ELEC-E8107 Stochastics models, estimation and Control (5 p)

Applications of the **state** estimation: **Kalman** filtering

- Tracking/ surveillance
- Control systems, power systems, failure detection
- Navigation and trajectory determination, robotics
  - Probabilistic robotics
- Signal processing, image processing, communication, operations research, econometric systems
- Remote sensing, geophysical research, Biomedical systems

Duality of estimation and control, stochastic optimal control

Course arrangements

- Lectures: Arto.Visala@aalto.fi on Tuesday 14:15-16:00
- Exercises and home assignments: Issouf.ouattara @aalto.fi on Tuesdays 8:15-10:00
- Slides mainly from textbook of Bar-Shalom et al
- In autumn 2022, evaluation consists of Exam (100%)
- In exam you can get extra points corresponding one grade by submitting in time valid Home assignments given in connection of exercises.
- It is possible to get the excellent grade without Home assignments, but it is not easy.

Field robotics; autonomous vehicles Data from ATV platforms used is in examples

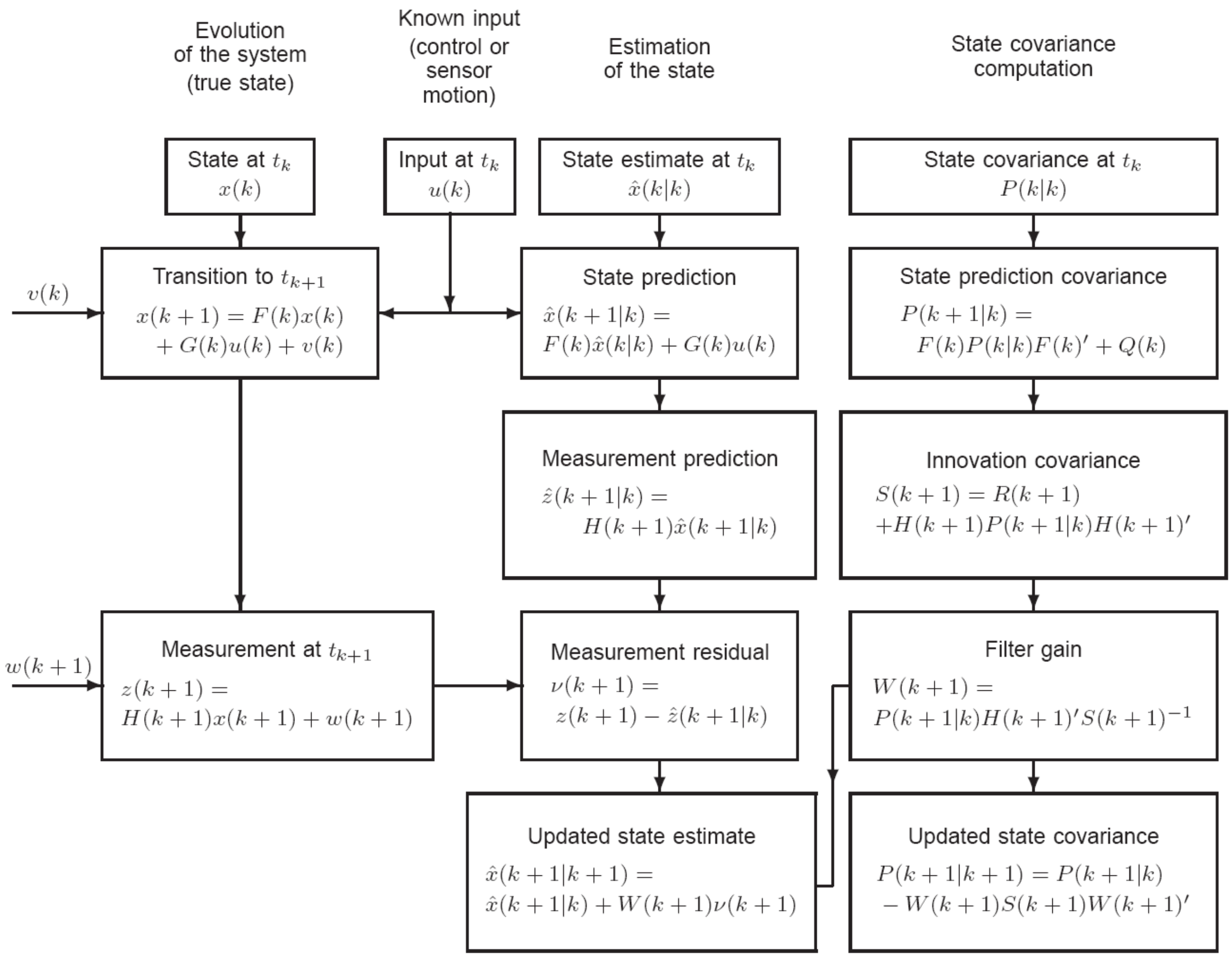
Aalto University

ELEC-E8107 Stochastics models, estimation and control
Lecture 1: Statistics and stochastics

- Gradient, Jacobian and Hessian; Eigenvalues, Eigenvectors, and Quadratic Forms
- Gaussian Random Variables; pdf, Mean and Covariance; Joint and Conditional Random Variables
- Fundamental Equations of Linear Estimation
- Stochastic processes; Correlation; White noise process
- Random Sequences, Markov Processes; Markov Sequences and Markov Chains

Statistics and stochastics in Kalman filter

Optimal fusion of Information from **Dynamic Model** and **Measurements**



Gradient, Jacobian, Hessian

Gradient operator of a scalar function $\nabla_x = \left[\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n}\right]'$

If Quadratic form $\nabla_x(x'Ax) = 2Ax$

For a vector function

$$\nabla_x f(x)' = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} [f_1(x) \cdots f_m(x)] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Jacobian matrix

$$f_x(x) \triangleq \frac{\partial f}{\partial x} = [\nabla_x f(x)']' = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Hessian

- Hessian of the scalar function

$$\phi_{xx}(x) \triangleq \frac{\partial^2 \phi(x)}{\partial x^2} = \nabla_x \nabla_x' \phi(x) = \begin{bmatrix} \frac{\partial^2 \phi}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 \phi}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \phi}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 \phi}{\partial x_n \partial x_n} \end{bmatrix}$$

Eigenvalues, Eigenvectors, and Quadratic Forms

A nxn matrix, eigenvectors and –values

$$Au_i = \lambda_i u_i \qquad i = 1, \ldots, n$$

Determinant
$$|A| = \prod_{i=1}^{n} \lambda_i$$

Trace
$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

Quadratic form
$$q = x' A x$$

A positive (semi)definite quadratic form

$$A > 0 \quad \Longleftrightarrow \quad x'Ax > 0 \quad \forall x \neq 0$$

$$A \geq 0 \quad \Longleftrightarrow \quad x'Ax \geq 0 \quad \forall x \neq 0$$

The inequality of two matrices is defined as follows:

the matrix A is smaller (not larger) than the matrix B if and only if the difference B − A is positive (semi) definite

Gaussian pdf, random variable, mean and covariance

- Scalar Gaussian or Normal pdf

$$p(x) = \mathcal{N}(x; \bar{x}, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

$$\boxed{x \sim \mathcal{N}(\bar{x}, \sigma^2)}$$

- Vector Gaussian pdf

$$\boxed{\mathcal{N}(x; \bar{x}, P) \triangleq |2\pi P|^{-1/2} e^{-\frac{1}{2}(x-\bar{x})' P^{-1} (x-\bar{x})}}$$

- Mean or expected value

$$\bar{x} = E[x]$$

- Covariance P

$$P = E[(x - \bar{x})(x - \bar{x})']$$

Mean, average, first moment

$$E[x] = \int_{-\infty}^{\infty} x p(x)\, dx \triangleq \bar{x}$$

The nth moment is

$$E[x^n] = \int_{-\infty}^{\infty} x^n p(x)\, dx$$

The second central moment or variance

$$\mathrm{var}(x) \triangleq E[(x - \bar{x})^2] =$$

$$\int_{-\infty}^{\infty} (x - \bar{x})^2 p(x)\, dx = E[x^2] - (\bar{x})^2 \triangleq \sigma_x^2$$

Mean Square MS

$$E[x^2] = [E(x)]^2 + \mathrm{var}(x) = \bar{x}^2 + \sigma_x^2$$

The Central Limit Theorem

If the sequence x_i, $i = 1, \ldots,$ consists of independent random variables, then under some reasonably mild conditions the pdf of the sum

$$z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i$$

will tend to a Gaussian pdf as $n \to \infty$.

The Law of Large Numbers

states loosely that the sum of a large number of random variables tends, under some fairly nonrestrictive conditions, to its expected value.

Joint and Conditional Gaussian random variables, Conditional pdf of x given z.

$$y = \begin{bmatrix} x \\ z \end{bmatrix} \qquad p(x,z) = p(y) = \mathcal{N}(y; \bar{y}, P_{yy}) \qquad \bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix}$$

$$\begin{aligned} P_{xx} &= \mathrm{cov}(x) = E[(x-\bar{x})(x-\bar{x})'] \\ P_{zz} &= \mathrm{cov}(z) = E[(z-\bar{z})(z-\bar{z})'] \\ P_{xz} &= \mathrm{cov}(x,z) = E[(x-\bar{x})(z-\bar{z})'] = P'_{zx} \end{aligned} \qquad P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}$$

$$p(x|z) = \frac{p(x,z)}{p(z)} = \frac{|2\pi P_{yy}|^{-1/2}\, e^{-\frac{1}{2}(y-\bar{y})' P_{yy}^{-1} (y-\bar{y})}}{|2\pi P_{zz}|^{-1/2}\, e^{-\frac{1}{2}(z-\bar{z})' P_{zz}^{-1} (z-\bar{z})}}$$

New zero mean random variables, the exponent q of Gaussian conditional pdf becomes

$$\xi \triangleq x - \bar{x}$$
$$\zeta \triangleq z - \bar{z}$$

$$\begin{aligned} q &= \begin{bmatrix} \xi \\ \zeta \end{bmatrix}' \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}^{-1} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \zeta' P_{zz}^{-1} \zeta \\ &= \begin{bmatrix} \xi \\ \zeta \end{bmatrix}' \begin{bmatrix} T_{xx} & T_{xz} \\ T_{zx} & T_{zz} \end{bmatrix} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} - \zeta' P_{zz}^{-1} \zeta \end{aligned}$$

continues, matrix inversion lemma is utilized; trick $TT^{-1}=I$

$$T_{xx}^{-1} = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}$$

$$P_{zz}^{-1} = T_{zz} - T_{zx}T_{xx}^{-1}T_{xz}$$

$$T_{xx}^{-1}T_{xz} = -P_{xz}P_{zz}^{-1}$$

$$\begin{aligned} q &= \xi' T_{xx}\xi + \xi' T_{xz}\zeta + \zeta' T_{zx}\xi + \zeta' T_{zz}\zeta - \zeta' P_{zz}^{-1}\zeta \\ &= (\xi + T_{xx}^{-1}T_{xz}\zeta)' T_{xx}(\xi + T_{xx}^{-1}T_{xz}\zeta) + \zeta'(T_{zz} - T_{zx}T_{xx}^{-1}T_{xz})\zeta - \zeta' P_{zz}^{-1}\zeta \\ &= (\xi + T_{xx}^{-1}T_{xz}\zeta)' T_{xx}(\xi + T_{xx}^{-1}T_{xz}\zeta) \end{aligned} \qquad (1.4.14\text{-}15)$$

completion of the squares in exponent of the conditional Gaussian pdf

$$\xi + T_{xx}^{-1}T_{xz}\zeta = x - \bar{x} - P_{xz}P_{zz}^{-1}(z - \bar{z})$$

$$\boxed{E(x|z) \triangleq \hat{x} = \bar{x} + P_{xz}P_{zz}^{-1}(z - \bar{z})}$$

$$\boxed{\mathrm{cov}(x|z) \triangleq P_{xx|z} = T_{xx}^{-1} = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}}$$

Fundamental equations of linear estimation

Estimation of Gaussian random vectors

x and z jointly Gaussian

z is the measurement

x random variable to be estimated

$$y \triangleq \begin{bmatrix} x \\ z \end{bmatrix} \qquad \bar{y} = \begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} \qquad P_{yy} = \begin{bmatrix} P_{xx} & P_{xz} \\ P_{zx} & P_{zz} \end{bmatrix}$$

$$y \sim \mathcal{N}[\bar{y}, P_{yy}] \qquad P_{xx} = E[(x-\bar{x})(x-\bar{x})'] \qquad P_{xz} = E[(x-\bar{x})(z-\bar{z})'] = P'_{zx}$$

Next slide set: The **MMSE Minimum Mean Square Error** –estimator is the conditional mean of x given z, for linear Gaussian case,

also **Maximum a Posteriori MAP** -estimator

$$\boxed{\hat{x} \triangleq E[x|z] = \bar{x} + P_{xz}P_{zz}^{-1}(z-\bar{z})}$$

$$\boxed{P_{xx|z} \triangleq E[(x-\hat{x})(x-\hat{x})'|z] = P_{xx} - P_{xz}P_{zz}^{-1}P_{zx}}$$

Fundamental equations of linear estimation - Interpretations

- *A priori* estimate is updated/corrected on the basis of measurement information in calculation of *a posteriori* estimate
- Correction gain depends directly on P_{xz} the crosscovariance between x and measurement z. P_{xz} must be unzero in order measurements contain information in general about the state x.
- Correction effect depends inversely proportional on P_{zz} . The better measurements, the ’smaller’ covariance, the bigger the correction gain.

Random process, stochastic process

A ***scalar random variable*** is a (real) number x determined by the outcome ω of a random experiment

$$x = x(\omega)$$

A (scalar) ***random process or a stochastic process*** is a function of time determined by the outcome of a random experiment

$$x(t) = x(t, \omega)$$

This is a family or ensemble of functions of time, in general different for each outcome ω

mean or ***ensemble average***

$$\bar{x}(t) = E[x(t)] = \int_{-\infty}^{\infty} \xi \, p_{x(t)}(\xi) \, d\xi$$

its ***autocorrelation***

$$R(t_1, t_2) \triangleq E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi\eta \, p_{x(t_1),x(t_2)}(\xi, \eta) \, d\xi d\eta$$

autocovariance of this random process

$$\begin{aligned} V(t_1, t_2) &\triangleq E[[x(t_1) - \bar{x}(t_1)][x(t_2) - \bar{x}(t_2)]] \\ &= R(t_1, t_2) - \bar{x}(t_1)\bar{x}(t_2) \end{aligned}$$

Stationarity process

$$R(t_1, t_2) = R(t_1 - t_2) = R(\tau)$$

The ***power spectrum*** or ***power spectral density*** of a *stationary* random process is the Fourier transform

$$S(\omega) = \mathcal{F}\{R(\tau)\} = \int_{-\infty}^{\infty} e^{-j\omega\tau} R(\tau) \, d\tau$$

White Noise

A (not necessarily stationary) random process whose autocovariance is zero for any two different times is called ***white noise***

$$V(t_1, t_2) = \sigma^2(t_1)\delta(t_1 - t_2) = S_0(t_1)\delta(t_1 - t_2)$$

$\delta(\cdot)$ the ***Dirac (impulse) delta function***

A ***stationary zero-mean white*** process

$$R(\tau) = E[x(t+\tau)x(t)] = S_0\delta(\tau)$$

A ***nonstationary zero-mean white*** process $x(t)$

$$E[x(t_1)x(t_2)] = S_0(t_1)\delta(t_1 - t_2)$$

Random Walk and the Wiener Process

The ***Wiener random process*** (or Wiener-Levy or Brownian motion) is a limiting form of the ***random walk***: the sum of independent steps of size $s \to 0$, equiprobable in each direction, taken at intervals $\Delta \to 0$ such that

$$\frac{s}{\sqrt{\Delta}} \to \sqrt{\alpha}$$

This yields a stochastic process **w**(t) with the following pdf [assuming **w**(0) = 0],

$$p[\mathbf{w}(t)] = \mathcal{N}[\mathbf{w}(t); 0, \alpha t]$$

the ***Wiener process*** is *nonstationary.* It relates to the zero-mean white noise, denoted here as $n(t)$, as follows

$$\mathbf{w}(t) = \int_0^t n(\tau)\, d\tau \qquad E[n(t_1)n(t_2)] = \alpha\delta(t_1 - t_2)$$

$$d\mathbf{w}(t) = n(t)\, dt$$

Stochastic sequences, Markov property

Markov processes are defined by ***Markov property***

$$p[x(t)|x(\tau), \tau \leq t_1] = p[x(t)|x(t_1)] \qquad \forall t > t_1$$

the past up to any t_1 is *fully characterized* by the value of the process at t_1

"The future is independent of the past if the present is known."

The Wiener process is Markov

Furthermore, the state *x*(*t*) of a (possibly time-varying) dynamic system driven by white noise *n*(*t*), is a Markov process

$$\dot{x}(t) = f[t, x(t), n(t)]$$

Random Sequences and Markov Sequences

Random Sequences,

$$X^k = \{x(j)\}_{j=1}^{k} \qquad k = 1, 2, \dots$$

a random sequence is ***Markov Sequences*** if

$$p[x(k)|X^j] = p[x(k)|x(j)] \qquad \forall k > j$$

The (real-valued) *zero-mean sequence v*(*j*), *j* = 1, . . . , is a ***discrete-time white noise*** (a ***white sequence***) if

$$E[v(k)v(j)] = q(k)\delta_{kj}$$

where the ***Kronecker delta function***

$$\delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases}$$

q(*k*) denotes its variance, *q*(*k*) =q

stationaty

The state *x*(*k*) of a dynamic system excited by white noise *v*(*k*)

$$x(k+1) = f[k, x(k), v(k)]$$

is a ***discrete-time Markov process*** or ***Markov sequence***. In general, both *x*(*k*) and *v*(*k*) are vector-valued.

The state of a linear dynamic system excited by white Gaussian noise

$$x(k+1) = F(k)x(k) + v(k)$$

is a ***Gauss-Markov sequence***

A special case, for a scalar *x*,

$$x(k+1) = x(k) + v(k)$$

x becomes the *integral (sum)* of the white noise sequence terms, and is called a ***discrete-time Wiener process***

Markov Chains

A ***Markov chain*** is a special case of a Markov sequence, in which the **state space** is **discrete** and **finite**

$$x(k) \in \{x_i,\, i = 1, \ldots, n\}$$

its characterization is given in full by the ***transition (jump) probabilities***

$$P\{x(k) = x_j \,|\, x(k-1) = x_i\} \triangleq \pi_{ij} \qquad i, j = 1, \ldots, n$$

Define the vector

$$\mu(k) \triangleq [\mu_1(k), \ldots, \mu_n(k)]'$$

$$\mu_i(k) \triangleq P\{x(k) = x_i\}$$

where the components are the probabilities of the chain being in state *i.*

The evolution in time of (1.4.22-11) is then given by

$$\mu_i(k+1) = \sum_{j=1}^{n} \pi_{ji}\mu_j(k) \qquad i = 1, \ldots, n$$

$$\mu(k+1) = \Pi'\mu(k)$$

with matrix notation $\Pi = [\pi_{ij}]$

the ***transition matrix of the Markov chain***

Markov property continues

discrete-time white noise (a white sequence) scalar

$$E[v(k)v(j)] = q(k)\delta_{kj} \qquad (1.4.22\text{-}3)$$

Kronecker delta function

$$\delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \qquad (1.4.22\text{-}4)$$

The state of linear dynamic equation with white noise process

$$\boxed{E[v(k)] = 0} \qquad (4.3.1\text{-}6)$$

$$\boxed{E[v(k)v(j)'] = Q(k)\delta_{kj}} \qquad (4.3.1\text{-}7)$$

$$\boxed{x(k+1) = F(k)x(k) + G(k)u(k) + v(k)} \qquad (4.3.1\text{-}9)$$

has the following analytical solution, next slide

The State as a Markov Process

$$x(k) = [\prod_{j=0}^{k-l-1} F(k-1-j)]x(l) + \sum_{i=l}^{k-1}[\prod_{j=0}^{k-i-2} F(k-1-j)][G(i)u(i) + v(i)] \quad (4.3.3\text{-}1)$$

Thus, since $v(i)$, $i = l, \ldots, k-1$, are independent of

$$X^l \triangleq \{x(j)\}_{j=0}^{l} \quad (4.3.3\text{-}2)$$

which depend only on $v(i)$, $i = 0, \ldots, l-1$, one has

$$p[x(k)|X^l, U^{k-1}] = p[x(k)|x(l), U_l^{k-1}] \qquad \forall k > l \quad (4.3.3\text{-}3)$$

$$U_l^{k-1} \triangleq \{u(j)\}_{j=l}^{k-1} \quad (4.3.3\text{-}4)$$

Thus, **the state vector is** a Markov process, or, more correctly, a **Markov sequence**.

State of a stochastic system described by a Markov process — **summarizes probabilistically its past**.