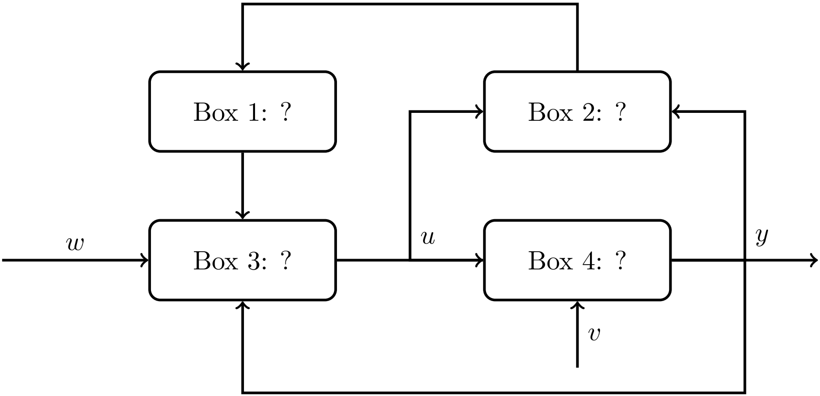
Stochastic Adaptive Control (02421)

Lecture 2

Tobias K. S. Ritschel

Section for Dynamical Systems, DTU Compute

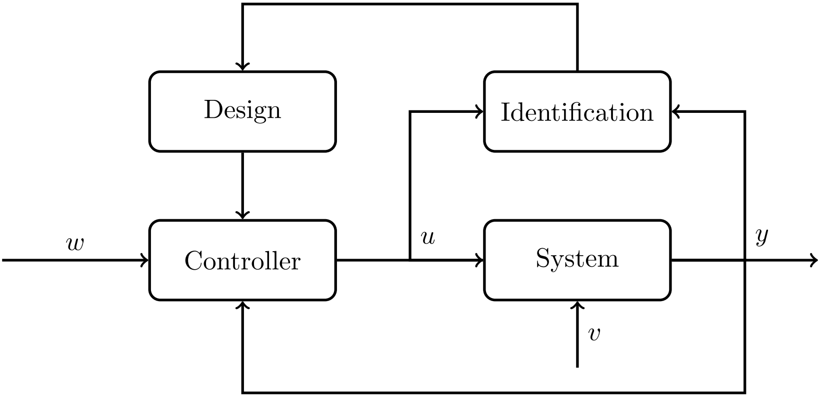
Stochastic Adaptive Control Diagram



What should the labels in the four boxes be and
can you explain the figure?

Think about it for yourself for one minute and
then discuss with the person next to you for one minute.

Stochastic Adaptive Control Diagram

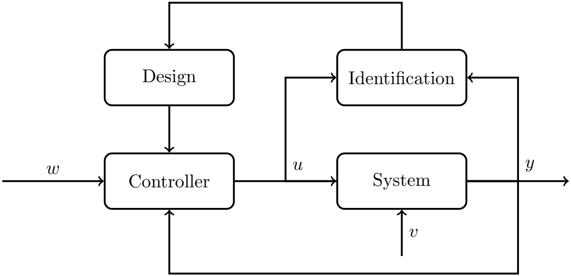


What should the labels in the four boxes be and
can you explain the figure?

Think about it for yourself for one minute and
then discuss with the person next to you for one minute.

Lecture Plan

1. Systems theory
2. **Stochastics**
3. State estimation - Kalman filter 1
4. State estimation - Kalman filter 2
5. Optimal control 1 - internal models
6. External models
7. Prediction + optimal control 1 - external models
8. Optimal control 2 - external models
9. System identification 1
10. System identification 2
11. System identification 3 + model validation
12. Adaptive control 1
13. Adaptive control 2



Today's Topics

- Follow-up from last lecture
- Stochastics
- Moments
- Confidence intervals
- Stochastic variables and vectors
- Stochastic processes and systems

Follow-up from Last Time: Exercises 1.1

You had to find the frequency domain correspondent of

$$\tau \dot{y} + y = u. \tag{1}$$

If we consider it a function $f(t)$ and remember that $\mathcal{L}\{\frac{\mathrm{d}y(t)}{\mathrm{d}t}\} = sY(s)$ holds, then the transformation becomes:

$$f(t) = \tau \dot{y}(t) + y(t) - u(t) = 0 \tag{2}$$

$$F(s) = \int_{-\infty}^{\infty} f(t) e^{-st} \, \mathrm{d}t = \int_{-\infty}^{\infty} (\tau \dot{y}(t) + y(t) - u(t)) e^{-st} \, \mathrm{d}t \tag{3}$$

$$= \tau \int_{-\infty}^{\infty} \dot{y}(t) e^{-st} \, \mathrm{d}t + \int_{-\infty}^{\infty} y(t) e^{-st} \, \mathrm{d}t - \int_{-\infty}^{\infty} u(t) e^{-st} \, \mathrm{d}t \tag{4}$$

$$= \tau s Y(s) + Y(s) - U(s) = 0 \tag{5}$$

The transfer function then becomes:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{1}{\tau s + 1} \tag{6}$$

Follow-up from Last Time: Exercises 5

You had to find eigenvalues, path to/from origin

$$x_{k+1} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} x_k + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u_k = Ax_k + Bu_k \tag{7}$$

1) The eigenvalues indicates asymptotic stability:

$$eig(A) = \{0, 0, 0\} \tag{8}$$

2-3)

$$A^2 = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad AB = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \tag{9}$$

Consequently, $U = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$ drives the states to the origin, whereas $U = \begin{bmatrix} -3 & 0 \end{bmatrix}$ does so in the minimum number of steps.

4) No, we cant affect state 3 at all.

Stochasticity

Real systems are usually stochastic in nature

$$\dot{x}_t = Ax_t + Bu_t + w_t \tag{10}$$

$$y_t = Cx_t + Du_t + e_t \tag{11}$$

stochastic: being uncertain, described by a random distribution and cannot be predicted precisely.

sources: measurements, model inaccuracy, unknown disturbances,...

Stochastic scalar variables

For a stochastic variable X

$$X \sim \mathcal{F}(\mathbf{p}) \tag{12}$$

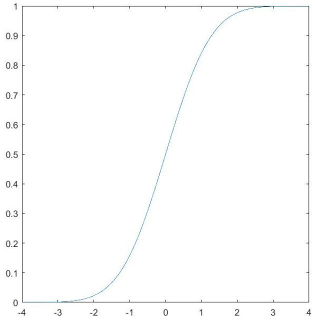
The cdf: cumulative distribution function $F_X(y)$:

$$F_X(y) = Pr\{X \leq y\} \in [0;1], \quad Pr\{a < X \leq b\} = F_X(b) - F_X(a) \tag{13}$$

The pdf: probability density function $f_X(y)$:

$$F_X(y) = \int_{-\infty}^{y} f_X(z)\,\mathrm{d}z, \tag{14}$$

$$f_X(y) \geq 0 \quad F(\infty) = 1 \tag{15}$$



Confidence Interval

A $1-p\%$ confidence interval $CI(p)$: the minimum interval $[a,b]$ which contains the $1-p\%$ most likely scenarios is

$$Pr\{a < X \leq b\} = 1 - p \tag{16}$$

For distributions that are symmetric around 0:

$$Pr\{X \leq a\} = p/2 \quad \text{or} \quad Pr\{X \leq b\} = 1 - p/2 \tag{17}$$

$$CI(p) = [F_X^{-1}(p/2), F_X^{-1}(1-p/2)] \tag{18}$$

Common usage is look-up tables for $F^{-1}(p/2)$ and the form:

$$X \in m_X \pm \sigma_X F^{-1}(p/2) \tag{19}$$

Example: Let us consider $X \in N(10,4)$. Then, a 95% CI is

$$10 - 2 \cdot 1.96 \leq X \leq 10 + 2 \cdot 1.96 \text{ or } 6.08 \leq X \leq 13.92 \tag{20}$$

Moments

For a real function $g(X)$:

$$\textbf{Nth moment of g(X):}\ E(g(X)^n) = \int_\Omega g(x)^n f(x) dx \tag{21}$$

Moments represent certain properties of stochastic variables.

$$\textbf{Mean (1st moment):}\ E(X) = m_x = \mu_x \tag{22}$$

$$\begin{aligned}\textbf{Variance (2nd central moment):}\ Var(X) &= E((X - m_x)^2) \\ &= E(X^2) - E(X)^2 = \sigma_x^2\end{aligned} \tag{23}$$

$$\textbf{Skewness (std.* 3rd central moment):}\ E((X - m_x)^3)/\sigma_x^3 \tag{24}$$

*Standardized.

Moments - Sample Moments

Consider having n samples of a variable X, the first two moments can be estimated by:

$$E(X) = \sum_{i=1}^{n} \frac{x_i}{n} \tag{25}$$

$$Var(X) = \sum_{i=1}^{n} \frac{(x_i - E(X))^2}{n} \tag{26}$$

alternative variance:

$$Var(X) = \sum_{i=1}^{n} \frac{(x_i - E(X))^2}{n-1} \tag{27}$$

$$\tag{28}$$

giving an unbiased estimate

Probabilities: Joint probability and independence

As mentioned the (marginal) probability of the statement: the variable X is less than x is true

$$Pr\{X \leq x\} = F_X(x) \tag{29}$$

The **joint probability** is then the chance of two (or more) statements are simultaneously true:

$$Pr\{X \leq x, Y \leq y\} = F_{X,Y}(x, y) \tag{30}$$

The **marginal distribution** can be computed from the joint distribution:

$$f_X(x) = \int_{\Omega_y} f_{X,Y}(x, y) dy \tag{31}$$

If the two variables are **independent:** $X \perp Y$:

$$F_{X,Y}(x, y) = F_X(x) F_Y(y), \quad f_{X,Y}(x, y) = f_X(x) f_Y(y) \tag{32}$$

Covariance

Covariance is a measure of how two stochastic variables varies relatively to each other

$$CoV(X,Y) = E((X - m_x)(Y - m_y)) \tag{33}$$

The variance is the just the self-covariance.

Correlation coefficient:

$$\rho = \frac{CoV(X,Y)}{\sqrt{Var(X)Var(Y)}}, \quad -1 \leq \rho \leq 1 \tag{34}$$

In the case of independent variables:

$$CoV(X,Y) = \rho = 0 \tag{35}$$

Note: The reverse if not true.

Dependency in probabilities - Bayes theorem

We can formulate probabilities under an assumption/condition
The likelihood given a condition:

$$Pr\{X \leq x|Y \leq y\} = \frac{Pr\{X \leq x, Y \leq y\}}{Pr\{Y \leq y\}} \quad (36)$$

$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y,X}(y|x)f_X(x) \quad (37)$$

The same can be done for the moments if $Var(X|Y) < \infty$ exists:

$$E(X|Y) = m_{x|y} = \int_{\Omega_x} x f_{X|Y}(x|y)dx \quad (38)$$

$$Var(X|Y) = E((X - m_{x|y})^2|Y) \quad (39)$$

Stochastic Vectors

For multiple variables we can utilize stochastic vectors:

$$\mathbf{X} = \left[X_1, \ldots, X_n\right]^T \tag{40}$$

$$\text{cdf: } F_{\mathbf{X}}(\mathbf{x}) = Pr(X_1 \leq x_1, \ldots, X_n \leq x_n,) \tag{41}$$

$$\text{marginal cdf: } F_{X_1}(x_1) = Pr(X_1 \leq x_1) \tag{42}$$

The 1st and 2nd moments are:

$$\mathbf{m}_x = E(\mathbf{X}) = \left[E(X_1), \ldots, E(X_n)\right]^T \tag{43}$$

$$P_x = P_x^T = Var(\mathbf{X}) = E((\mathbf{X} - \mathbf{m}_x)(\mathbf{X} - \mathbf{m}_x)^T) \geq 0 \tag{44}$$

Variance matrix being symmetric and positive semi-definite $y^T P_x y \geq 0$.
P_x is also diagonalizable.
for $n = 2$

$$P_x = \begin{bmatrix} Var(X_1) & CoV(X_1, X_2) \\ CoV(X_2, X_1) & Var(X_2) \end{bmatrix} \tag{45}$$

Mathematical properties of Moments: 1st and 2nd

Consider the constant matrix A and vector **m**

$$E(\mathbf{X} + \mathbf{m}) = E(\mathbf{X}) + \mathbf{m} \tag{46}$$
$$E(A\mathbf{X}) = AE(\mathbf{X}) \tag{47}$$
$$E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y}) \tag{48}$$
$$E(\mathbf{X}^T A\mathbf{X}) = tr(AVar(\mathbf{X})) + E(\mathbf{X})^T AE(\mathbf{X}) \tag{49}$$

$$Var(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X})E(\mathbf{X})^T \tag{50}$$
$$Var(\mathbf{X} + \mathbf{m}) = Var(\mathbf{X}) \tag{51}$$
$$Var(A\mathbf{X}) = AVar(\mathbf{X})A^T \tag{52}$$
$$Var(\mathbf{X} + \mathbf{Y}) = Var(\mathbf{X}) + Var(\mathbf{Y}) + CoV(\mathbf{X}, \mathbf{Y}) + CoV(\mathbf{X}, \mathbf{Y})^T \tag{53}$$

Hint: Check out the Matrix Cookbook.
Link: `https://www2.imm.dtu.dk/pubdb/pubs/3274-full.html`.

Vector Covariance and Variance

$$CoV(\mathbf{X},\mathbf{Y}) = E((\mathbf{X}-\mathbf{m}_x)(\mathbf{Y}-\mathbf{m}_y)^T) \tag{54}$$

$$CoV(\mathbf{X},\mathbf{X}) = Var(\mathbf{X}) = P_x \tag{55}$$

$$CoV(\mathbf{Y},\mathbf{X}) = CoV(\mathbf{X},\mathbf{Y})^T \tag{56}$$

$$CoV(A\mathbf{X},\mathbf{Y}) = ACoV(\mathbf{X},\mathbf{Y}) \tag{57}$$

$$CoV(\mathbf{X},A\mathbf{Y}) = CoV(\mathbf{X},\mathbf{Y})A^T \tag{58}$$

$$CoV(\mathbf{X}+\mathbf{V},\mathbf{Y}) = CoV(\mathbf{X},\mathbf{Y}) + CoV(\mathbf{V},\mathbf{Y}) \tag{59}$$

The principal directions of the variance (PCA):

$$[\Lambda,\mathbf{V}] = eig(P_x) \tag{60}$$

$$P_x\mathbf{V}_i = \lambda_i\mathbf{V}_i \tag{61}$$

where the vectors in **V** indicate the main directions of the variation, Λ indicating the variance associated with these directions.

Different Distributions - Gaussian and χ^2

Gaussian or normal distribution

$$X \in N(m_x, \sigma_x^2) \tag{62}$$

$$Y = \frac{X - m_x}{\sigma_x} \in N(0,1) \quad \text{standard Gaussian} \tag{63}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp(-\frac{(x-m_x)^2}{2\sigma_x^2}) \tag{64}$$

$$F_X(x) = F_Y(\frac{x - m_x}{\sigma_x}) \tag{65}$$

Different Distributions - Gaussian and χ^2

Gaussian or normal distribution

$$X \in N(m_x, \sigma_x^2) \tag{62}$$

$$Y = \frac{X - m_x}{\sigma_x} \in N(0,1) \quad \text{standard Gaussian} \tag{63}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp(-\frac{(x-m_x)^2}{2\sigma_x^2}) \tag{64}$$

$$F_X(x) = F_Y(\frac{x - m_x}{\sigma_x}) \tag{65}$$

We use the normal distribution a lot in stochastic modeling, but is there something about it that can be unrealistic?

Think about it for yourself for one minute and then discuss with the person next to you for one minute.

Different Distributions - Gaussian and χ^2

Gaussian or normal distribution

$$X \in N(m_x, \sigma_x^2) \tag{62}$$

$$Y = \frac{X - m_x}{\sigma_x} \in N(0,1) \quad \text{standard Gaussian} \tag{63}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} exp(-\frac{(x-m_x)^2}{2\sigma_x^2}) \tag{64}$$

$$F_X(x) = F_Y(\frac{x - m_x}{\sigma_x}) \tag{65}$$

χ^2-distribution

$$X = \sum_{i=1}^{n} \psi_i^2 \in \chi^2(n), \quad \psi_i \in N(0,1), \quad \psi_i \perp \psi_j \tag{66}$$

$$f(x) = \frac{1}{\Gamma(n/2)} x^{n/2-1} exp(-\frac{x}{2}) \tag{67}$$

$$E(X) = n \quad Var(X) = 2n \tag{68}$$

Different Distributions - Gamma

A Generalization of $\chi^2(n) = \Gamma(n/2, 2)$

$$X \in \Gamma(k, \theta), \quad 0 < X < \infty \tag{69}$$

$$f_X(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} exp(-\frac{x}{\theta}) \tag{70}$$

$$E(X) = k\theta, \quad Var(X) = k\theta^2 \tag{71}$$

The Gamma Function:

$$\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt \tag{72}$$

$$\Gamma(k+1) = k\Gamma(k) \tag{73}$$

$$\Gamma(1) = 1, \quad \Gamma(\frac{1}{2}) = \sqrt{\pi} \tag{74}$$

For integer values of k:

$$\Gamma(k) = (k-1)!, \quad \Gamma(k+\frac{1}{2}) = \frac{(2k-1)!}{2^k}\sqrt{\pi} \tag{75}$$

Other Related Distributions

The F-distribution:

$$X = \frac{Zm}{Yn} \in F(n, m) \tag{76}$$

$$Z \in \chi^2(n), Y \in \chi^2(m), \quad Z \perp Y \tag{77}$$

Student's t-distribution:

$$X = \frac{Z}{\sqrt{Y}} \sqrt{n} \in t(n) \tag{78}$$

$$Z \in N(0, 1), Y \in \chi^2(n), \quad Z \perp Y \tag{79}$$

The Rayleigh distribution:

$$X = \sqrt{Y_1^2 + Y_2^2} \in Ray(\sigma_y^2) \quad Y_i \in N_{iid}(0, \sigma_y^2) \tag{80}$$

Stochastic Tricks

Stochastic variable change:

$$X \in N(m, P), \quad Z \in N(0, I) \tag{81}$$

$$X = SZ + m \tag{82}$$

$$P = SS^T \tag{83}$$

The Cholesky Factorization:

$$S^T = chol(P) \tag{84}$$

The eigenvector approach:

$$Pv_i = \lambda_i v_i \quad \rightarrow PV = VD \tag{85}$$

$$S = V\sqrt{D} \tag{86}$$

Stochastics in Matlab

Let us take a look how we can generate stochastic in Matlab.

A Stochastic System

We will now extend the discrete-time (deterministic) systems

$$x_{k+1} = Ax_k + Bu_k \tag{87a}$$

$$y_k = Cx_k + Du_k, \tag{87b}$$

to discrete-time stochastic systems of the form

$$x_{k+1} = Ax_k + Bu_k + G\xi_k \tag{88a}$$

$$y_k = Cx_k + Du_k + F\eta_k. \tag{88b}$$

The noise is split into two components:

- $\{\xi_k,\ k \in \mathbb{N}\}$ denotes the process noise.
- $\{\eta_k,\ k \in \mathbb{N}\}$ denotes the measurement/sampling noise.

In discrete-time stochastic systems, x_k a stochastic variable, and the evolution of the system is a stochastic process.

Stochastic Processes

Stochastic process:
A sequence of stochastic variables $\{x(t,\omega), t \in T, \omega \in \Omega\}$, where Ω denotes the sample space of the uncertainty, and T denotes the time range

For a fixed t, $x(t,\cdot)$ is a random variable and for a fixed $\omega \in \Omega$, $x(\cdot,\omega)$ is a realization of the stochastic process; also called a time-series.

A stochastic process can be described using a marginal CDF or pdf

$$F_{X_t}(x_t, t) = Pr\{X_t \leq x_t\} \tag{89}$$

$$f_{X_t}(x_t, t) = \nabla_{x_t} F_{X_t}(x_t, t) \tag{90}$$

or if the different times are related, using joint probabilities

$$F_{X_t,X_s}(x_t, x_s, t, s) = Pr\{X_t \leq x_t, X_s \leq x_s\} \tag{91}$$

Properties of Stochastic processes

Three important ways to describe the statistical properties of a process are

The mean:

$$m_x(t) = \mathtt{E}\Big[x(t)\Big] = \int_{-\infty}^{\infty} z f_{x(t)}(z)\,\mathrm{d}z, \tag{92}$$

The variance:

$$P_x(t) = \mathtt{V}\Big[x(t)\Big] = \mathtt{E}\Big[\big(x(t) - \mathtt{E}[x(t)]\big)\big(x(t) - \mathtt{E}[x(t)]\big)'\Big], \tag{93}$$

The auto-covariance function:

$$r_x(t_1, t_2) = \mathtt{Cov}\Big[x(t_1), x(t_2)\Big] = \mathtt{E}\Big[\big(x(t_1) - \mathtt{E}[x(t_1)]\big)\big(x(t_2) - \mathtt{E}[x(t_2)]\big)'\Big], \tag{94}$$

where we have the identity that $r_x(t, t) = P_x(t)$.

Auto-covariance and auto-correlation

Similar to the general covariance function, an auto-correlation function can be defined as

$$\rho_x(t_1, t_2) = \frac{r_x(t_1, t_2)}{\sqrt{P_x(t_1)P_x(t_2)}}. \tag{95}$$

The auto-correlation function is often used in the model design and validation phase.

If the underlying process is stationary, then the auto-covariance function (and the auto-correlation) is only dependent on the difference $t_1 - t_2$. Let τ denote this time-difference, then the auto-covariance and auto-correlation functions are considered as univariate functions defined according to

$$r_x(\tau) = \texttt{Cov}\Big[x(t), x(t+\tau)\Big] \tag{96a}$$

$$\rho_x(\tau) = \frac{r_x(\tau)}{P_x(\tau)}. \tag{96b}$$

Stationarity

In deterministic systems, we work with stationary points. Similarly, in stochastic systems, we work with **stationary distributions:** a time-invariant distribution.

A process $x(t)$ is said to be **strongly stationary** of order n if the distribution functions are time-invariant - i.e. when

$$f_{x(t_1),\ldots,x(t_n)}(z_1,\ldots,z_n) = f_{x(t_1+h),\ldots,x(t_n+h)}(z_1,\ldots,z_n), \tag{97}$$

for any $n \in \mathbb{N}$ and $h \in \mathbb{R}$.

A process is said to be **weakly stationary** if the first two moments (the mean and covariance) are time-invariant. while the auto-covariance is

$$r_x(s,t) = r_x(s-t) \tag{98}$$

Stationarity with ergodicity

If a process is ergodic, only a single realization of the stochastics is needed to compute the statistic properties (moments).

A stationary process is said to be **weakly ergodic**, if its ensemble averages equal appropriate time averages of samples.

$$E\{x_t\} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} x_t \quad (1^{\text{st}} \text{ ord.}) \tag{99}$$

$$E\{x_t x_s^T\} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} x_t x_s^T \quad (2^{\text{nd}} \text{ ord.}) \tag{100}$$

Let g be an arbitrary function for which $E\{g(x_t)\}$ exists. Then, a stationary process is said to be **strongly ergodic** if

$$E\{g(x_t)\} = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} g(x_t) \tag{101}$$

Normal process and Markov process

A process $x(t)$ is said to be a **normal process** (or Gaussian process) if any finite dimensional distribution function $f_{x(t_1),\ldots,x(t_n)}(z_1,\ldots,z_n)$ is a multivariate normal distribution for any $n \in \mathbb{N}$.

If Y follows an n-dimensional multivariate normal distribution with mean μ and covariance Σ, then the distribution function, f, is given by

$$f_Y(y) = \frac{1}{(2\pi)^{n/2}\sqrt{\mathtt{det}(\Sigma)}}\mathtt{exp}\left(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right). \tag{102}$$

A process $x(t)$ is said to be a **Markov process** if for any $t_1 < t_2 < \cdots < t_n$, the distribution of $x(t_n)$ given $(x(t_1),\ldots,x(t_{n-1}))$ is the same as the distribution of $x(t_n)$ given $x(t_{n-1})$. Thus

$$\mathtt{Prob}\Big(x(t_n) \leq x \mid x(t_{n-1}),\ldots,x(t_1)\Big) = \mathtt{Prob}\Big(x(t_n) \leq x \mid x(t_{n-1})\Big) \tag{103}$$

Stochastic State-Space Models

We now return briefly to stochastic state-space systems in the form

$$x_{k+1} = Ax_k + Bu_k + G\xi_k, \quad \xi_k \sim F(\mu_\xi, P_\xi) \tag{104a}$$

$$y_k = Cx_k + Du_k + F\eta_k, \quad \eta_k \sim F(\mu_\eta, P_\eta) \tag{104b}$$

The mean, μ_k, and the covariance, P_k, of this process evolve according to

$$\mu_{k+1} = A\mu_k + Bu_k + G\mu_\xi, \quad \mu_0 = \mathtt{E}\left[x_0\right] \tag{105a}$$

$$P_{k+1} = AP_kA^T + GP_\xi G^T, \quad P_0 = \mathtt{Cov}\left[x_0\right] \tag{105b}$$

where u_k is assumed to be deterministic.

Stochastic State-Space Models

We now return briefly to stochastic state-space systems in the form

$$x_{k+1} = Ax_k + Bu_k + G\xi_k, \quad \xi_k \sim F(\mu_\xi, P_\xi) \tag{104a}$$
$$y_k = Cx_k + Du_k + F\eta_k, \quad \eta_k \sim F(\mu_\eta, P_\eta) \tag{104b}$$

The mean, μ_k, and the covariance, P_k, of this process evolve according to

$$\mu_{k+1} = A\mu_k + Bu_k + G\mu_\xi, \quad \mu_0 = \mathtt{E}\left[x_0\right] \tag{105a}$$
$$P_{k+1} = AP_kA^T + GP_\xi G^T, \quad P_0 = \mathtt{Cov}\left[x_0\right] \tag{105b}$$

where u_k is assumed to be deterministic.

How do the different terms on the right-hand side of (104a) affect the distribution of the states over time?

Think about it for yourself for one minute and then discuss with the person next to you for one minute.

Stochastic State-Space Models

We now return briefly to stochastic state-space systems in the form

$$x_{k+1} = Ax_k + Bu_k + G\xi_k, \quad \xi_k \sim F(\mu_\xi, P_\xi) \tag{104a}$$

$$y_k = Cx_k + Du_k + F\eta_k, \quad \eta_k \sim F(\mu_\eta, P_\eta) \tag{104b}$$

The mean, μ_k, and the covariance, P_k, of this process evolve according to

$$\mu_{k+1} = A\mu_k + Bu_k + G\mu_\xi, \quad \mu_0 = \mathtt{E}\left[x_0\right] \tag{105a}$$

$$P_{k+1} = AP_kA^T + GP_\xi G^T, \quad P_0 = \mathtt{Cov}\left[x_0\right] \tag{105b}$$

where u_k is assumed to be deterministic.

From this we can define the stationary mean, μ_∞, and variance, P_∞, according to

$$\mu_\infty = A\mu_\infty + Bu_\infty + G\mu_\xi \tag{106a}$$

$$P_\infty = AP_\infty A^T + GP_\xi G^T. \tag{106b}$$

Stochastic State-Space Models

If A has full-rank and the eigenvalues lie within the unit circle, then the stationary mean is the zero-vector. Under such conditions, the auto-covariance of the stationary process is given by

$$r_x(\tau) = A^\tau P_\infty. \tag{107}$$

Continuous-Time Stochastic Processes (SDE)

In continuous time, the inclusion of stochastics into the system is more tricky. Consider the ODE:

$$\frac{\partial x(t)}{\partial t} = f(x(t), u(t), t). \tag{108}$$

The rate of change of the state is uniquely determined by the time and current value of the state and input.

Initially, we assume that this rate of change is a random variable whose probability distribution is uniquely determined by the time and the current value of the state vector. Hence,

$$\frac{\partial x(t)}{\partial t} = f(x(t), u(t), t) + g(x(t), u(t), t)v(t) \tag{109}$$

where $\{v(t),\ t \in \mathbb{R}\}$ is a scalar stochastic process with:

- $v(t) \perp v(s)$ for any $t \neq s$ (independence)
- $v(t)$ is strongly stationary
- $E\{v(t)\} = 0, \forall t$ (zero-mean)

Continuous-Time Stochastic Processes (SDE)

Unfortunately, the process $v(t)$ cannot be continuous with those properties. In fact, $v(t)$ is almost surely discontinuous everywhere. Consequently, it is not integrable.

An approximation can be achieved by representing $v(t)$ by a suitable white noise process.
Alternatively we can represent the system on its difference form (discrete):

$$x(t+\Delta t) - x(t) = f(x(t),t)\Delta t + g(x(t),t)v(t)\Delta t + o(\Delta t). \tag{110}$$

We can replace $v(t)\Delta t$ with $\Delta W(t) = W(t+\Delta t) - W(t)$, where $W(t)$ has stationary independent increments with zero-mean and continuous: i.e. a Brownian motion (Wiener Process).
Taking the limit $(\Delta t \to 0)$, we obtain the SDE

$$\mathrm{d}x(t) = f(x(t),t)\,\mathrm{d}t + g(x(t),t)\,\mathrm{d}W(t) \tag{111}$$

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau),\tau)\,\mathrm{d}\tau + \int_{t_0}^{t} g(x(\tau),\tau)\,\mathrm{d}W(\tau) \tag{112}$$

Continuous-Time Stochastic Processes (SDE)

A stochastic integral can be defined using many different limiting schemes (all with different probabilistic properties). Using the Itô-interpretation implies that the SDE will be a Markov process.

The two first (conditional) moments of this difference process is given by

$$\mathrm{E}\Big[x(t+\Delta t) - x(t) \big|\, x(t)\Big] = f(x(t), \Delta t)\Delta t + o(\Delta t) \tag{113a}$$

$$\mathrm{V}\Big[x(t+\Delta t) - x(t) \big|\, x(t)\Big] = \sigma(x(t), t)\sigma(x(t), t)'\Delta t + o(\Delta t), \tag{113b}$$

where we have used the identity

$$\mathrm{E}\Big[W(t+\Delta t) - W(t)\Big]^2 = I\Delta t, \quad (I \text{ is the identity matrix}). \tag{114}$$

We notice that the variance is proportional to Δt and not Δt^2!

Brownian motion has the property that it's difference is zero-mean and normally distributed:

$$\Delta W(t) = W(t+\Delta t) - W(t) \in N(0, I\Delta t) \tag{115}$$

Linear Stochastic Differential Equations

We will now restrict our attention to linear stochastic differential equations. These attain the form

$$\mathrm{d}x(t) = (A(t)x(t) + B(t)u(t))\,\mathrm{d}t + G(t)\,\mathrm{d}w(t), \tag{116}$$

where we will assume that the initial distribution, x_0, follows a normal distribution with mean, m_0, and covariance, R_0, and the matrices $A(t)$ and $B(t)$ are continuous functions of time.

The expectation of such linear stochastic differential equation is given by

$$\mathtt{E}x(t) = \mathtt{E}x_0 + \mathtt{E}\int_{t_0}^{t} A(\tau)x(\tau) + B(\tau)u(\tau)\,\mathrm{d}\tau + \mathtt{E}\int_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau) \tag{117}$$

$$= \mathtt{E}x_0 + \int_{t_0}^{t} A(\tau)\mathtt{E}x(\tau) + B(\tau)u(\tau)\,\mathrm{d}\tau = m_x(t) \tag{118}$$

Thus, the mean $m_x(t)$, satisfies the ordinary differential equation

$$\frac{\partial m_x(t)}{\partial t} = Am_x(t) + B(t)u(t), \quad m_x(t_0) = m_0. \tag{119}$$

Linear Stochastic Differential Equations

To compactly define the auto-covariance and covariance function of the solution, x, we need the state-transition matrix Φ:

$$\frac{\partial \Phi(t;t_0)}{\partial t} = A(t)\Phi(t;t_0), \quad \Phi(t_0;t_0) = I. \tag{120}$$

Let $R(s,t)$ denote the auto-covariance of x, if $s \geq t$ then

$$R(s,t) = \texttt{Cov}\Big[x(s), x(t)\Big] = \Phi(s,t)P(t) \tag{121}$$

where $P(t)$ is the covariance of $x(t)$.

The resulting variance $P(t)$ is given by the ordinary differential equation

$$\frac{\partial P(t)}{\partial t} = A(t)P(t) + P(t)A(t)^T + G(t)G(t)^T, \quad P(t_0) = R_0, \tag{122}$$

which can be derived in similar way as for the mean (though requiring a few more calculations).

Sampling a Linear Stochastic Differential Equation

We will now consider the process described by the equations

$$\mathrm{d}x(t) = A(t)x(t)\,\mathrm{d}t + B(t)\,\mathrm{d}w(t) \tag{123a}$$

$$y(t) = C(t)x(t) + D(t)u(t) + F(t)e(t), \quad e(t) \sim F(m_e, P_e) \tag{123b}$$

where $w(t)$ is a Brownian motion, and $e(t)$ is following some distribution F. Integrating the above system over a single period with constant input yields

$$x(t_{k+1}) = \Phi(t_{k+1}; t_k)x(t_k) + \int_{t_k}^{t_{k+1}} \Phi(\tau; t_k)B(\tau)\,\mathrm{d}\tau u(t_k) + \tilde{w}(t_k) \tag{124a}$$

$$y(t_{k+1}) = C(t_k)x(t_k) + D(t_k)u(t_k) + F(t_k)e(t_k), \quad e(t_k) \sim F(m_e, P_e) \tag{124b}$$

where $\Phi(t_{k+1}; t_k)$ is the state-transition matrix from t_k to t_{k+1}.

Sampling a Linear Stochastic Differential Equation

The new discrete random input $\tilde{w}(t_k)$ satisfies

$$\tilde{w}(t_k) = \int_{t_k}^{t_{k+1}} \Phi(\tau; t_k) G(\tau) \, \mathrm{d}w(\tau) \tag{125a}$$

with the moments:

$$\mathtt{E}\{\tilde{w}(t_k)\} = 0 \tag{126a}$$

$$\mathtt{E}\{\tilde{w}(t_k)\tilde{w}(t_k)^T\} = \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}; \tau) G(\tau) G(\tau)' \Phi(t_{k+1}; \tau)' \, \mathrm{d}\tau \tag{126b}$$

If we compare with the first approach for a stochastic description:

$$\dot{x}_t = A_t x_t + B_t u_t + G_t w_t, \quad w_t \in N(0, R_1) \tag{127}$$

The resulting discretization is almost identical, with the difference being

$$\mathtt{E}\{\tilde{w}(t_k)\tilde{w}(t_k)^T\} = \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}; \tau) G(\tau) R_1 G(\tau)' \Phi(t_{k+1}; \tau)' \, \mathrm{d}\tau \tag{128}$$

Computing the Discretization of Stochastic LTI Systems

We have now seen how to discretize an SDE, but let us now consider how to actuallu do the discretization computationally.

Let us consider the continuous-time stochastic LTI systems given by:

$$\dot{x}_t = Ax_t + Bu_t + Gw_t, \quad w_t \in N(0, R_1) \tag{129}$$

$$y_t = Cx_t + Du_t + Fe_t, \quad e_t \in N(0, R_2) \tag{130}$$

In discrete-time, it is given by

$$x_{k+1} = A_d x_k + B_d u_k + w_k, \quad w_k \in N(0, R_{1d}) \tag{131}$$

$$y_k = C_d x_k + D_d u_k + e_k, \quad e_k \in N(0, R_{2d}) \tag{132}$$

As always, the output equation is not discretized:

$$C_d = C, \quad D_d = D, \quad R_{2d} = FR_2F^T \tag{133}$$

Computing the Discretization of Stochastic LTI Systems

From the previous discussion we can derive the discrete-time system as:

$$x_{k+1} = A_d x_k + B_d u_k + w_k \tag{134}$$

$$A_d = e^{A(T_s)} \tag{135}$$

$$B_d = \int_{kT_s}^{(k+1)T_s} e^{A((k+1)T_s - s)} B ds \tag{136}$$

$$w_k = \int_{kT_s}^{(k+1)T_s} e^{A((k+1)T_s - s)} G w_s ds \tag{137}$$

with

$$E\{w_k\} = \int_{kT_s}^{(k+1)T_s} e^{A((k+1)T_s - s)} G E\{w_s\} ds = 0 \tag{138}$$

$$Var\{w_k\} = \int_{kT_s}^{(k+1)T_s} e^{A((k+1)T_s - s)} G R_1 G^T e^{A^T((k+1)T_s - s)} ds \tag{139}$$

Computing the Discretization of Stochastic LTI Systems

In the discussion on discretization of deterministic systems, the controlled part was obtained by

$$\begin{bmatrix} A_d & B_d \\ 0 & I \end{bmatrix} = \exp\left(\begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} T_s \right) \tag{140}$$

For stochastic systems, we can use a similar trick:

$$\begin{bmatrix} A_d & \tilde{R}_1 \\ 0 & A_d^{-T} \end{bmatrix} = \exp\left(\begin{bmatrix} A & GR_1G^T \\ 0 & -A^T \end{bmatrix} T_s \right) \qquad R_{1d} = \tilde{R}_1 A_d^T \tag{141}$$

$$\begin{bmatrix} A_d^{-1} & \tilde{R}_1^T \\ 0 & A_d^T \end{bmatrix} = \exp\left(\begin{bmatrix} -A & GR_1G^T \\ 0 & A^T \end{bmatrix} T_s \right) \qquad R_{1d} = A_d \tilde{R}_1^T \tag{142}$$

Both approaches are equivalent.

Proof of discretization trick

From the form of the trick, we can derive 3 differential equations:

$$W = \exp\left(\begin{bmatrix} F & G \\ 0 & H \end{bmatrix} t\right) = \begin{bmatrix} x & y \\ 0 & z \end{bmatrix}, \quad \dot{W} = CW \tag{143}$$

$$\dot{x} = Fx, \quad \dot{z} = Hx \tag{144}$$

$$\dot{y} = Fy + Gz \tag{145}$$

We know the solution to these are

$$x = e^{Ft}x_0, \quad z = e^{Ht}z_0 \tag{146}$$

$$y = e^{Ft}y_0 + \int_0^t e^{F(t-s)}Ge^{Hs}z_0 ds \tag{147}$$

Set the initial value as $x_0 = I, y_0 = 0, z_0 = I$ and rearrange:

$$x = e^{Ft}, \quad z = e^{Ht} \tag{148}$$

$$y = e^{Ft}\int_0^t e^{-Fs}Ge^{Hs}ds \tag{149}$$

Proof of discretization trick

We can now show that the computed matrices gives the discrete variance, we are after.
Let us define $F = -A$, $H = A^T$, and $G = R_1$, then the solutions at $t = T_s$ becomes

$$x = e^{-AT_s} = A_d^{-1} \tag{150}$$

$$z = e^{A^T T_s} = A_d^T \tag{151}$$

$$y = e^{-AT_s} \int_0^{T_s} e^{As} R_1 e^{A^T s} ds = A_d^{-1} R_{1d} \tag{152}$$

With the resulting definition of the variance:

$$R_{1d} = \int_0^{T_s} e^{As} R_1 e^{A^T s} ds = z^T y \tag{153}$$

For the other approach, the proof is similar, but has one more step in the end: a variable change of the integral.

Questions

Questions?