t_i < t_j \text{ if } i < j, \{x(t_1), \cdots, x(t_m)\} \in G$$

Gaussian pdf is motivated by the central limit theorem, which states that scaled sum of a sequence of independent RVs converges to a Gaussian distribution.

Bernoulli Process

Another common process is a Bernoulli process. This is used often in information theory and can model e.g. a stream of bits in a communication channel.

> **Definition** (Bernoulli Process). $x : \Omega \times T \to \{0, 1\}$, $\{x(0), x(1), \cdots\}$ *is a sequence of i.i.d. random variables such that*
>
> $$\begin{aligned} q(t) &= \\ 1 - q(t) &= \end{aligned}$$

Poisson Process

Properties of Stochastic Process

We want to know if a process is integrable and square intebrable. It is is integrable if

$$\forall t \in T, \forall i \in \mathbb{Z}_{n_x}, \mathbb{E}|x_i(t)| < \infty;$$
$$m_x(t) = \mathbb{E}\left[x(t)\right], m_x : T \to \mathbb{R}^{n_x}$$

Call m_x the mean value function of x. Square integrable if

[insert definition square integrable]

A common idea that we need is positive-definite functions. A function $W : T \times T \to \mathbb{R}^{n_x \times n_x}$ is called a positive definite function if all entries of the matrix generated by the function are positive, that is

$$0 \leq \sum_{i=1}^{m} \sum_{j=1}^{m} c_i^T W((t_i, t_j) x_j$$ [1]

The function W on $T = \mathbb{N}$ is a covariance function if and only if

1. $W(s, t) = W(t, s)^T \ \forall s, t \in T$ (closed w.r.t. transposition)
2. W is positive-definite

[1] Note that if the inequality here is strict, the it strictly positive definite

Another important property is stationarity. A stochastic process is stationary if any FDPDF remains the same after a time-shift operation.

$$\begin{aligned} &x : \Omega \times T \to \mathbb{R}^{n_x}, T \subseteq \mathbb{Z} \\ &\text{if } \forall m \in \mathbb{Z}_+, \forall t_i \in T \text{ s.t. } i \in \mathbb{N}_m, \\ &\forall s \in \mathbb{Z} \text{ s.t. } t_i + s \in T, \\ &p(x(t_1), \cdots, x(t_m)) = p(x(t_1 + s), \cdots, x(t_m + s)) \end{aligned}$$

Note that often, we may need to remove a trend or something similar to "detrend" the data and transform it to a stationary process. Closely related to stationarity is time-invertibility. The definition is nearly the same, but instead we say reverse around a point in time s. Proving this is relatively straightforward. Because of this derivation, time-reversibility implies stationarity.

$$\begin{aligned} &x : \Omega \times T \to \mathbb{R}^{n_x}, T \subseteq \mathbb{Z} \\ &\text{if } \forall m \in \mathbb{Z}_+, \forall t_i \in T \text{ s.t. } i \in \mathbb{N}_m, \\ &\forall s \in \mathbb{Z} \text{ s.t. } s - t_i \in T, \\ &p(x(t_1), \cdots, x(t_m)) = p(s - x(t_1), \cdots, x(s - t_m)) \end{aligned}$$

Markov Processes

Conditional Independence

A different perspective on independence is to formulate it in terms of expected values. It can be shown to be equivalent to the independence of σ-algebras definition.

Conditional indepedence is a generalisation of independence. Conditional independence is widely used used in Engineering and Mathematics.

> **Theorem** (Conditional Independence Relation). *Given probability space (Ω, F, P), $F_1, F_2, G \subseteq F$, sub-σ-algebras; we write*
>
> $$\mathbb{E}\left[x_1 x_2 | G\right] = \mathbb{E}\left[x_1 | G\right] \mathbb{E}\left[x_2 | G\right]$$
>
> *for all $x_1 \in L(\Omega, F_1, \mathbb{R}_+)$, $x_2 \in L(\Omega, F_2, \mathbb{R}_+)$. Notation $(F_1, F_2 | G) \in CI$ where CI denotes conditionally independent relation of sub-σ-algebras*

As an elemantary property, $(F_1, F_2 \vee G)$ independent $\implies$ $(F_1, F_2 | G) \in CI$. The proof is relatively straightforward.

For Gaussian RVs, we have

$$(y_1, y_2, x) \in G(0, Q_{(y_1, y_2, x)})$$

with $y_1 : \Omega \to \mathbb{R}^{n_{y_1}}$, $y_2 : \Omega \to \mathbb{R}^{n_{y_2}}$ and $x : \Omega \to \mathbb{R}^{n_x}$, $Q_x \succ 0$.

Markov Processes Definition

A stochastic process is called a Markov Process if, for all times, the future and the past of the process are conditionally independent when conditioned on the present. Equivalently we write

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

where $x : \Omega \times T \to X$, (Ω, F, P), (X, G) and $F_t^{x+} = \sigma(\{x(s), \forall s \geq t\}$, $F_{t-1}^{x-} = \sigma(\{x(s), \forall s \leq t-1\}$. The future of the random process x *only* depends on the current state and not any past states. This holds for arbitrary non-linear transformations.

Interpretation of a Markov process in terms of measurable map from a state to a conditional measure on a future state

$$x(s) \mapsto \text{cpdf}(x(t)|F_s^{x-}) \; \forall s, t \in T, s < t$$

where cpdf denotes a *conditional probability distribution function*.

We can represent Markov processes recursively. If a Markov process is integrable, then there is a recursive representation of the form

$$\begin{aligned}
&x : \Omega \times T \to \mathbb{R}^{n_x}, \\
&x(t+1) = f(t, x(t)) + \Delta m(t), x(0) = x_0, \\
&f(t, x(t)) = \mathbb{E}\left[x(t+1)|F^{x(t)}\right] \\
&m(t) = \sum_{s=1}^{t} \Delta m(s), m : \Omega \times T \to \mathbb{R}^{n_x}
\end{aligned}$$

This process $\{m(t), F_t^x, t \in T\}$ is called a Martingale.

Gaussian Processes

A process is Gaussian if every member of its family of FDPFs is Gaussian pdf. Gaussian processes are square-integrable and define

$$\begin{aligned}
m_x(t) &= \mathbb{E}\left[x(t)\right], m_x : T \to \mathbb{R}^{n_x} \\
W_x(t,s) &= \mathbb{E}\left[(x(t) - m_x(t))(x(s) - m_x(s))^T\right], \\
&W_x : T \times T \to \mathbb{R}^{n_x \times n_x} \\
Q_x(t) &= \mathbb{E}\left[(x(t) - m_x(t))(\cdot)^T\right], \\
&Q_x : T \to \mathbb{R}_{pds}^{n_x \times n_x}
\end{aligned}$$

Note that $Q_x(t) = W_x(t,t)$ for all $t \in T$.

For stationary Gaussian processes, we have $W_x(t) = W_x(t, 0) = W_x(t+r, r)$ for all $r \in \mathbb{Z}$.

Some common types of Gaussian processes are

- Gaussian White Noise, $v(t) \in G(m_v(t), Q_v(t))$
- Stationary Gaussian White Noise, $v(t) \in G(m_v, Q_v)$
- Standard Stationary Gaussian White Noise, $v(t) \in G(0, I_{n_v})$

Theorem (Representation of a Gauss-Markov Process). *Consider a Gaussian process with the notation $x : \Omega \times T \to \mathbb{R}^{n_x}$, $x(t) \in G(0, Q_x(t))$. We assume that for all $t \in T$, $Q_x(t) \succ 0$. If x has the representation*

$$\begin{aligned}
&x(t+1) = A(t)x(t) + M(t)v(t), x(0) = x_0 \\
&x_0 : \Omega \to \mathbb{R}^{n_x}, x_0 \in G(0, Q_{x_0}), Q_{x_0} \succ 0 \\
&v : \Omega \to \mathbb{R}^{n_v}, \textit{ Standard Gaussian White Noise} \\
&F^{x_0}, F_\infty^v \textit{ are independent} \\
&A : T \to \mathbb{R}^{n_x \times n_x}, M : T \to \mathbb{R}^{n_x \to n_v}
\end{aligned}$$

then it is a Gauss-Markov process.

A natural question to ask is, given a Gaussian process, when is it Markov? We can derive that the Gaussian process if Markov if we can factorize the covariance matrix as

$$W_x(t,s) = W_x(t,r)W_x(r,r)^{-1}W_x(r,s)$$

where $s < r < t$ and $s, r, t \in T$.

Finite-Valued Processes

Consider

$$x : \Omega \times T \to \mathbb{Z}_{n_{i_x}} = \{1, 2, \cdots, n_{i_x}\} \subset \mathbb{Z}$$

We define an indicator process of the finite valued process x according to

$$i_{x,j}(\omega, t) = \begin{cases} 1, & \text{if } x(\omega, t) = j \\ 0, & \text{otherwise} \end{cases}$$

Now we can define a vector of indicator variables and represent

$$x(t) = C_x i_x(t)$$

We again find that the representation of a Finite-valued Markov process is represented in the form

$$\begin{aligned}
&i_x(t+1) = Ai_x(t) + \Delta m(t), i_x(0) = i_{x,0} \\
&x(t) = C_x i_x(t) \\
&\text{with } A \in \mathbb{R}_{st,+}^{n_{i_x} \times n_{i_x}} \text{ a stochastic matrix} \\
&Ai_x(t) = \mathbb{E}\left[i_x(t+1)|F^{x(t)}\right] = \mathbb{E}\left[i_x(t+1)|F^{i_x(t)}\right] \\
&0 = \mathbb{E}\left[\Delta m(t)|F_t^x\right] \; \forall t \in T \\
&\Delta m : \Omega \times T \to \mathbb{R}^{n_{i_x}}
\end{aligned}$$

We can prove that $\Delta m(t)$ is what is called a martingale increment at time $t \in T$.

Lecture 03: Stochastic Systems

Concept of Stochastic Systems

Definition (Stochastic System). *A stochastic system is 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,

$(\Omega\mathcal{F}, P)$ *complete probability space*

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

$(Y, B_Y), (X, B_X),$ *output and state space*

$y : \Omega \times T \to Y, x : \Omega \times T \to X$

And were we define the sigma algebras

$$\begin{aligned} F_t^{x-} &= \sigma(\{x(s), \forall s \leq t\} \\ F_t^{x+} &= \sigma(\{x(s), \forall s \geq t\} \\ F_t^{y+} &= \sigma(\{y(s), \forall s \geq t\} \\ F_{t-1}^{y-} &= \sigma(\{x(s), \forall s \leq t-1\} \\ F_{-1}^{y-} &= \{\Omega, \emptyset\} \end{aligned}$$

We denote the entire collection

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

The future and past are conditionally independent given the current state

See alternative conditions slide 13/60 lecture 03. All of these equivalently define independence of future and past given the state (Markov property).

Gaussian System Representation

Time-varying Gaussian system representation in discrete time, forward representation if the output and the state process are define by

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

In the case that $A(t), C(t), M(t), N(t)$ are time independent we define the time-invariant Gaussian system representation of the form

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

Where $v(t) \in G(0, I_{n_v})$, $x_0 \in G(m_{x_0}, Q_{x_0})$. Notation $(n_y, n_x, n_v mA, C, M, N)$.

In the literature, commonly, the generalised disturbance is used ($v(t) = [r(t), w(t)]^T$). Note that another common description that is *not* equivalent (and hence results may *not* be generalised) is given as

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

The Gausian system is state-output conditionally indpendent if

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

In the Gaussian representation we write

$$\mathbb{E}\left[M(t)v(t)(N(t)v(t))^T\right] = 0 \quad \forall t \in T$$

as a sufficient condition for independence. Gaussian system can also be written in terms of a state transition function

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

One can prove that

$$\begin{aligned} x(t) &= \Phi(t,s)x(s) + \sum_{r=s}^{t-1} \Phi(t-1,t)M(r)v(r) \\ y(t) &= C(t)\Phi(t,s)x(s) + \sum_{r=s}^{t-1} C(t)\Phi(t-1,t)M(r)v(r) \\ &\quad + N(t)v(t) \end{aligned}$$

This follows relative directly from a proof by induction. We can show using this representation that $x(t), y(t)$ are jointly Gaussian processes.

Slide 26/60 shows that a Gaussian system representation defines a stochastic system. One can observe slide 30/60 for the derivation of the output process representation.

Forward and Backward Representation

Systems can also be written in a backwards representation. They start at time t and recurse backward in time rather than forward. We define

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

We can reconstruct the forward representation from this backward representation. The forward and backward representation are related through the covariance of the processes X, Y and V.

Observability of Deterministic Linear Systems

Definition (Injective Functions). *Let* $h : \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$, $n_x, n_y \in \mathbb{Z}_+$. *If*

$$\forall x_a, x_b \in \mathbb{R}^{n_x}, h(x_a) = h(x_b) \implies x_a = x_b$$

This implies that x_a *can be directly determined from the output* $h(x_a)$.

In the linear case we write

$$h(x) = Cx$$

Where $\ker(C) = \{x_a \in \mathbb{R}^{n_x} | Cx_a = 0\}$ and $\mathrm{Im}(C) = \{Cx_b \in \mathbb{R}^{n_y} \ |\forall x_b \in \mathbb{R}^{n_x}\}$.

Theorem (Observability of a Linear System)**.** *Consider the time varying linear system*

$$\begin{aligned} x(t+1) &= A(t)x(t), x(0) = x_0 \\ y(t) &= C(t)x(t), \\ T &= \mathbb{N}, x_0 \in \mathbb{R}^{n_x} \end{aligned}$$

Define for all $t_0 \in T$*,* $t_1 \in \mathbb{Z}_+$ *such that* $t_0 + t_1 - 1 \in T$*. The state* $x(t_0) \in \mathbb{R}^{n_x}$ *to be observable from the future outputs on the interval*

$$\{t_0, t_0 + 1, \cdots, t_0 + t_1 - 1\} \subseteq T$$

if the following state-to-output map is injective

$$x(t_0) \mapsto \{y(t_0), y(t_0 + 1), \cdots, y(t_0 + t_1 - 1)\}$$

We call this system observable if the condition above holds for all $t_0 \in T$ *and* $x(t_0) \in \mathbb{R}^{n_x}$*.*

If we now define

$$\bar{y} = \begin{bmatrix} y(t_0) \\ y(t_0 + 1) \\ \vdots \\ y(t_0 + t_1 - 1) \end{bmatrix}$$

Which is equivalent to writing $n_x = \mathrm{Rank}(\mathcal{O})$, where $\mathcal{O}$ is defined as

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

In the case that $A(t), C(t)$ are time invariant we write

$$\mathcal{O}(A, C) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n_x - 1} \end{bmatrix}$$

Stochastic Observability and Stochastic Co-Observability

In the stochastic case, observability is not as straightforward to define as the state-to-output map is stochastic. The system is stochastically observable on the interval $(t_0, t_0 + 1, \cdots, t_0 + t_1 - 1) \subseteq T$ if the stochastic state-to-output map is injective on the support of $x(t_0)$,

$$x(t_0) \mapsto \mathrm{cpdf}(\{y(t_0), y(t_0+1), \cdots, y(t_0+t_1-1) \ |F^{x(t_0)}\})$$

This means that by measurements one can *in principle* approximate the conditional measure used above. Support of the Gaussian random variable $x(t_0)$ is

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

The system is considered stochastic co-observable if the stochastic state-to-past-output map is injective on the support of $x(t_0)$:

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

Note that stochastic observability and stochastic co-observability are **distinct** concepts.

Rank condition for stochastic observability

Given stochastic forward system $(A(t), C(t), M(t), N(t))$ we can show that this system is stochastically observable if

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

See the slides for the backwards representation. This reduces to also a rank condition of the backward representation.

Lecture 04: Time-Invariant Stochastic Systems

Controllability

Controllability is a necessary and sufficient condition for the existence of a stabalizing control law. It has been defined for the following:

1. Sets and Maps
2. Deterministic Linear Systems
3. Stochastic Systems
4. Gaussian Systems

Definition (Controllability of a map)**.** *Consider a tuple of sets and maps $(U, X, Y, g : U \to X, h : X \to Y)$. Call $g : U \to X$ the input-to-state map. Call the tuple controllable with respect to the subset $X_{co} \subseteq X$ if*

$$g : U \to X_{co} \subseteq X \text{ is surjective}$$
$$\iff \forall x_c \in X_{co} \subseteq X, exists u_c \in U : x_c = g(u_c)$$

If in addition $X_{co} = X$ then g is completely surjective and the tuple is completely controllable.

Interpetation: For each state there exists an input to reach that state. Note that in the case that

$$g(u) = Gu$$

we have complete surjectivity iff

$$\text{Rank}(G) = \text{dim}(\text{Im}(G)) = n_x$$

For time varying linear systems we say that $x_a \in X_{co} \subseteq X$ controllable on $T = \{t_0, t_0 + 1, \cdots, t_0 + t_1\}$ if the following input-to-state map is surjective:

$$\{u(t_0), u(t_0 + 1), \cdots, u(t_0 + t_1\} \mapsto x(t_0 + t_1) = x_a$$

If this holds for arbitrary t_0, t_1 and $X_{co} = X$, then we have complete controllability. We can again define this in terms of matrices of the linear time varying system if

$$\mathcal{C}(A, B : t_0 : t_0 + t_1 - 1) = \cdots$$

Where we have controllability if

$$\mathcal{C}(A, B, t_0 : t_0 + t_1 - 1) = n_x$$

In the LTI case this reduces to the well known controllability condition.

Theorem (Controllability of LTI Systems)**.** *LTI system (A, B) is controllable if and only if*

$$\text{Rank}(\mathcal{C}(A, B)) = n_x$$

Where

$$\mathcal{C}(A, B) = \begin{bmatrix} B & AB & \cdots & A^{n_x - 1}B \end{bmatrix}$$

General advice: Look at the singular values of the controllability matrix. If the singular values are further apart than a factor 10, this may indicate poor controllability and further investigation may be needed.

An interesting note is that by Kalman's work, we can decompose a system which is not controllable into a controllable normal form, where we split the system in a part that is and isn't controllable. We have

$$x(t + 1) = Ax(t) + Bu(t), \qquad x(0) = x_0$$

Which under $z(t) = Sx(t)$ yields

$$z(t + 1) = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} z(t) + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u(t), \qquad z(0) = z_0$$

Where (A_{11}, B_1) is controllable.

Stabilizability

(A, B) is a stabilizable type if one one of the following holds:

a. spec$(A_{22}) \subset \mathbb{D}_O$ after transfomation into Kalman form
b. $\forall \lambda \in \Lambda(A)$ we have $\lambda \in \mathbb{D}_O$ or Hautus' test (HJB lemma) holds ($n_x = \text{Rank}([A - \lambda \mathcal{I} \; B])$
c. $\lambda \in \Lambda(A)$ are spectrally assignable

Note: $\mathbb{D}_O$ is the complex open unit disc:

$$\mathbb{D}_O = \{c \in \mathbb{C} \mid |c| < 1\}$$

Note that for LTI systems to be stable we must have

$$\Lambda(A) \subset \mathbb{D}_O$$

Time-Invariant Gaussian Systems

Supportable Pairs

Closely related but distinct from controllable pairs

Definition (Supportable Pair)**.** *Consider*

$$x(t + 1) = Ax(t) + Mv(t)$$

Call the matrix tuple (A, M) a supportable pair if

$$n_x = \text{Rank}(\mathcal{C}(A, M))$$

Lyapunov Equation

Definition (Lyapunov Matrix Equation)**.** *Consider a time-invariant Gaussian sstems. Define the recursion of the state variance function and the discrete-time Lyapunov equations by the respective formulas:*

$$Q_x(t + 1) = AQ_x(t)A^T + MM^T, \qquad Q_x(0) = Q_{x_0}$$
$$Q = AQA^T + MM^T$$

With $m_x(t) = \mathbb{E}[x(t)]$ and $Q_x(t) = \mathbb{V}(x(t))$

A very important theorem on the Lyapunov equation is the following: For time invariant Gaussian systems if the matrix A is exponentially (asymptotically) stable, then

$$Q = \lim_{t\to\infty} Q_x(t)$$
$$Q = AQA^T + MM^T$$

Thus the covariance convergence to a unique solution of the Lyapunov equation iff A is stable. Note that we also have that $Q \succ 0$.

We also have that any 2 of the following 3 statements imply the 3rd:

1. $\Lambda(A) \subseteq \mathbb{D}_O$
2. (A, M) is supportable Pair
3. $Q \succ 0$

Exponential stability implies asymptotic convergence.

$$\exists c \in \mathbb{R}_+, \exists r \in (0,1) : \|Q_x(t) - Q_x(\infty)\|_2 \le c|r|^t$$

where we define

$$Q_x(\infty) = \lim_{t\to\infty} \frac{1}{t}\sum_{s=0}^{t-1} Q_x(s)$$
$$Q_x(\infty) = AQ_x(\infty)A^T + MM^T$$

Note that this is also related to the observability Grammian such that if we have

$$Q_c = AQ_cA^T + MM^T$$
$$Q_o = A^TQ_oA + C^TC\text{tr}(CQ_cC^T) = \text{tr}(M^TQ_oM)$$

Slide 27/56 lecture 04 highlights some useful inequalities when proving lower bounds and upper bounds for e.g. Ricatti equations.

Invariant Probability Measures

We can define probability measures on an image space. Consdier a stochastic system with state and output (x, y). For time $t \in T$, consider

$$\begin{bmatrix} x(\omega, t+1) \\ y(\omega, t) \end{bmatrix} : \Omega \to \mathbb{R}^{n_x+n_y}$$

This map induces a probability measure on the image space according to

$$(X \times Y, B(X \times Y)) = (\mathbb{R}^{n_x+n_y}, B(\mathbb{R}^{n_x+n_y}))$$

For $A \in B(\mathbb{R}^{n_x+n_y})$ we have

$$P_{(x^+,y),t}(A) = P(\{\omega \in \Omega | (x(\omega, t+1), y(\omega, t)) \in A\}$$

Then P is a probability measure on $X \times Y$ if this concers a Gaussian system then that measure is a Gaussian measure. We call the measure invariant if

$$\exists P_{(x^+,y)} : B(X) \otimes B(Y) \to [0, 1]$$

such that

$$P_{(x^+,y),t} = P_{(x^+,y)} \quad \forall t \in T$$

and we define $P_x = P_{(x^+,y)|B(X)}$ and $P_y = P_{(x^+,y)|B(Y)}$

Consider now a time-invariant Gaussian system (A, B, M, N) with $x_0 \in G(0, Q_{x_0})$ and $\Lambda(A) \in \mathbb{D}_O$. There exists an invariant measure of the system which is a Gaussian measure and which may be constructed as defined below:

$$Q_x = AQ_xA^T + MM^T,$$
$$Q_y = AQ_xA^T + NN^T,$$
$$Q_{x^+y} = AQ_xC^T + MN^T$$
$$Q_{(x^+,y)} = \begin{bmatrix} Q_x & Q_{x^+y} \\ Q_{x^+y}^T & Q_y \end{bmatrix}$$

Where $G(0, Q_{(x^+,y)})$ is the invariant probability measure.

Note that if the initial state x_0 does not have the invariant state probability measure, the in the limit the distribution converges to the invariant Gaussian measure:

$$\begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix} \in G$$

And we have

$$D - \lim_{t\to\infty} G\left(\begin{bmatrix} m_x(t+1) \\ m_y(t) \end{bmatrix}, \begin{bmatrix} Q_{x^+}(t+1) & Q_{x^+y}(t) \\ Q_{x^+y}(t)^T & Q_y(t) \end{bmatrix}\right)$$
$$= G(0, Q_{(x^+,y)}$$

Note that the support of the invariant state pdf $x(t)$ equals the state space $\mathbb{R}^{n_x}$ is equivalent to the condition that $Q_x \succ 0$ as well as the condition that (A, M) is a supportable pair. To prove this we look at the characteristic equation $\mathbb{E}\left[\exp(jw(x, y)^T | F_t^x)\right]$.

Definition (Backward Supportable Pair)**.** *Consider a backward time-invariant Gaussian System*

$$x(t-1) = A_bx(t) + M_bv_b(t), \qquad x(0) = x_0$$

Call the tuple (A_b, M_b) a backward supportable pair if

$$\text{Rank}(\mathcal{C}(A_b, M_b) = n_x$$

Call the system representation a backward supportable system representation if (A_b, M_b) is a backward-supportbale pair.

We can use the matrix Q_x to derive a relation between the forward and backward representation, as Q_x is invariant under the system representation. We have Q_x solve

$$Q_x = A_fQ_xA_f^T + M_fM_f^T$$

Then we have that the system representations are similarity transformations of one another

$$\begin{aligned}
A_f &= Q_x A_b^T Q_x^{-1} \\
C_f &= C_b Q_x A_b^T Q_x^{-1} + N_b Q_{v_b} M_b^T Q_x^{-1} \\
A_b &= Q_x A_f^T Q_x^{-1} \\
C_b &= C_f Q_x A_f^T Q_x^{-1} + N_f Q_{v_f} M_f^T Q_x^{-1}
\end{aligned}$$

Note that $\Lambda(A_f) = \Lambda(A_b)$.

Stochastic Observability and Co-Observability

Consider a time-invariant stochastic system and assume that there exists an invariant probability measure of the system, and that the state and output have the invariant measure. Call the state $x(t_0) \in C$ stochastically observable if the distribution of the future outputs is injective on the support of x. For the co-observability case, we have the same condition but on the past outputs.

Note that this all reduces in the time-invariant case to the check that

$$\text{Rank}(\mathcal{O}(A, C)) = n_x$$

Theorem (Characterization of Stochastic Observability). *Consider a forward time-invariant Gaussian System*

$$\begin{aligned}
x(t+1) &= Ax(t) + Mv(t), \quad x(0) = x_0 \\
y(t) &= Cx(t) + Nv(t) \quad v \in G(0, I) \\
&\Lambda(A) \in \mathbb{D}_O \\
&\textit{Then } \exists Q_x : \\
Q_x &= AQ_xA^T + MM^T; \\
&x_0 \in G(0, Q_x) \implies x(t) \in G(0, Q_x) \forall t
\end{aligned}$$

This system is stochastically observable if one of the following holds:

1. *ker*$(Q_x) =$ *ker*$(\mathcal{O}(A, C)Q_x$
2. (A, M) *is a supportable pair and* (A, C) *is an observable pair*

Note that like for controllability we can do a Kalman decomposition to write the system in Kalman Observable form

$$\begin{aligned}
x(t+1) &= \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} x(t) + \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} v(t) \\
y(t) &= \begin{bmatrix} C_1 & 0 \end{bmatrix} + Nv(t) \\
&\Lambda(A) \subset \mathbb{D}_o, (A_{11}, C_1) \text{ an observable pair}
\end{aligned}$$

Note that x_2 is excited by the noise but not present in y directly or via x_1, so we cannot "see" x_2 at the output side of the system.

Note. We can write out the same proof and theorems but for the backwards representation of the system. Using Q_x as a coordinate transformation we can show these to be equivalent.

We can use these outlined properties to show that some parts of the state space model are considered non-relevant if they are not supportable and not visible in the output. Then we can prove that a state-space realisation is in some sense minimal if we have a time-invariant Gaussian system with

- (A, M) supportable
- (A, C) observable
- (A_b, C_b) observable

Time-Invariant Stochastic Systems

We can represent the system in terms of indicator functions. We can extend many results from Gaussian systems to these systems, however care needs to be taken [insert system representation]

We have probability measure $p_x(t) = \mathbb{E}[i_x(t)]$ and $p_y(t) = \mathbb{E}[i_y(t)]$.

The (sub)systems might be irreducible and nonperiodic. There are some questions of existence and uniqueness of steady state equation $p_{x_s} = Ap_{x_s}$. Stochastic observability and stochastic co-observability are characterized in terms of the system matrices.

Lecture 05: Weak Stochastic Realisation of Gaussian Systems - I

Motivation and Problem Definition

Motivating Example

Given that we measured an output process, how do we obtain a time-invariant Gaussian system of the form? There are 2 main problems:

- System Identification, find *approximate* system parameters based on measurements
- Stochastic Realisation, Find an *exact* representation of a Gaussian system based on a covariance function

Example (Paper Machine Realisation)**.** Consider observations without inputs and compute

$$\{z(s) \in \mathbb{R}^{n_y}, \forall t \in T_1 = \{1, 2, , \cdots, t_1\}\}$$

Where we estimate the average of the series as

$$z_a = \frac{1}{t_1} \sum_{s=1}^{t_1 - 1} z(s)$$

we can then make an estimate of the covariance function such that

$$\hat{W}(t) = \frac{1}{(t_1 - t) - 1} \sum_{s=1}^{t_1 - t} (z(t+s) - z_a)(z(s) - z_a)^T$$

The function $\hat{W}$ is a covariance function if and only if it is PD function and if and only if it satisfies a block Toeplitz matrix condition for all times. We now want a representation of the form

$$\begin{aligned} x(t+1) &= Ax(t) + Mv(t), x(0) = x_0 \in G(0, Q_x) \\ y(t) &= Cx(t) + Nv(t), v(t) \in G(0, I_{n_v}) \end{aligned}$$

$*\text{spec}(A) \subset \mathbb{D}_o$,

$$Q_x = AQ_xA^T + MM^T$$

The probability measure associated with ouput y is equal to the probability measure associated with the stochastic process obtained from the measurements.

Problem Definition

Problem (Weak Gaussian Stochastic Realisation Problem: Consider a stationary Gaussian process taking values in measurable space $(\mathbb{R}^{n_y}, B(\mathbb{R}^{n_y}))$, $m(t) = 0$ for all $t \in T$ and with covariance function $W : T \to \mathbb{R}^{n_y \times n_y}$.

(a) Does there exists a stable time invariant Gaussian system representation?

(b) Characterize weak Gaussian stochastic realisation which are of minimal state-space dimension

(c) Classify or describe all weak Gaussian stochastic realisation of the considered process. Relate any two minimal weak stochastic realisation to one another

(d) Formulate a procedure by which one of can construct all weak Gaussian Stochastic Realisations

Stochastic Realisation Theory

Consider the covariance function of a stationary process $W : T \to \mathbb{R}^{n_y \times n_y}$ on $T = \mathbb{N}$. Define the finite Hankel matrix of W with k row block and m column blocks as

$$H_w(k, m) = \begin{bmatrix} W(1) & W(2) & \cdots & W(m-1) & W(m) \\ W(2) & W(3) & \cdots & W(m) & W(m+1) \\ W(3) & W(4) & \cdots & W(m+1) & W(m+2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ W(k) & W(k+1) & \cdots & W(k+m-2) & W(k+m-1) \end{bmatrix}$$

We can extend this to an infinite Hankel matrix which leads to the following: Define

$$\text{Rank}(H_W) = \sup_{k,m} \text{Rank}(H_W(k, m))$$

Where H_w is finite rank if the rank above is not infinite for the infinite Hankel matrix. This is a theoretical condition.

We define a set of state variance matrices associated with a parameterization of a covariance function by the formula

$$\begin{bmatrix} Q - FQF^T & G - FQH^T \\ (G0FQH^T)^T & J + J^T - HQH^T \end{bmatrix} \succeq 0$$

Which is an LMI on Q.

The weak Gaussian stochastic realisation and assume that

$$T = \mathbb{Z}, m(t) = 0, W(0) \succ 0, \lim_{t \to \infty} W(t) = 0$$

There exists a time-invariant Gaussian system representation such that the output process y equal the considered process in probability; equivalently if

$$w(t) = W_y(t) \quad \forall t \in T \iff \text{Rank}(H_W) < \infty$$

Note that the rank of the infinite Hankel matrix is in fact equal to the minimal dimension of the state space realisation of the process. Call this system a weak Gaussian Stochastic realisation. for any weak Gaussian realisation there exists some state variance $Q_x \in \mathbb{Q}_{lsdp}$. If a weak realiastion, then there exists a covariance function of the form

$$W(t) = \begin{cases} HF^{t-1}G & t > 0 \\ J + J^T & t = 0 \\ (HF^{t-1}G)^T & t < 0 \end{cases}$$

A weak Gaussian stochastic realisation with spec$(A) \subset \mathbb{D}_o$ is of minimal dimension over all such realisations if and only if the following holds :

- Support of the Gaussian measure $G(0, Q_x)$ of the state process equal $\mathbb{R}^{n_x}$ ($Q_x \succ 0$ iff (A, M) Supportable)
- If the system is stochastically observable

- If the system is stochastically co-observable

We now want to classify the description. Fix a covariance relation. Define the classification map as the function

$$c_{lsp} : Q_{lsdp} \to \text{WGSRP}_{\min} \text{ is a bijection}$$

We write

$$\begin{bmatrix} M \\ N \end{bmatrix} \begin{bmatrix} M & N \end{bmatrix} = \begin{bmatrix} Q - FQF^T & G - FQH^T \\ (G0FQH^T)^T & J + J^T - HQH^T \end{bmatrix}$$

Stochastic realiations have a unique minimal element and a unique maximal element.

$$\exists Q^-, Q^+ \in Q_{lsdp}, \forall Q \in Q_{lsdp} : Q^- \preceq Q \preceq Q^+$$

We also want to classify the realisation as minimal. We relate 2 realisation by a linear map:

$$\begin{aligned} A_1 &= L_x A_2 L_x^{-1} \\ C_1 &= C_2 L_x^{-1} \\ M_1 &= L_x M_2 U_v \\ N_1 &= N_2 U_v \end{aligned}$$

The realisation is minimal if it is a linear transformation of a minimal realisation. The realisation can be considered minimal if we take (F, G) controllable and (H, J) observable.

Theorem (Weak Gaussian Stochastic Realisation)**.** *Consider a stationary Gaussian process with zero mean value function and with covariance function W. Determine a minimal covariance realisation of W such*

$$W(t) = \begin{cases} HF^{t-1}G & t > 0 \\ J + J^T & t = 0 \\ G^T(F^T)^{-t-1}H^T & t < 0 \end{cases}$$

Determine a matrix $Q \in Q_{lsdp}$ which implies that $Q \in \mathbb{R}^{n_x \times n_x}_{pds}$ where n_x is the rank of the infinite Hankel matrix. Then construct $A = F$, $C = H$ and $n_v = n_x + n_y$, $M \in \mathbb{R}^{n_x \times n_v}$ and $N \in \mathbb{R}^{n_y \times n_v}$ such that

$$\begin{bmatrix} M \\ N \end{bmatrix} \begin{bmatrix} M & N \end{bmatrix}^T = \begin{bmatrix} Q - FQF^T & G - FQH^T \\ (G - FQH^T)^T & J + J^T - HQH^T \end{bmatrix} \succeq 0$$

$\Omega = \mathbb{R}^{n_x} \times (\mathbb{R}^{n_v})^T$, $\omega = (\omega_1, \omega_2)$ such that $x_0(\omega) = \omega_1$, $v(\omega, t) = \omega_2(t)$ such that $x_0 \in G(0, Q)$, $v(t) \in G(0, I_{n_v})$, F^{x_0}, F^v_∞ independent then

$$(n_y, n_x, n_v, A, C, M, N) \in \mathit{WGSRP}_{\min}$$

Kalman Realisation and the Kalman Filter

The Kalman realization is a time-invariant Gaussian system such that

- (A, M) is supportable pair ($Q_x \succ 0$)
- (A, C) is observable pair
- (A_b, C_b) is observable pair (co-observability)
- spec$(A - MN^{-1}C) \subset \mathbb{D}_o$

Consider a Kalman realisation and define

$$Q_x = AQ_xA^T + MM^T$$

then $Q_x \succ 0$and $Q_x = Q^- \in Q_{lsdp}(F, G, H, J)$. The state variance matrix Q_x of the Kalman realisation equal the minimal state variance. Kalman realised that any Kalman realisation can be written as the Kalman filter linear system

$$\begin{cases} x(t+1) &= Ax(t) + MN^{-1}[y(t) - Cx(t)] \\ v(t) &= N^{-1}[y(t) - Cx(t)] \end{cases}$$

Where we note that MN^{-1} is the Kalman gain. Note also that $F^{x(t+1)} \subseteq F^{y-}_t \vee F^{x_0}$.

Explanation of Minimality

Consider that if we have that (A, M) not supportable pair. In the long run x_2 is irrelevant, implying that $Q_x \not\succ 0$ and hence it cannot be minimal.

In the case that (A, C) is not observable there are components of x that do not show up in y. Thus we do not have that these states are relevant for the input-output behaviour and hence we can disregard the states.

Be careful about the following false conjecture:

- (A, M) is supportable
- (A, C) is stochastically observable
- These 2 facts together do NOT imply minimal weak Gaussian realisation of the output process. We also NEED co-observability

The Kalman filter is another Stochastic Realisation. Consider the time invariant Kalamn filter of time invariant Gaussian system

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + K[y(t) - C\hat{x}(t)] \\ \bar{v}(t) &= y(t) - C\hat{x}(t) \quad v(t) \in G(0, Q_{\bar{v}}) \end{aligned}$$

hence

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + K\bar{v}(t) \\ y(t) &= C\hat{x}(t) + \bar{v}(t) \end{aligned}$$

Note that these 2 representations are two stochastic realisations of the same output process y. The Kalman filter is also a weak Gaussian stochastic realisation.

We can transform system to a minimal realisation. We start with $(n_y, n_x, n_v, A, C, M, N) \in$ WGSRP and we assume A Schur stable. Suppose that (A, M) is not a supportable pair. Then transform the system to the Kalman form:

$$x(t+1) = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} x(t) + \begin{bmatrix} M_1 \\ 0 \end{bmatrix} v(t)$$

Then (A_{11}, M_1) with corresponding C_1 and N is a reduced order representation. This again holds for (A, C) not observable. Then there exists a Kalman decomposition such that

$$\begin{aligned} x(t+1) &= \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} x(t) + \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} v(t) \\ y(t) &= \begin{bmatrix} C_1 & 0 \end{bmatrix} x(t) + Nv(t) \end{aligned}$$

Where the pair (A_{11}, C_1) is observable and the representation

$$x(t+1) = A_{11}x(t) + M_1v(t)$$
$$y(t) = C_1x(t) + Nv(t)$$

is again of lower dimension. Finally, we can also do this for the backward representation for co-observability. After all 3 steps the realisation will always be minimal.

Strong Stochastic Realisation of Gaussian Process

Consider stationary stochastic process $\bar{y}$. Call a Gaussian system a strong Gaussian stochastic realisation of $\bar{y}$ if there exists a time invariant Gaussian system such that

$$(x, y) \text{ and } F^{x(t)} \subseteq F^{y}_{\infty}$$

such that $\bar{y} = y$ almost surely for all t. To construct a strong realisation in the σ-algebraic setting. Construct $x(t) \in X$ such that

$$(F^{y+}_{t}, F^{y-}_{t-1} | F^{x(t)}) \in CI$$

We have $F^{x(t)} \subseteq F^{y-}_{t-1}$. Paricular case y_+, y_- are jointly Gaussian finite dimensional vectors.

$$x = \mathbb{E}\left[y_+ | F^{y-}\right] = Q_{y+y-}Q^{-1}y_-$$

Note that $(F^{y+}_{t}, F^{y-}_{t-1} | F^{x(t)}) \in CI$ then (x, y) are state and output process of a stochastic system.

System Identification

Original motivation of Kalman to research stochastic realisation was system identification. For a system to be identifiable we need a characterization of minimality and the description of the equivalence class of stochastic realisation. A procedure to determine an approximat eweak stochastic realisation called the subspace identification algorithm. Construct of a canonical form is needed. For example Observable canonical form.

Consider stationary Gaussian process on a finite horizon

1. Fix time $t_0 \in T$. Restrict attention to finite future and past. Construct

$$F^{y-}_{t_0-1} \implies y_-(t_0 - t : t_0 - 1)$$
$$F^{y+}_{t_0} \implies y_+(t_0 - t : t_0 + t_1 - 1)$$

We have $(y_+, y_-) \in G$ and

$$x(t_0) = \mathbb{E}\left[y_+ | F^{y-}\right] = Ly_-(t_0 - t_1 : t_0 - 1$$

using $\hat{W}$ then $(F^{y+}, F^{y-} | F^{x(t_0)}) \in CI$. We compute

$$x(t_0 + 1) = Ly_-(t_0 - t_1 + 1 : t_0)$$

2. Construct the system matrices and the noise porcess.

$$(x(t_0), x(t_0+1), y(t_0)) \in G \implies$$
$$\begin{bmatrix} A \\ C \end{bmatrix} x(t_0) = \mathbb{E}\left[\begin{bmatrix} x(t_0+1) \\ y(t_0) \end{bmatrix} \middle| F^{x(t_0)}\right]$$

where $v : \Omega \times T \to \mathbb{R}^{n_x+n_y}$ Gaussian white noise. We have

$$v(t) = \begin{bmatrix} x(t+1) - Ax(t) \\ y(t) - Cx(t) \end{bmatrix} \in G(0, Q_v)$$
$$M = \begin{bmatrix} I_{n_x} & 0 \end{bmatrix}, N = \begin{bmatrix} 0 & I_{n_y} \end{bmatrix}$$

3. One obtains a Gaussian system representation with output process $y(t)$ which is an approximation of the Gaussian process.

Lecture 06: Weak Stochastic Realisation of Gaussian Systems - II

Realisation of Deterministic Linear Systems

Definition (Impulse Response Function). *Define the impulse response function of LTI system according to*

$$x(t+1) = Ax(t) + Bu(t), \quad x(0) = x_0$$
$$y(t) = Cx(t) + Du(t),$$
$$W_s(t) = \begin{cases} D, & t = 0 \\ CA^{t-1}B, & t \geq 1 \end{cases}$$

Where we can define the input as a unit impulse $\delta(t)$ and the output is given in terms of convolution sum of the impulse response and input.

Consider an impulse response function $W : \mathbb{N} \to \mathbb{R}^{n_y \times n_u}$ If there exists an LTI system $(n_y, n_x, n_u, A, B, C, D) \in LSP$ such that $W(t) = W_s(t)$ where $W_s(t)$ defined above then we call this system a realisation. The realisation is minimal if $x(t)$ is of minimal dimensions.

The realisation construct an internal description based on an external description:

- **External**: input-output behaviour of a linear system (n_y, n_u, W)
- **Internal**: input-output behaviour of a linear system $(n_y, n_x, n_u, A, B, C, D)$

Needed to quantify minimality:

$$\text{Rank}(\mathcal{C}(A,B)) = n_x, \qquad \text{Rank}(\mathcal{O}(A,C)) = n_x,$$

Theorem (Existence of a Realisation (Kalman, 1963)). *There exists a realisation $(n_y, n_x, n_u, A, B, C, D) \in LSP$ with finite state-space dimension if Rank(H_W) $< \infty$ and $W(0) = D$ where H_W is the infinite Hankel matrix and $W(0)$ is the impulse response function.*

Theorem (Minimality of a Realisation). *The realisation $(n_y, n_x, n_u, A, B, C, D) \in LSP$ with finite state-space dimension. The following statements are equivalent:*

- *The realisation is miniaml*
- $n_x =$ *Rank*$(H_W) < \infty$
- (A, B) *controllable,* (A, C) *observable*

Classify:

$$LSP_{\min} = \left\{ \begin{array}{c} (n_y, n_x, n_u, A, B, C, D) \in LPS \mid \\ \textit{Rank}(\mathcal{C}(A,B)) = n_x, \ \textit{Rank}(\mathcal{O}(A,C)) = n_x \\ W(0) = D, \ W(t) = CA^{t-1}B \ \forall t \in T \end{array} \right\}$$

Note that all minimal realisation can be related through a similarity transformation of the form

$$\tilde{A} = TAT^{-1}, \tilde{B} = TB, \tilde{C} = CT^{-1}, \tilde{D} = D$$

Note that if there exists a realisation, we can decompose H_W as

$$H_W = \sup_{k,m \in \mathbb{Z}_+} \mathcal{O}_k(A,C)\mathcal{C}_m(A,B)$$

Note that we also have

$$\begin{aligned} & n_x = \text{Rank}(H_W(k,m)) \\ \iff^{(1)} \quad & n_x = \text{Rank}(\mathcal{O}_k(A,C)), n_x = \text{Rank}(\mathcal{C}_m(A,B)) \\ \iff^{(2)} \quad & n_x = \text{Rank}(\mathcal{O}(A,C)), n_x = \text{Rank}(\mathcal{C}(A,B)) \\ \iff \quad & (A,C) \text{ observable, } (A,B) \text{ controllable} \end{aligned}$$

(1) follows from Sylvesters inequality and (2) follows from Cayley-Hamilton.

Note that we can reduce a realisation to a minimal realisation by sequential Kalman decomposition into an observable and controllable sub-space.

Note (Infinite Hankel Matrix Problem). The rank of the infinite Hankel matrix in terms of computation theory is undecidable, however we can make adequete numerical approximations on order to draw useful conclusions

Covariance Functions and Dissipative Systems

Definition (Dissipative Linear Systems). *Consider* $(n_y, n_x, n_u, F, G, H, J) \in LSP$. *Define the supply rate* $h : \mathbb{R}^{n_y} \times \mathbb{R}^{n_y} \to \mathbb{R}$*such that*

$$h(t) = u(t)^\top y(y)$$

Call this system dissipative with supply rate h if there exists a storage function $S(x)$ satisfying the dissipation inequality

$$S(x(t)) - S(x(t)) - \sum_{r=s}^{t-1} h(u(r), y(r)) \leq 0$$

The intuition here is the following $S(x(t)) - S(x(s))$ is the storage. $\sum_{r=s}^{t-1} h(u(r), y(r))$ is the energy supplied. The sum should be negative and hence energy is dissipated. Now 2 questions arise:

a. When is a linear system dissipative?
b. If a system is dissipative, classify all storage functions.

We interpret available storage as the maximal amount of energy which can be extracted from the system over the future interval. We write:

$$S^-(x) = \sup_{(t,x_1), u \in (F(T_d, U, 0, x, t, x_1)} \left[-\sum_{r=0}^{t-1} h(u(r), y(r)) \right]$$

The system is dissipative if and only if

- The available storage is a finite valued function
- The available storage is a storage function
- $0 \leq S^0 \leq S$

We can also define the required supply. Consider a linear control system and supply rate h. Define the required supply as

$$S^+(x) = \inf \sum_{r=s}^{-1} h(u(r), y(r))$$

We have that $S^+ \geq S \geq S^0 \geq 0$.

Note that we can relate dissipativity to covariance functions. A system is dissipative if and only if W is a PD function which happens if and only if W is a covariance function. See the notes/book for proof.

Define the set of state variance matrices. Consider a linear system with the assumption

$$
\begin{aligned}
lsp =& (n_y, n_x, n_u, F, G, H, J) \in LSP \\
& J + J^\top \succ 0, \text{spec}(F) \subset \mathbb{D}_o \\
lspd =& (n_y, n_x, n_u, F^\top, H^\top, G^\top, J^\top) \in LSP \\
& J + J^\top \succ 0, \text{spec}(F^\top) \subset \mathbb{D}_o
\end{aligned}
$$

Where lsp is a linear system and $lspd$ is the dual representation. Define the set of state matrices of storage based on lsp, define the set state variance matrices based on $lspd$.

Theorem (Algebraic Characterization of Dissipativeness)**.** *If the system is dissipative then there exists a minimal state-variance matrix which is a solution of the ARE of stochast realisation. Note that the ARE does not have a unique solution unless we impose*

$$
\begin{aligned}
& Q \succ 0 \\
& J + J^\top - HQH^\top \succ 0 \\
& \textit{spec}(F - (G - FQH^\top)(J + J^\top - HQH^\top)^{-1} \\
& \quad \times (G - FQH^\top)^\top) \subset \mathbb{D}_o
\end{aligned}
$$

Proof of Theorem Weak Gaussian Stochastic Realisation

Assume there exists a realisation

$$
\begin{aligned}
x(t+1) &= Ax(t) + Mv(t), \quad x(0) = x_0 \in G(0, Q_{x_0}) \\
y(t) &= Cx(t) + Nv(t) \quad \text{spec}(A) \subset D_o \\
\exists Q_x &\in \mathbb{R}_{pds}^{n_x \times n_x} : \\
Q_x &= AQ_xA^\top + MM^\top
\end{aligned}
$$

Then define

$$
\begin{aligned}
Q_{x^+,y} &= AQ_xC^\top + MN^\top \\
Q_y &= CQ_xC^\top + NN^\top
\end{aligned}
$$

and $F = A$, $H = C$, $G = Q_{x^+,y}$, $J + J^\top = Q_y$. Then define the impulse response function in terms of these functions. We can show that

$$H_W(k, m) = \mathcal{O}_k(H, F)\mathcal{C}_m(F, G)$$

which has rank $\leq n_x$ implying the infinite Hankel matrix has finite rank. This can be used to show that $Q_x = Q_W$ where W is impulse response function and it follows that the realised system is a weak Gaussian Stochastic Realisation.

Note that controllability of (F, G) is equivalent to (A_b, C_b) is an observable pair. Hence we need this to show minimality.

Canonical Forms

For a considered covariance function there exists a set of minimal weak Gaussian stochastic realisations. This set, in general, may be large. Mostly one selects the Kalman realisation.

Consider a set X with an equivalence relation E defined on it:

$$E \subseteq X \times X$$

Where we have

1. $(x, x) \in E$
2. $(x, y) \in E \implies (y, x) \in R$
3. $(x, y) \in E, (y, z) \in E \implies (x, z) \in E$

We define a canonical form of (X, E) as

$$X_{cf} \subseteq X : \forall x \in X, \exists x_{cf} \in X_{cf} : (x, x_{cf}) \in E$$

Note that one can interpret this geometrically.

Observable Canonical Form

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

$$
A_{cf} = \begin{bmatrix}
0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & & & \ddots & & \vdots \\
0 & 0 & 0 & \cdots & 1 & 0 \\
0 & 0 & 0 & \cdots & 0 & 1 \\
-a_0 & -a_1 & -a_2 & \cdots & -a_{n_x-2} & -a_{n_x-1}
\end{bmatrix}
$$
$$
C_{cf} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}
$$

Where a_i follows from the Cayley-Hamilton theorem.

Lecture 07: Stochastic Control Systems and Stochastic Control Problems

Stochastic Control Systems

Definition (Stochastic Control System). *Formally, we define a stochastic control system as a collection satisfying the relation*

$$\text{cpm}(\ (x(t+1), y(t)) \mid F_t^{x-} \vee F_{t-1}^{y-} \vee F_t^{u-}) \\ = \text{cpm}(\ (x(t+1), y(t)) \mid F^{x(t)} \vee F^{u(t)})\ \forall t \in T$$

which implies that we have that the next state and current measurement and the previous measurements, states and control inputs are conditionally independent given the current state and input:

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

Where $u : \Omega \times T \to U$ is the input process, $c : \Omega \times T \to X$ the state process and $y : \Omega \times T \to Y$ the output process. If the above does not explicitly depend on time it is time-invariant and if $x_0 \in G$ and cpm[2] is conditionally Gaussian, the system in Gaussian. We denote

$$\{\Omega, F, P, T, Y, B_Y, X, B_X, U, B_U, y, x, u\} \in \textit{StocCS}$$

Less formally, we can represent the systems defined above in terms of a recursive system:

$$x(t+1) = f(t, x(t), u(t), v(t)),\ x(0) = x_0$$

where F^{x_0}, F_∞^v are independent for all $t \in T$. We also have that $F^{v(t)}, F_t^u$ also independent. This system is time invariant if f is not explicitly dependent on time. In the Gaussian (linear) case we can write this recursion as

$$x(t+1) = A(t)x(t) + B(t)u(t) + M(t)v(t),\ x(t_0) = x_0, \\ y(t) = C(t)x(t) + D(t)u(t) + N(t)v(t),$$

Where (A, M) is a supportable pair and (A, B) a controllable pair. Additionally, we often have $n_y \leq n_v$ and Rank$(N) = n_y$, Rank$([M^T, N^T]) = n_v$.

One can use the characteristic equation to prove that this Gaussian representation conforms to the abstract definition above.

Definition (Controlled Output). *Define the controlled output on a finite horizon of a Gaussian control system representation*

$$x(t) = C_z(t)x(t) + D_z(t)u(t),\ \forall t \in T \setminus \{t_1\} \\ z(t_1) = C_z(t_1)x(t_1)$$

Stochastic Controllability

Informally, this concept is needed to define the set of a reachable states in finite time t_1. The idea is to go from input process on an interval and initial conditions to a conditional measure on the state at time t_1. See slide 21/62 of lecture 07 for more details on notation. Intuitively we know that the set of probability measures in generally strictly smaller than the set of all probability measures on the set X. We consider the set P_{co} of control-objective probability measures, where

$$P_{co}(X, B(X)) \subseteq P_c(t_1, X, B(X))$$

Where P_c is the set of reachable measures and P_{co} the set of measures that our control objective states we *want* to reach.

We consider a system $\Sigma \in$ StocCS. This is consdier stochasticalyl controllable in the control interval T_c with respect to the control objective probability measure P_{co} if

$$P_{co}(X, B(X)) \subseteq P_c(t_1, X, B(x)),\ T_c = \{t_0 : t_1\} \subseteq T$$

We also define stochastic co-controllability via a similar argument but this is practically not been applied in literature so far.

It can be shown that the set of reachable measures depends on $F^{x(t_0)}$, $F_{t_1-1}^u$ and the conditional covariance of x given the mentioned σ-algebras. We can control the mean, *but not the variance, as the variance does not depend on $u(t)$.*

Practically speaking, we can check whether the pair (A, B) is controllable using

$$\text{Rank}(\mathcal{C}(A, B)) = \text{Rank}\left(\begin{bmatrix} B & AB & \cdots & A^{n_x-1}B \end{bmatrix}\right) = n_x$$

Note that we need the property that (A, M) is a supportable pair. If this is not the case, then part of the stochastic system is deterministic and we need to check other properties. As before, if the system is *not* stochastically controllable, we can do a Kalman decomposition

$$x(t+1) = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} x(t) + \begin{bmatrix} B_1 \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} v(t) \\ \text{s.t. } (A_{11}, B_1) \text{ a controllable pair}$$

Control Laws

For control we distinguish:

- Input process $y : \Omega \times T \to U$
- Control law $g : X \times T \to U$

In general

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

In general a control law is more useful than an input trajectory. A control law is a mapping which specifies the control input for each state.

We specify the information structure as a σ-algebra family $\{G_t, t \in T\}$ such that for all $t \in T$. G_t specifies all the information available for the input $u(t)$. These are very useful for decentralized/distributed control.

Definition (Special Information structuresc). *Past-state information structure*

$$\{F_t^{x-}, \forall t \in T\},\ F_t^{x-} = \sigma(\{x(s), t_0 \le s \le t\}$$

Markov information structure

$$\{F^{x(t)}, \forall t \in T\},\ F^{x(t)} = \sigma(\{x(t)\})$$

Past-state information structure

$$\{F_{t-1}^{y-}, \forall t \in T\},\ F_{t-1}^{y-} = \sigma(\{y(s), t_0 \le s \le t-1\})$$

Classical information structure

$$\{H_t, \forall t \in T\}$$

if there is only one controller with one information structure and 2 satisfies perfect recall:

$$\forall t \in T,\ H_t \subseteq H_{t+1}$$

We will generally work with past-state control laws (depend on past states) and Markov control laws (depend on current state).

Closed-Loop systems

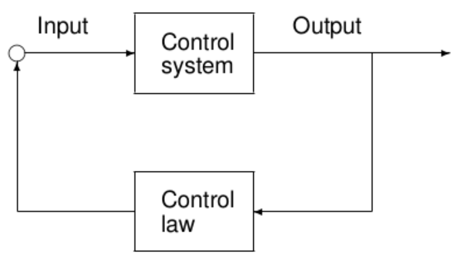

Figure 1: *Classical Closed-loop system*

Given Gaussian stochastic control system

$$\begin{cases} x(t+1) = A(t)x(t) + B(t)u(t) + M(t)v(t) \\ z(t) = C_z x(t) + D_z(t)u(t) \end{cases}$$

with control law $g_t : X^{t+1} \to U$, define recursively the closed-loop Gaussian stochastic system

$$\begin{cases} x^g(t+1) = A(t)x^g(t) + B(t)g_t(x^g(0:t)) + M(t)v(t) \\ z(t) = C_z x^g(t) + D_z(t)g_t(x^g(0:t)) \end{cases}$$

Note that the closed loop system x^g is a Markov process under a Markov control law. Proof is on slide 49/62 lecture 7.

Control Objectives

A control objective is a property that a control system can have, and that an engineer strives to attain. Important ones are:

- **Stability**, finite and bounded variance asymptotically
- **Assignment of Dynamics**, pole placement
- **Optimal Control**, minimize the cost function over all control laws
- **Robustness under uncertainty**, satisfactory dynamics under different operating conditions, unmodelled dynamics and exogenous pertubations.
- **Adaption**, satisfactory performance under slow variations over time (e.g. power systems where load varies over 24h each day).

Note. The general stochastic control problem is then, given a stochastic control system, an information structure, a set of admissible control laws and a set of control objectives, synthesize a control law such that the closed-loop system satisfies the control objectives as well as possible.

More on optimal control later, but generally the goal is to minimize a cost function $J(g)$ over all control laws g such that the expected value of the (quadratic) cost function is minimized. More formally:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + Mv(t) \\ y(t) &= Cx(t) + Du(t) + Nv(t) \\ z(t) &= C_z x(t) + D_z u(t) + Nv(t) \\ z(t_1) &= C_z x(t_1); \\ J &: G \to \mathbb{R}_+, \\ J(g) &= \mathbb{E}\left[z(t_1)^T z(t_1) + \sum_{s=0}^{t_1-1} z(s)^T z(s) \right] \end{aligned}$$

Then we solve

$$J^* = \inf_{g \in G} J(g) = J(g^*)$$

Distinguish:

- Control Synthesis: Develop control theory and design procedures. Develop approaches and procedures for finding control laws
- Control Design: Develop and compute actual control laws. Domain dependent and use simulation and testing