TECHNISCHE UNIVERSITÄT DARMSTADT

DSP

Prof. Dr.-Ing. Abdelhak M. Zoubir

SIGNAL PROCESSING GROUP

Copyright

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline I

TECHNISCHE
UNIVERSITÄT
DARMSTADT

1. **Random Variables and Stochastic Processes**
   - Random Variables
   - Operations on Random Variables
   - Expected Value and Variance
   - Covariance
   - Random Processes
   - Stationarity and Ergodicity
   - Second-Order Moment Function
   - Power Spectral Density
   - Linear filtering
   - Covariance Function and Spectrum
   - Interaction of Two Linear Systems
   - Linear filtered process plus noise

Outline II

TECHNISCHE UNIVERSITÄT DARMSTADT

2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation
8. Application: Airplane Tracking using Kalman Filter
9. Discrete Kalman Filter

Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

- A random variable is a number $X(\zeta)$ assigned to every outcome $\zeta \in \mathcal{S}$ of an experiment.
- Examples: the gain in a game of chance, the voltage of a random source, the cost of a random component.
- Given $\zeta \in \mathcal{S}$, the mapping $\zeta \to X(\zeta) \in \mathbb{R}$ is a random variable, if for all $x \in \mathbb{R}$ the set $\{\zeta : X(\zeta) \leq x\}$ is also an event

$$\{X \leq x\} \triangleq \{\zeta : X(\zeta) \leq x\}$$

with probability

$$\mathrm{P}\left[\{X \leq x\}\right] \triangleq \mathrm{P}\left[\{\zeta : X(\zeta) \leq x\}\right]$$

Distribution Functions

TECHNISCHE UNIVERSITÄT DARMSTADT

- The mapping $x \in \mathbb{R} \to F_X(x) = \mathrm{P}(\{X \leq x\})$ is called the probability distribution function of the random variable X.
- The probability distribution function has the following properties
  1. $$0 \leq F_X(x) \leq 1$$ $$F_X(\infty) = 1$$ $$F_X(-\infty) = 0$$
  2. $F_X(x)$ is continuous from the right i.e., $$\lim_{\epsilon \to 0} F_X(x+\epsilon) = F_X(x), \quad \epsilon > 0$$

Distribution Functions

TECHNISCHE UNIVERSITÄT DARMSTADT

3. If $x_1 < x_2$ then $F_X(x_1) \leq F_X(x_2)$, i.e., $F_X(x)$ is a non-decreasing function of x

$$\mathrm{P}\left[\{x_1 < X \leq x_2\}\right] = F_X(x_2) - F_X(x_1) \geq 0$$

4. The probability density function is defined as:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

and if it exists has the property

$$f_X(x) \geq 0.$$

Distribution Functions

To find $F_X(x)$, we integrate $f_X(x)$, i.e.,

$$\mathrm{P}\left(\{X \leq x\}\right) = F_X(x) = \int_{-\infty}^{x} f_X(u) du$$

and because $F_X(\infty) = 1$,

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

Further,

$$\mathrm{P}\left(\{x_1 < X \leq x_2\}\right) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\ dx$$

Distribution Functions

TECHNISCHE UNIVERSITÄT DARMSTADT

5. For a continuous variable X, $\mathrm{P}(x_1 < X < x_2) = \mathrm{P}(x_1 < X \leq x_2)$ because

$$\mathrm{P}(\{x\}) = \mathrm{P}\left((-\infty, x]\right) - \mathrm{P}\left((-\infty, x)\right) = F_X(x) - F_X(x-0) = 0$$

Uniform Distribution

- X is said to be uniformly distributed on $[a, b)$ if it admits the probability density function

$$f_X(x) = \frac{1}{b-a}\left[u(x-a) - u(x-b)\right], \; a < b$$

- The probability distribution function is

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u)du \\ &= \frac{1}{b-a}\left[(x-a)u(x-a) - (x-b)u(x-b)\right] \end{aligned}$$

Uniform Distribution

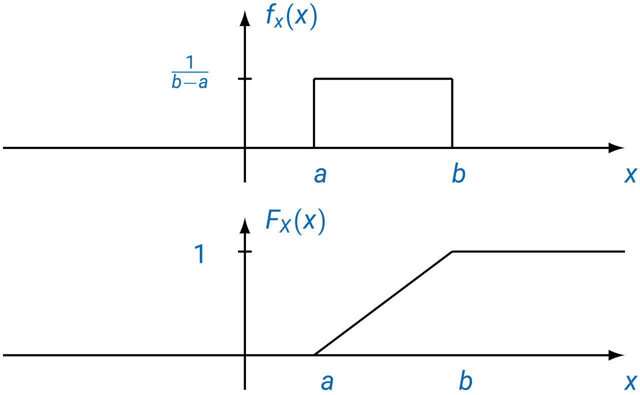


Figure: Uniform Distribution

Normal Distribution

TECHNISCHE UNIVERSITÄT DARMSTADT

- X is normally distributed with mean μ and variance σ^2 $(X \sim \mathcal{N}(\mu, \sigma^2))$ if

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad \sigma^2 > 0, \ -\infty < \mu < \infty$$

- A closed-form expression for $F_X(x)$ does not exist. For $\mu = 0, \ \sigma^2 = 1, \ X$ is said to be standard normally distributed, and its distribution function is denoted by $\Phi_X(x)$, i.e., we have

$$F_X(x) = \Phi_X\left(\frac{x-\mu}{\sigma}\right)$$

Normal Distribution

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example:
  Show that for the normal distribution $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx$$

If we let $s = \frac{x-\mu}{\sigma}$, then

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2}\, ds$$

Normal Distribution

TECHNISCHE UNIVERSITÄT DARMSTADT

Because $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} > 0$, we calculate

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \right)^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\, dx\, dy \\
&= \int_{0}^{2\pi} \int_{0}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r\, dr\, d\varphi \\
&= \int_{0}^{\infty} e^{-\frac{1}{2}r^2} r\, dr \\
&= \int_{0}^{\infty} e^{-s}\, ds \\
&= 1
\end{aligned}$$

where $r = \sqrt{x^2 + y^2}$ and $\varphi = arctan\left(\frac{y}{x}\right)$.

Multiple Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

The Joint Distribution Function and its Properties

- The probabilities of two events $A = \{X_1 \leq x_1\}$ and $B = \{X_2 \leq x_2\}$ have already been defined as functions of x_1 and x_2, respectively called probability distribution functions

$$
\begin{aligned}
F_{X_1}(x_1) &= \mathrm{P}\left[\{X_1 \leq x_1\}\right] \\
F_{X_2}(x_2) &= \mathrm{P}\left[\{X_2 \leq x_2\}\right]
\end{aligned}
$$

- We introduce a new concept to include the probability of the joint event $\{X_1 \leq x_1, X_2 \leq x_2\} = \{(X_1, X_2) \in D\}$, where x_1 and x_2 are two arbitrary real numbers and D is the quadrant shown in Figure 2.

Multiple Random Variables

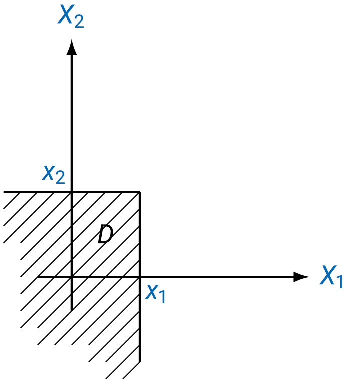


Figure: The quadrant D of two arbitrary real numbers x and y.

Multiple Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

- The joint distribution $F_{X_1X_2}(x_1, x_2)$ of two random variables X_1 and X_2 is the probability of the event $\{X_1 \leq x_1, X_2 \leq x_2\}$,

$$F_{X_1X_2}(x_1, x_2) = \mathrm{P}\left(\{X_1 \leq x_1, X_2 \leq x_2\}\right).$$

- For n random variables $X_1, \ldots, X_n$ the joint distribution function is denoted by

$$F_{X_1X_2\ldots X_n}(x_1, \ldots, x_n) = \mathrm{P}\left(\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n\}\right)$$

- For now, only bivariate random variables, i.e. the case $n = 2$ is considered

Bivariate Distributions

TECHNISCHE UNIVERSITÄT DARMSTADT

The bivariate probability distribution function $F_{X_1X_2}(x_1, x_2)$ has the following properties:

1) $F_{X_1X_2}(x_1, x_2)$ is bounded between 0 and 1, especially

$$
\begin{aligned}
\lim_{x_1 \to \infty, x_2 \to \infty} F_{X_1X_2}(x_1, x_2) &= 1 \\
\lim_{x_1 \to -\infty} F_{X_1X_2}(x_1, x_2) = \lim_{x_2 \to -\infty} F_{X_1X_2}(x_1, x_2) &= 0
\end{aligned}
$$

2) $F_{X_1X_2}(x_1, x_2)$ is continuous from the right in each component when the other component is fixed, i.e.

$$
\lim_{\epsilon \to 0} F_{X_1X_2}(x_1 + \epsilon, x_2) = \lim_{\epsilon \to 0} F_{X_1X_2}(x_1, x_2 + \epsilon) = F_{X_1X_2}(x_1, x_2), \quad \epsilon > 0
$$

Bivariate Distributions

3) The event $\{x_{11} < X_1 \leq x_{12}, X_2 \leq x_2\}$ consists of all points (X_1, X_2) in the vertical half-strip D_1, and the event $\{X_1 \leq x_1, x_{21} < X_2 \leq x_{22}\}$ consists of all points in the horizontal half-strip D_2. See Figure 3.
Thus, we have

$$P(\{x_{11} < X_1 \leq x_{12}, X_2 \leq x_2\}) = F_{X_1X_2}(x_{12}, x_2) - F_{X_1X_2}(x_{11}, x_2)$$

$$P(\{X_1 < x_1, x_{21} < X_2 \leq x_{22}\}) = F_{X_1X_2}(x_1, x_{22}) - F_{X_1X_2}(x_1, x_{21})$$

Bivariate Distributions

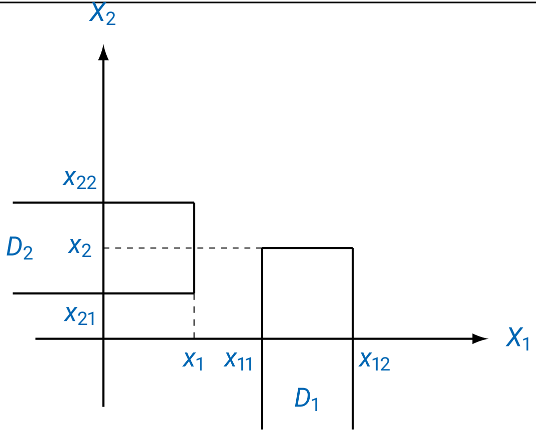


Figure: Two half-strips D_1 and D_2 of a bivariate random variable (X_1, X_2).

4) The event $\{x_{11} < X_1 \leq x_{12}, x_{21} < X_2 \leq x_{22}\}$ consists of all points (X_1, X_2) in the rectangle D_3. See Figure 4.
Therefore we have,

$$\mathrm{P}(\{x_{11} < X_1 \leq x_{12}, x_{21} < X_2 \leq x_{22}\}) = F_{X_1X_2}(x_{12}, x_{22}) - F_{X_1X_2}(x_{11}, x_{22}) - F_{X_1X_2}(x_{12}, x_{21}) + F_{X_1X_2}(x_{11}, x_{21}) > 0$$

Bivariate Distributions

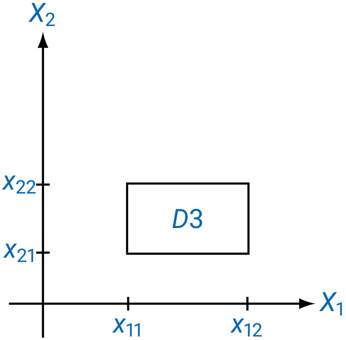


Figure: Probability that (X_1, X_2) is in the rectangle D_3.

Bivariate Distributions

5) The joint probability density function of $X_1$and X_2 is defined as

$$f_{X_1X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1X_2}(x_1, x_2)}{\partial x_1 \, \partial x_2}$$

and if it exists has the property $f_{X_1X_2}(x_1, x_2) \geq 0$.
To find $F_{X_1X_2}(x_1, x_2)$, we can integrate $f_{X_1X_2}(x_1, x_2)$,

$$F_{X_1X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1X_2}(u_1, u_2) du_1 du_2.$$

Because $F_{X_1X_2}(\infty,\infty)=1$, we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)dx_1dx_2 = 1.$$

The probability of the event $\{x_{11} < X_1 \leq x_{12}, x_{21} < X_2 \leq x_{22}\}$, i.e. are all points in the rectangle D_3 in Figure 4, can be also calculated using

$$\mathrm{P}(\{x_{11} < X_1 \leq x_{12}, x_{21} < X_2 \leq x_{22}\}) = \int_{x_{11}}^{x_{12}}\int_{x_{21}}^{x_{22}} f_{X_1X_2}(x_1,x_2)dx_1dx_2.$$

Bivariate Distributions

6) The marginal probability distribution functions can be expressed in terms of the joint probability distribution function

$$
\begin{aligned}
F_{X_1}(x_1) &= \lim_{x_2 \to \infty} F_{X_1 X_2}(x_1, x_2) \\
F_{X_2}(x_2) &= \lim_{x_1 \to \infty} F_{X_1 X_2}(x_1, x_2).
\end{aligned}
$$

Similarly, the marginal probability density function can be expressed in terms of the joint probability density functions,

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_2$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1 X_2}(x_1, x_2) dx_1.$$

Operations on Random Variables

- The previously introduced random variable concept is a means of defining events on a sample space.
- It forms a mathematical model for describing characteristics of some real, physical world random variables, which are mostly based on a single concept – expectation.

Operations on Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

Expectation

- The mathematical expectation of a continuous random variable X,

$$\mathrm{E}[X]$$

which may be read "the expected value of X" or "the mean value of X" is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

where $f_X(x)$ is the probability density function of X.

Operations on Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

- If X happens to be discrete with N possible values x_i having probabilities $\mathrm{P}(x_i)$ of occurrence, then

$$
\begin{aligned}
f_X(x) &= \sum_{i=1}^{N} \mathrm{P}(x_i) \cdot \delta(x - x_i) \\
\mathrm{E}[X] &= \sum_{i=1}^{N} x_i \mathrm{P}(x_i)
\end{aligned}
$$

- Example: Normal Distribution
  If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$
f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2}(x-\mu)^2/\sigma^2\right\}
$$

Operations on Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot \exp\{-\frac{1}{2}(x-\mu)^2/\sigma^2\} dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu) \cdot \exp\{-\frac{1}{2}u^2\} \sigma du \\
&= \frac{1}{\sqrt{2\pi}} \left[\sigma \int_{-\infty}^{\infty} u \cdot \exp\{-\frac{1}{2}u^2\} du + \mu \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}u^2\} du \right] \\
&= \frac{1}{\sqrt{2\pi}} \left[\sigma \cdot 0 + \mu\sqrt{2\pi} \right] \\
&= \mu
\end{aligned}
$$

Expected Value of a Function of Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

- Suppose we are interested in the mean of the random variable $Y = g(X)$

$$\mathrm{E}\left[Y\right] = \int_{-\infty}^{\infty} y \cdot f_Y(y) dy$$

and we are given $f_X(x)$. Then,

$$\mathrm{E}\left[Y\right] = \mathrm{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx$$

Expected Value of a Function of Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example: The average power in a 1 Ω resistor

$$\mathrm{E}\left[V^2\right] = \int_{-\infty}^{\infty} v^2 \cdot f_V(v) dv$$

  V and $f_V(v)$ are respectively the random voltage and its probability density function.

- In particular, if we apply the transformation $g(\cdot)$ to the random variable X, defined as

$$g(X) = X^n, \;\; n = 0, 1, 2, \ldots$$

Expected Value of a Function of Random Variables

TECHNISCHE UNIVERSITÄT DARMSTADT

- The expectation of $g(X)$ is known as the nth order moment of X,

$$\mathrm{E}\left[X^n\right] = \int_{-\infty}^{\infty} x^n f_X(x) dx$$

  and denoted by μ_n.
- It is also of importance to use central moments of X around the mean.

$$\mu_n' = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])^n\right], \quad n = 0, 1, 2 \ldots$$

$$\int_{-\infty}^{\infty} (x - \mathrm{E}\left[X\right])^n \cdot f_X(x) dx = \int (x - \mu)^n f_X(x) dx$$

- Clearly the first order central moment is zero.

The Variance

TECHNISCHE UNIVERSITÄT DARMSTADT

- The variance of a random variable X is by definition

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 \cdot f_X(x)dx$$

- The positive constant σ, is called the standard deviation of X.
- From the definition, it follows that σ^2 is the mean of the random variable $(X - \mathrm{E}[X])^2$.

The Variance

TECHNISCHE UNIVERSITÄT DARMSTADT

- The variance can also be expressed as

$$
\begin{aligned}
\sigma^2 &= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] = \mathrm{E}\left[X^2 - 2X\cdot(\mathrm{E}[X]) + (\mathrm{E}[X])^2\right] \\
&= \mathrm{E}\left[X^2\right] - 2\mathrm{E}[X\mathrm{E}[X]] + (\mathrm{E}[X])^2 \\
&= \mathrm{E}\left[X^2\right] - 2(\mathrm{E}[X])^2 + (\mathrm{E}[X])^2 \\
&= \mathrm{E}\left[X^2\right] - (\mathrm{E}[X])^2
\end{aligned}
$$

- Example: Exponential distribution

$$f_X(x) = \lambda \exp\{-\lambda x\}u(x),\ \lambda > 0$$

$$\mathrm{E}[X] = \frac{1}{\lambda}$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

Transformation of a Random Variable

TECHNISCHE UNIVERSITÄT DARMSTADT

- In practice, one may wish to transform one random variable X into a new random variable Y

$$Y = g(X),$$

where the probability density function $f_X(x)$ or distribution function of X is known.

- The problem is to find the probability density function of $f_Y(y)$ or distribution function $F_Y(y)$ of Y.
- Let us assume that $g(\cdot)$ is continuous and differentiable at all values of x for which $f_X(x) \neq 0$.

Transformation of a Random Variable

TECHNISCHE UNIVERSITÄT DARMSTADT

- Furthermore, we assume that $g(\cdot)$ is monotone for which the inverse $g^{-1}(\cdot)$ exists. Then:

$$
\begin{aligned}
F_Y(y) &= \mathrm{P}\left[\{Y \leq y\}\right] = \mathrm{P}\left[\{g(X) \leq y\}\right] = \mathrm{P}\left[\{X \leq g^{-1}(y)\}\right] \\
&= \int_{-\infty}^{g^{-1}(y)} f_X(x)dx
\end{aligned}
$$

holds. The density function of $Y = g(X)$ is obtained via differentiation which leads to

$$
f_Y(y) = f_X\left(g^{-1}(y)\right) \cdot \left| \frac{dg^{-1}(y)}{dy} \right|
$$

Transformation of a Random Variable

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example:
  Let
  $$Y = aX + b$$
  To find the probability density function of Y knowing $F_X(x)$ or $f_X(x)$, we calculate
  $$\begin{aligned} F_Y(y) &= \mathrm{P}\left[\{Y \leq y\}\right] = \mathrm{P}\left[\{aX + b \leq y\}\right] \\ &= \mathrm{P}\left[\left\{X \leq \frac{y-b}{a}\right\}\right], \; a > 0 \\ &= F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

Transformation of a Random Variable

And

$$F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right), \; a < 0$$

By differentiating $F_Y(y)$, we obtain the probability density function

$$f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{y-b}{a}\right)$$

Covariance

TECHNISCHE UNIVERSITÄT DARMSTADT

- If the statistical properties of the two random variables X_1 and X_2 are described by their joint probability density function $f_{X_1X_2}(x_1, x_2)$, then we have

$$\mathrm{E}\left[g(X_1, X_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2) \cdot f_{X_1X_2}(x_1, x_2) dx_1 dx_2$$

- For $g(X_1, X_2) = X_1^n X_2^k$

$$\mathrm{E}\left[X_1^n X_2^k\right] = \mu_{n,k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^n x_2^k f_{X_1X_2}(x_1, x_2) dx_1\, dx_2$$

For $n, k = 0, 1, 2, \ldots$ are called the joint moments of X_1 and X_2. Clearly $\mu_{0k} = \mathrm{E}\left[X_2^k\right]$, while $\mu_{n0} = \mathrm{E}\left[X_1^n\right]$.

Covariance

TECHNISCHE UNIVERSITÄT DARMSTADT

- The second order moment $\mathrm{E}[X_1X_2]$ of X_1 and X_2 we denote by $r_{X_1X_2}$,

$$r_{X_1X_2} = \mathrm{E}[X_1X_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2f_{X_1X_2}(x_1,x_2)dx_1dx_2$$

- If $r_{X_1X_2} = \mathrm{E}[X_1X_2] = \mathrm{E}[X_1]\cdot\mathrm{E}[X_2]$, then X_1 and X_2 are known to be uncorrelated. If $r_{X_1X_2} = 0$, then X_1 and X_2 are orthogonal.
- The moments μ'_{nk}, $k = 0,1,2,\ldots$ are known to be joint central moments and defined by (for $n,k = 0,1,2,\ldots$)

$$\begin{aligned}\mu'_{nk} &= \mathrm{E}\left[(X_1-\mathrm{E}[X_1])^n(X_2-\mathrm{E}[X_2])^k\right]\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1-\mathrm{E}[X_1])^n(x_2-\mathrm{E}[X_2])^kf_{X_1X_2}(x_1,x_2)dx_1dx_2\end{aligned}$$

Covariance

- The second order central moments

$$\begin{aligned} \mu'_{20} &= \sigma^2_{X_1} \\ \mu'_{02} &= \sigma^2_{X_2} \end{aligned}$$

  are the variances of X_1 and X_2, respectively.
- The joint moment μ'_{11} is called covariance of X_1 and X_2 and denoted by $c_{X_1 X_2}$

$$\begin{aligned} c_{X_1 X_2} &= \mathrm{E}\left[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])\right] \\ &= \int_{-\infty}^{\infty} (x_1 - \mathrm{E}[X_1])(x_2 - \mathrm{E}[X_2]) \cdot f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 \end{aligned}$$

- Clearly

$$c_{X_1 X_2} = r_{X_1 X_2} - \mathrm{E}[X_1]\,\mathrm{E}[X_2]\,.$$

Covariance

TECHNISCHE UNIVERSITÄT DARMSTADT

- If two random variables are independent, i.e. $f_{X_1X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$, then $c_{X_1X_2} = 0$. The converse however is not true (only for the Gaussian case).
- Example:
  U is uniformly distributed on $[-\pi, \pi)$

$$X_1 = \cos U, \; X_2 = \sin U$$

  X_1 and X_2 are not independent because $X_1^2 + X_2^2 = 1$

$$\begin{aligned} c_{X_1X_2} &= \mathrm{E}\left[(X_1 - \mathrm{E}[X_1])(X_2 - \mathrm{E}[X_2])\right] \\ &= \mathrm{E}[X_1X_2] = \mathrm{E}[\cos U \sin U] \\ &= \frac{1}{2}\mathrm{E}[\sin 2U] \\ &= 0 \end{aligned}$$

  Thus X_1 and X_2 are uncorrelated but not independent.

Random Processes

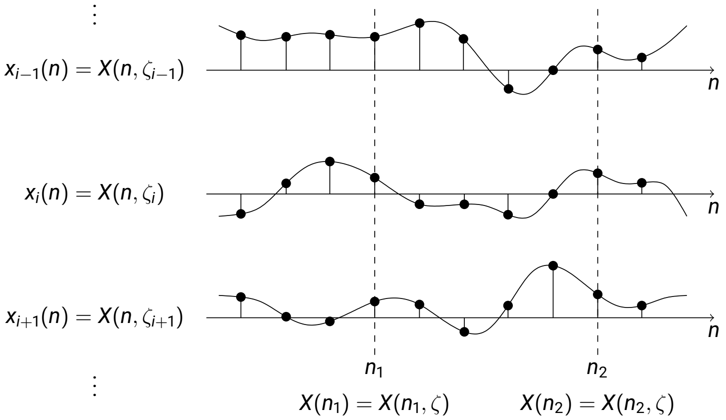
TECHNISCHE UNIVERSITÄT DARMSTADT

- A real-valued random process is an indexed set of real-valued functions of time that has certain statistical properties.
- We characterise a real-valued random process by a set of real-valued functions and associated probability description.
- In general, $x_i(n) = X(n, \zeta_i)$, $n \in \mathbb{Z}$, $\zeta_i \in \mathcal{S}$, denotes the waveform that is obtained when the event ζ_i of the process occurs.
- $x_i(n)$ is called a sample function of the process.
- The set of all possible sample functions $\{X(n, \zeta_i)\}$ is called the ensemble and defines the random process $X(n)$ that describes a noise source.

Random Processes

- A random process can be seen as a function of two variables, time n and elementary event ζ.
- When n is fixed, $X(n, \zeta)$ is simply a random variable. When ζ is fixed, $X(n, \zeta) = x(n)$ is a function of time known as "sample function" or realisation.
- If $n = n_1$ then $X(n_1, \zeta) = X(n_1) = X_1$ is a random variable.

Random Processes



Voltage waveforms emitted from a noise source

Random Processes

TECHNISCHE UNIVERSITÄT DARMSTADT

- To describe the random processes, we define the joint cumulative distribution function (cdf)

$$\begin{aligned} F_X\left(x_1, \ldots, x_N; n_1, \ldots, n_N\right) &= F_{X(n_1),\ldots,X(n_N)}\left(x_1, \ldots, x_N\right) \\ &= \mathrm{P}(\{X(n_1) \leq x_1, \ldots, X(n_N) \leq x_N\}). \end{aligned}$$

- We assume that the following derivative of the cdf exists, and we define the joint probability density function (pdf) as

$$f_X\left(x_1, \ldots, x_N; n_1, \ldots, n_N\right) = \frac{\partial^N F_X\left(x_1, \ldots, x_N; n_1, \ldots, n_N\right)}{\partial x_1 \ldots \partial x_N}.$$

Random Processes

TECHNISCHE UNIVERSITÄT DARMSTADT

- The defined joint cdf or pdf describe a random process completely.
- Practical application is only feasible in special cases.
- Common assumption: identical distribution $f_X(x; n)$ for all time indexes of interest $n_1, \dots, n_N$
- Describe random processes with their moments.

Random Processes

- The mean of process $X(n)$ is given by

$$\mu(n) = \mathrm{E}\left[X(n)\right] = \int_{-\infty}^{\infty} x \cdot f_X(x;n)dx$$

- The variance is given by

$$\sigma^2(n) = \mathrm{E}[X(n)^2] - \mathrm{E}[X(n)]^2 = \int_{-\infty}^{\infty} x^2 \cdot f_X(x;n)dx - \mu(n)^2.$$

- Because in general, $f_X(x;n)$ depends on time index n, the mean and variance of random process $X(n)$ will also depend on time index n.

Second-Order Moment Function

- The second-order moment function (SOMF) of a real-valued stationary process $X(n)$ is defined as:

$$\begin{aligned} r_{XX}(n_1, n_2) &= \mathrm{E}[X(n_1)X(n_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_{X_1X_2}(x_1, x_2; n_1, n_2)\, dx_1\, dx_2. \end{aligned}$$

- The central SOMF is also called the covariance function is defined by:

$$c_{XX}(n_1, n_2) = \mathrm{E}\left[(X(n_1) - E[X(n_1)])(X(n_2) - E[X(n_2)])\right]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - \mu(n_1))(x_2 - \mu(n_2)) f_{X_1X_2}(x_1, x_2; n_1, n_2)\, dx_1\, dx_2.$$

Stationarity

- A random process $X(n)$, $n \in \mathbb{Z}$ is said to be stationary to the order N if, for any n_1, n_2, $\ldots$ n_N, and $x_1, x_2, \ldots, x_N$, and any n_0:

$$
\begin{aligned}
&F_X\left(x_1, \ldots, x_N; n_1, \ldots, n_N\right) \\
&\quad = \mathrm{P}\left[\{X(n_1) \leq x_1, X(n_2) \leq x_2, \ldots, X(n_N) \leq x_N\}\right] \\
&\quad = \mathrm{P}\left[\{X(n_1 + n_0) \leq x_1, X_2(n_2 + n_0) \leq x_2, \ldots, \ X(n_N + n_0) \leq x_N\}\right] \\
&\quad = F_X\left(x_1, \ldots, x_N; n_1 + n_0, \ldots, n_N + n_0\right)
\end{aligned}
$$

Stationarity

- Equivalently, in terms of the probability density function:

$$\begin{aligned} & f_X\left(x_1, x_2, \ldots, x_N; n_1, \ldots, n_N\right) \\ & \qquad = f_X\left(x_1, x_2, \ldots, x_N; n_1 + n_0, \ldots, n_N + n_0\right) \end{aligned}$$

- The process is said to be strictly stationary if it is stationary to the infinite order.
- If a process is stationary, it can be translated in time without changing its statistical description.

Wide-Sense Stationarity

TECHNISCHE UNIVERSITÄT DARMSTADT

- A random process is said to be wide-sense stationary if:
  1. $\mathrm{E}\left[X(n)\right]$ is a constant.
  2. $r_{XX}(n_1, n_2) = r_{XX}(\kappa)$, where $\kappa = |n_2 - n_1|$
- A process that is stationary to order 2 or greater is certainly wide-sense stationary. However, a finite order stationary process is not necessarily strictly stationary.
- If a process is wide-sense stationary, the SOMF is only a function of the time difference $\kappa = |n_2 - n_1|$, and

$$r_{XX}(\kappa) = \mathrm{E}\left[X(n+\kappa) \cdot X(n)\right]$$

Ergodicity

TECHNISCHE UNIVERSITÄT DARMSTADT

- A random process is said to be ergodic if all time averages of any sample function are equal to the corresponding ensemble averages (expectation).

Example: Determine the mean $\mu(n)$ of a process $X(n)$

- Observe a large number of samples $X(n, \zeta_i)$, $i = 1, \ldots, L$
- Use the ensemble average as the estimate for $\mu(n) = \mathrm{E}[X(n)]$, i.e.

$$\hat{\mu}(n) = \frac{1}{L} \sum_{i=1}^{L} X(n, \zeta_i)$$

Ergodicity

TECHNISCHE UNIVERSITÄT DARMSTADT

- Suppose only a single sample $x_i(n) = X(n, \zeta_i)$ of $X(n)$ for $0 \leq n \leq N-1$ is available.
- Can one use the time average as the estimate for $\mu(n)$?.

$$\overline{X(n)} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} X(n, \zeta_i)$$

- Impossible if $\mu(n)$ depends on n
- If $\mu(n) = \mu$ is a constant, then under the general conditions, $\overline{X(n)}$ approaches μ.

Second-Order Moment Function

TECHNISCHE UNIVERSITÄT DARMSTADT

Properties of the SOMF of a Stationary Real-valued Process

1. $r_{XX}(0) = \mathrm{E}\left[X(n)^2\right] = \sigma_X^2 + \mu_X^2$ is the second order moment
2. $r_{XX}(\kappa) = r_{XX}(-\kappa)$
3. $r_{XX}(0) \geq |r_{XX}(\kappa)|$, $|\kappa| > 0$

- Proofs for 1 and 2 follow from the definition. The third property holds because

$$\mathrm{E}\left[(X(n+\kappa) \pm X(n))^2\right] \geq 0$$

$$\mathrm{E}\left[X(n+\kappa)^2\right] \pm 2\mathrm{E}\left[X(n+\kappa)\cdot X(n)\right] + \mathrm{E}\left[X(n)^2\right] \geq 0$$

$$r_{XX}(0) \pm 2\, r_{XX}(\kappa) + r_{XX}(0) \geq 0.$$

Cross-SOMF

- The "cross-SOMF" for two-valued real processes $X(n)$ and $Y(n)$ is:

$$r_{XY}(n_1, n_2) = \mathrm{E}\left[X(n_1) \cdot Y(n_2)\right]$$

- If $X(n)$ and $Y(n)$ are jointly stationary, the cross-SOMF becomes:

$$r_{XY}(n_1, n_2) = r_{XY}(n_1 - n_2) = r_{XY}(\kappa)$$

where $\kappa = n_1 - n_2$.

Properties of the Cross-SOMF

TECHNISCHE UNIVERSITÄT DARMSTADT

Properties of the Cross-SOMF of a Jointly Stationary Real-valued Process

1. $r_{XY}(-\kappa) = r_{YX}(\kappa)$
2. $|r_{XY}(\kappa)| \leq \sqrt{r_{XX}(0) \cdot r_{YY}(0)}$.

- Property 2 holds because

$$\mathrm{E}\left[(X(n+\kappa) - aY(n))^2\right] \geq 0 \quad \forall a \in \mathbb{R}$$

$$\mathrm{E}\left[X(n+\kappa)^2\right] + a^2\mathrm{E}\left[Y(n)^2\right] - 2a\mathrm{E}\left[X(n+\kappa)Y(n)\right] \geq 0 \quad \forall a \in \mathbb{R}$$

$$r_{XX}(0) + a^2 r_{YY}(0) - 2a r_{XY}(\kappa) \geq 0$$

Properties of the Cross-SOMF

- The previous quadratic form is valid if the roots of a are not real (except as a double real root). Therefore

$$r_{XY}(\kappa)^2 - r_{XX}(0)r_{YY}(0) \leq 0$$

$$r_{XY}(\kappa)^2 \leq r_{XX}(0)r_{YY}(0) \Rightarrow |r_{XY}(\kappa)| \leq \sqrt{r_{XX}(0)r_{YY}(0)}$$

- Sometimes

$$\frac{r_{XY}(\kappa)}{\sqrt{r_{XX}(0)r_{YY}(0)}}$$

is denoted by $\rho_{XY}(\kappa)$.

Properties of the Cross-SOMF

3.

$$|r_{XY}(\kappa)| \leq \frac{1}{2}\,[r_{XX}(0) + r_{YY}(0)]$$

Property 2 constitutes a tighter bound than that of property 3, because the geometric mean of two positive numbers cannot exceed the arithmetic mean, that is

$$\sqrt{r_{XX}(0) \cdot r_{YY}(0)} \leq \frac{1}{2}\,[r_{XX}(0) + r_{YY}(0)]$$

Properties of the Cross-SOMF

- Two random processes $X(n)$ and $Y(n)$ are said to be uncorrelated if:

$$r_{XY}(\kappa) = \mathrm{E}\left[X(n+\kappa)\right] \cdot \mathrm{E}\left[Y(n)\right] = \mu_X \cdot \mu_Y \; \forall \kappa \in \mathbb{Z}$$

- Two random processes $X(n)$ and $Y(n)$ are orthogonal if:

$$r_{XY}(\kappa) = 0, \; \forall \kappa \in \mathbb{Z}$$

- If $X(n) \equiv Y(n)$, the cross-SOMF becomes the SOMF.
- If the random processes $X(n)$ and $Y(n)$ are jointly ergodic, the time average may be used to replace the ensemble average.

Properties of the Cross-SOMF

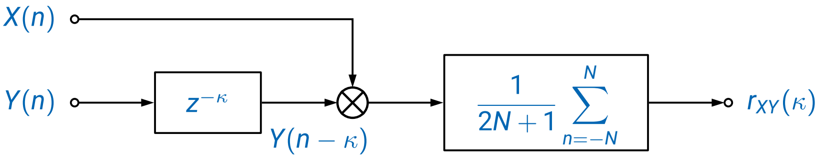
- If $X(n)$ and $Y(n)$ are jointly ergodic

$$
\begin{aligned}
r_{XY}(\kappa) &\triangleq \mathrm{E}\left[X(n+\kappa)\cdot Y(n)\right] = \mathrm{E}\left[X(n)\cdot Y(n-\kappa)\right] \\
&\equiv \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} X(n)\, Y(n-\kappa)
\end{aligned}
$$

- In this case, cross-SOMFs and auto-SOMFs may be measured by using a system composed of a delay element, a multiplier and an accumulator.

Properties of the Cross-SOMF

- In the figure below, the block $z^{-\kappa}$ is a delay line of order κ.



Measurement of SOMF.

The Central SOMF

TECHNISCHE UNIVERSITÄT DARMSTADT

- The central second-order moment function is called the covariance function and is defined by

$$c_{XX}(n+\kappa,n) = \mathrm{E}\left[(X(n+\kappa) - \mathrm{E}\left[X(n+\kappa)\right])(X(n) - \mathrm{E}\left[X(n)\right])\right]$$

- This can be re-written in the form

$$c_{XX}(n+\kappa,n) = r_{XX}(n+\kappa,n) - \mathrm{E}\left[X(n+\kappa)\right]\mathrm{E}\left[X(n)\right]$$

- The covariance function for two processes $X(n)$ and $Y(n)$ is defined by

$$c_{XY}(n+\kappa,n) = \mathrm{E}\left[(X(n+\kappa) - \mathrm{E}\left[X(n+\kappa)\right])(Y(n) - \mathrm{E}\left[Y(n)\right])\right]$$

The Central SOMF

TECHNISCHE UNIVERSITÄT DARMSTADT

- The covariance function may also be expressed by

$$c_{XY}(n+\kappa,n) = r_{XY}(n+\kappa,n) - \mathrm{E}\left[X(n+\kappa)\right]\mathrm{E}\left[Y(n)\right]$$

- For processes that are at least jointly wide sense stationary, we have

$$c_{XX}(\kappa) = r_{XX}(\kappa) - \left(\mathrm{E}\left[X(n)\right]\right)^2$$

and

$$c_{XY}(\kappa) = r_{XY}(\kappa) - \mathrm{E}\left[X(n)\right]\mathrm{E}\left[Y(n)\right]$$

The Central SOMF

- The variance of a random process is obtained from $c_{XX}(\kappa)$ by setting $\kappa = 0$.
- If the process is stationary

$$\begin{aligned}\sigma_X^2 &= \mathrm{E}\left[(X(n) - \mathrm{E}[X(n)])^2\right] \\ &= r_{XX}(0) - (\mathrm{E}[X(n)])^2 \\ &= r_{XX}(0) - \mu_X^2\end{aligned}$$

- For two random processes, if $c_{XY}(n+\kappa, n) = 0$, then they are called uncorrelated, which is equivalent to $r_{XY}(n+\kappa, n) = \mathrm{E}[X(n+\kappa)]\,\mathrm{E}[Y(n)]$.

Complex-valued Random Processes

TECHNISCHE UNIVERSITÄT DARMSTADT

- A complex-valued random process is defined as $Z(n) \triangleq X(n) + j\,Y(n)$ where $X(n)$ and $Y(n)$ are real-valued random processes.
- A complex-valued process is stationary if $X(n)$ and $Y(n)$ are jointly stationary, i.e.

$$
\begin{aligned}
&f_Z\left(x_1, \ldots, x_N, y_1, \ldots, y_N; n_1, \ldots, n_{2N}\right) \\
&\qquad = f_Z\left(x_1, \ldots, x_N, y_1, \ldots, y_N; n_1 + n_0, \ldots, n_{2N} + n_0\right)
\end{aligned}
$$

  for all $x_1, \ldots, x_N$, $y_1, \ldots, y_N$ and all $n_1, \ldots, n_{2N}$, n_0 and N.
- The mean of a complex-valued random process is defined as:

$$
\mathrm{E}\left[Z(n)\right] = \mathrm{E}\left[X(n)\right] + j\mathrm{E}\left[Y(n)\right]
$$

Complex-valued Random Processes

- The SOMF for a complex-valued random process is:

$$r_{ZZ}(n_1, n_2) = \mathrm{E}\left[Z(n_1) \cdot Z(n_2)^*\right]$$

- The complex-valued random process is stationary in the wide sense if $\mathrm{E}\left[Z(n)\right]$ is a complex-valued constant and $r_{ZZ}(n_1, n_2) = r_{ZZ}(\kappa)$ where $\kappa = n_1 - n_2$.
- The SOMF of a wide-sense stationary complex-valued process has the symmetry property:

$$r_{ZZ}(-\kappa) = r_{ZZ}(\kappa)^*$$

as it can be easily seen from the definition.

Complex-valued Random Processes

- The cross-SOMF for two complex-valued random processes $Z_1(n)$ and $Z_2(n)$ is:

$$r_{Z_1Z_2}(n_1, n_2) = \mathrm{E}\left[Z_1(n_1) \cdot Z_2(n_2)^*\right]$$

- If the complex-valued random processes are jointly wide-sense stationary, the cross-SOMF becomes:

$$r_{Z_1Z_2}(n_1, n_2) = r_{Z_1Z_2}(\kappa) \text{ where } \kappa = n_1 - n_2$$

- The covariance function of a complex-valued random process is defined as

$$c_{ZZ}(n+\kappa, n) = \mathrm{E}\left[\left(Z(n+\kappa) - \mathrm{E}\left[Z(n+\kappa)\right]\right)\left(Z(n) - \mathrm{E}\left[Z(n)\right]\right)^*\right]$$

and is a function of κ only if $Z(n)$ is wide-sense stationary.

Complex-valued Random Processes

TECHNISCHE UNIVERSITÄT DARMSTADT

- The cross-covariance function of $Z_1(n)$ and $Z_2(n)$ is given as:

$$c_{Z_1Z_2}(n+\kappa, n) = \mathrm{E}\left[\left(Z_1(n+\kappa) - \mathrm{E}\left[Z_1(n+\kappa)\right]\right)\left(Z_2(n) - \mathrm{E}\left[Z_2(n)\right]\right)^*\right]$$

- $Z_1(n)$ and $Z_2(n)$ are called uncorrelated if $c_{Z_1Z_2}(n+\kappa, n) = 0$.
- They are orthogonal if $r_{Z_1Z_2}(n+\kappa, n) = 0$.

Power Spectral Density and Spectrum

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Find a frequency domain representation for random processes
- The discrete-time Fourier transform (DTFT) of a random process does not exist because it is an ensemble of deterministic processes
- A frequency domain representation is given by the Power Spectral Density (PSD) or the spectrum
- The spectrum describes the distribution of power over frequencies
- It is a fundamental tool to analyse, e.g. random phenomena whose behavior is best described in the frequency domain

The Wiener-Khintchine Theorem

TECHNISCHE UNIVERSITÄT DARMSTADT

- Wiener-Khintchine theorem:
  When $X(n)$ is a wide-sense stationary process, the PSD can be obtained from the Fourier Transform of the SOMF:

$$S_{XX}(e^{j\omega}) = \mathcal{F}\{r_{XX}(\kappa)\} = \sum_{\kappa=-\infty}^{\infty} r_{XX}(\kappa)\, e^{-j\omega\kappa}$$

and conversely,

$$r_{XX}(\kappa) = \mathcal{F}^{-1}\left\{S_{XX}(e^{j\omega})\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(e^{j\omega})\, e^{j\omega\kappa}\, d\omega$$

The Wiener-Khintchine Theorem

- We now define the spectrum of a stationary random process as the Fourier Transform of the covariance function

$$C_{XX}(e^{j\omega}) = \mathcal{F}\left\{c_{XX}(\kappa)\right\} = \sum_{\kappa=-\infty}^{\infty} c_{XX}(\kappa)\, e^{-j\omega\kappa}$$

and conversely,

$$c_{XX}(\kappa) = \mathcal{F}^{-1}\left\{C_{XX}(e^{j\omega})\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{XX}(e^{j\omega})\, e^{j\omega\kappa}\, d\omega$$

PSD - Examples

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example: Sinusoid with random phase

$$X(n) = A\sin(\omega_0 n + \Phi)$$

where $A,\ \omega_0$ are constants and Φ is uniformly distributed on $[0, 2\pi)$.

- The mean of $X(n)$ is given by

$$\mathrm{E}\left[X(n)\right] = \int_0^{2\pi} \frac{1}{2\pi} A\sin(\omega_0 n + \varphi)\, d\varphi = 0$$

- The SOMF of $X(n)$ is

$$r_{XX}(\kappa) = \mathrm{E}\left[X(n+\kappa)X(n)\right] = \frac{A^2}{2}\, \cos(\omega_0\kappa)$$

PSD - Examples

TECHNISCHE UNIVERSITÄT DARMSTADT

- Taking the Fourier transform of $r_{XX}(\kappa)$ we obtain the PSD

$$S_{XX}(e^{j\omega}) = \mathcal{F}\{r_{XX}(\kappa)\} = \frac{A^2}{2}\pi\left\{\eta(\omega - \omega_0) + \eta(\omega + \omega_0)\right\}$$

where

$$\eta(\omega) = \sum_{k=-\infty}^{\infty} \delta(\omega + k2\pi)$$

PSD - Examples

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example: Multiple frequency components:

$$Z(n) = \sum_{i=1}^{P} A_i \exp\{j\omega_i n\}.$$

- The complex random variables $A_i \in \mathbb{C}$ are uncorrelated with zero mean and variance σ_i^2, $i = 1, \ldots, P$:

$$\begin{aligned} r_{ZZ}(\kappa) &= \sum_{i=1}^{P} \sigma_i^2 \exp\{j\omega_i \kappa\} \\ S_{ZZ}(e^{j\omega}) &= \sum_{i=1}^{P} \sigma_i^2 \eta(\omega - \omega_i) \end{aligned}$$

PSD - Examples

- Example: Multiple frequency components:

$$X(n) = \sum_{i=1}^{P} \left[A_i \cos(\omega_i n) + B_i \sin(\omega_i n) \right].$$

The real-valued r.v.s A_i and B_i, $i = 1, \ldots, P$ are pairwise uncorrelated with zero mean and common variance σ_i^2.

$$\begin{aligned} r_{XX}(\kappa) &= \sum_{i=1}^{P} \sigma_i^2 \cos(\omega_i \kappa) \\ S_{XX}(e^{j\omega}) &= \pi \sum_{i=1}^{P} \sigma_i^2 \left[\eta(\omega - \omega_i) + \eta(\omega + \omega_i) \right] \end{aligned}$$

Properties of the PSD

- The PSD has as number of important properties:
  1. The PSD is always real-valued, even if $X(n)$ is complex-valued:

$$
\begin{aligned}
S_{XX}(e^{j\omega})^* &= \sum_{\kappa=-\infty}^{\infty} r_{XX}(\kappa)^* e^{+j\omega\kappa} = \sum_{\kappa=-\infty}^{\infty} r_{XX}(-\kappa) e^{+j\omega\kappa} \\
&= \sum_{\kappa'=-\infty}^{\infty} r_{XX}(\kappa') e^{-j\omega\kappa'} = S_{XX}(e^{j\omega})
\end{aligned}
$$

  2. $S_{XX}(e^{j\omega}) \geq 0$, even if $X(n)$ is complex-valued.
  3. When $X(n)$ is real-valued, $S_{XX}(e^{-j\omega}) = S_{XX}(e^{j\omega})$.
  4. $\int_{-\pi}^{\pi} S_{XX}(e^{j\omega}) \frac{d\omega}{2\pi} = P_{XX}$ is the total normalised average power.

White Noise Process

- The PSD of a white noise process, $X(n)$, is constant over all frequencies:

$$S_{XX}(e^{j\omega}) = \sigma_X^2$$

where σ_X is a positive constant.

- Taking the inverse Fourier transform of the PSD, we obtain the SOMF:

$$r_{XX}(\kappa) = \sigma_X^2\,\delta(\kappa)$$

where $\delta(\kappa) = \begin{cases} 1, & \kappa = 0 \\ 0, & \kappa \neq 0 \end{cases}$ is Kronecker's delta function.

Cross-Power Spectral Density

- For two jointly stationary processes $X(n)$ and $Y(n)$ the cross-power spectral density is given by

$$S_{XY}(e^{j\omega}) = \mathcal{F}\{r_{XY}(\kappa)\} = \sum_{\kappa=-\infty}^{\infty} r_{XY}(\kappa) e^{-j\omega\kappa}$$

  and is in general complex-valued.
- Equivalently, the cross-spectrum of two jointly stationary processes $X(n)$ and $Y(n)$ is given by:

$$C_{XY}(e^{j\omega}) = \mathcal{F}\{c_{XY}(\kappa)\} = \sum_{\kappa=-\infty}^{\infty} c_{XY}(\kappa) e^{-j\omega\kappa}$$

Cross PSD - Properties

TECHNISCHE UNIVERSITÄT DARMSTADT

- The cross-PSD has a number of important properties:

 1. $S_{XY}(e^{j\omega}) = S_{YX}(e^{j\omega})^*$. In addition, if $X(n)$ and $Y(n)$ are real-valued, then $S_{XY}(e^{j\omega}) = S_{YX}(e^{-j\omega})$. This is from the definition of the Fourier Transform.
 2. $\mathrm{Re}\{S_{XY}(e^{j\omega})\}$ and $\mathrm{Re}\{S_{YX}(e^{j\omega})\}$ are even functions of ω if $X(n)$ and $Y(n)$ are real-valued. This is from the definition of the Fourier Transform.
 3. $\mathrm{Im}\{S_{XY}(e^{j\omega})\}$ and $\mathrm{Im}\{S_{YX}(e^{j\omega})\}$ are odd functions of ω if $X(n)$ and $Y(n)$ are real-valued. This is from the definition of the Fourier Transform.
 4. $S_{XY}(e^{j\omega}) = 0$ and $S_{YX}(e^{j\omega}) = 0$ if $X(n)$ and $Y(n)$ are orthogonal.

Linear filtering

- Consider $X(n)$ and $Y(n)$ are stationary, $h(n)$ is the unit sample response of a linear time-invariant (LTI) system, and the filter is stable $(\sum |h(n)| < \infty)$:

$$X(n) \longrightarrow \boxed{h(n)} \longrightarrow Y(n)$$

- For such a system

$$Y(n) = \sum_k h(k)X(n-k) = \sum_k h(n-k)X(k)$$

exists with probability one, i.e. $\sum_{k=-N}^{M} h(k)X(n-k)$ converges with probability one to $Y(n)$ for $N, M \to \infty$.

Linear filtering

TECHNISCHE UNIVERSITÄT DARMSTADT

- If $X(n)$ is stationary and $\mathrm{E}\left[|X(n)|\right] < \infty$, then $Y(n)$ is stationary.
- If $X(n)$ is white noise, i.e. $\mathrm{E}\left[X(n)\right] = 0$, $\mathrm{E}\left[X(n)X(k)\right] = \sigma_X^2\delta(n-k)$, with σ_X^2 the power of the noise and $\delta(n)$ the Kronecker delta function

$$\delta(n) = \begin{cases} 1, & n = 0 \\ 0, & \text{else} \end{cases}$$

- $Y(n)$ is called a linear process.

Linear filtering

- Assume now that the filter is not necessarily stable but $\int |H(e^{j\omega})|^2 \, d\omega < \infty$
- If for $X(n)$, $\sum |c_{XX}(n)| < \infty$, then

$$Y(n) = \sum_k h(k)X(n-k) = \sum_k h(n-k)X(k)$$

  exists in the mean square sense (MSS), i.e, $\sum_{k=-N}^{M} h(k)X(n-k)$ converges in the MSS to $Y(n)$ for $N, M \to \infty$.
- $Y(n)$ is stationary in the wide sense with

$$\mu_Y \triangleq \mathrm{E}\left[Y(n)\right] = \sum_k h(k)\mathrm{E}\left[X(n-k)\right] = \mu_X H(e^{j0})$$

Linear filtering

- Example: Hilbert Transforms

$$h(n) = \begin{cases} 0, & n = 0 \\ (1 - \cos(\pi n))/(\pi n), & \text{else} \end{cases} \quad \Leftrightarrow \quad H(e^{j\omega}) = \begin{cases} -j, & 0 < \omega < \pi \\ 0, & \omega = 0, -\pi \\ j, & -\pi < \omega < 0 \end{cases}$$

- The filter is not stable, but $\int |H(e^{j\omega})|^2 \, d\omega < \infty$.
- $Y(n)$ is stationary in the wide sense.
- For example, if $X(n) = a\cos(\omega_o n + \phi)$, then $Y(n) = a\sin(\omega_o n + \phi)$.

Output covariance function and spectrum

TECHNISCHE UNIVERSITÄT DARMSTADT

- We first define $\tilde{X}(n) \triangleq X(n) - \mu_X$ and

$$\begin{aligned}
\tilde{Y}(n) \triangleq Y(n) - \mu_Y &= \sum_k h(k)X(n-k) - \sum_k h(k)\mathrm{E}\left[X(n-k)\right] \\
&= \sum_k h(k)\left(X(n-k) - \mu_X\right) \\
&= \sum_k h(k)\tilde{X}(n-k)
\end{aligned}$$

- In general, the system unit sample response and the input process may be complex-valued.

Output covariance function and spectrum

- The covariance function of the output is determined according to

$$
\begin{aligned}
c_{YY}(\kappa) &= \mathrm{E}\left[\tilde{Y}(n+\kappa)\tilde{Y}(n)^*\right] \\
&= \mathrm{E}\left[\left(\sum_k h(k)\tilde{X}(n+\kappa-k)\right)\left(\sum_l h(l)\tilde{X}(n-l)\right)^*\right] \\
&= \sum_k \sum_l h(k)h(l)^*\mathrm{E}\left[\tilde{X}(n+\kappa-k)\tilde{X}(n-l)^*\right] \\
&= \sum_k \sum_l h(k)h(l)^* c_{XX}(\kappa-k+l)
\end{aligned}
$$

Output covariance function and spectrum

- Taking the Fourier transform of both sides, leads to

$$C_{YY}(e^{j\omega}) = |H(e^{j\omega})|^2 C_{XX}(e^{j\omega})$$

- Applying the inverse Fourier transform formula, yields

$$c_{YY}(\kappa) = \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) e^{j\omega\kappa} \frac{d\omega}{2\pi}$$

  - The integral exists because $C_{XX}(e^{j\omega})$ is bounded and

$$\int |H(e^{j\omega})|^2 \, d\omega < \infty$$

.

Output covariance function and spectrum

- Example: FIR filter
  Let $X(n)$ be white noise with power σ^2 and

$$Y(n) = X(n) - X(n-1).$$

- From the above, one can see that

$$h(n) = \begin{cases} 1, & n = 0 \\ -1, & n = 1 \\ 0, & \text{else} \end{cases}$$

Output covariance function and spectrum

- Taking the Fourier transform leads to

$$H(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h(n)e^{-j\omega n} = 1 - e^{-j\omega}$$

- The spectrum of the output process is therefore

$$\begin{aligned} C_{YY}(e^{j\omega}) &= |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) \\ &= |1 - e^{-j\omega}|^2 \sigma^2 \\ &= |2je^{-j\omega/2} \sin(\omega/2)|^2 \sigma^2 \\ &= 4 \sin(\omega/2)^2 \sigma^2 \end{aligned}$$

Cross-covariance function and cross-spectrum

TECHNISCHE UNIVERSITÄT DARMSTADT

- The cross-covariance function of the output signal is determined according to

$$
\begin{aligned}
c_{YX}(\kappa) &= \mathrm{E}\left[\tilde{Y}(n+\kappa)\tilde{X}(n)^*\right] \\
&= \mathrm{E}\left[\left(\sum_k h(k)\tilde{X}(n+\kappa-k)\right)\tilde{X}(n)^*\right] \\
&= \sum_k h(k)\mathrm{E}\left[\tilde{X}(n+\kappa-k)\tilde{X}(n)^*\right] \\
&= \sum_k h(k)c_{XX}(\kappa-k)
\end{aligned}
$$

Cross-covariance function and cross-spectrum

- Taking the Fourier transform of both sides leads to

$$C_{YX}(e^{j\omega}) = H(e^{j\omega})C_{XX}(e^{j\omega})$$

- Applying the inverse Fourier transform formula yields

$$c_{YX}(\kappa) = \int_{-\pi}^{\pi} H(e^{j\omega})C_{XX}(e^{j\omega})\mathrm{e}^{j\omega\kappa}\,\frac{d\omega}{2\pi}.$$

Cross-covariance function and cross-spectrum

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example: Consider the following system where $X(n)$ is a stationary, zero-mean white noise process with variance σ_X^2.

$$X(n) \longrightarrow \boxed{h(n)} \longrightarrow Y(n)$$

Represenation of a discrete-time system

- The unit sample response of the system is given by

$$h(n) = \left[\frac{1}{3}\left(\frac{1}{4}\right)^n + \frac{2}{3}\left(-\frac{1}{2}\right)^n\right] u(n)$$

where $u(n)$ is the unit step function $u(n) = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$

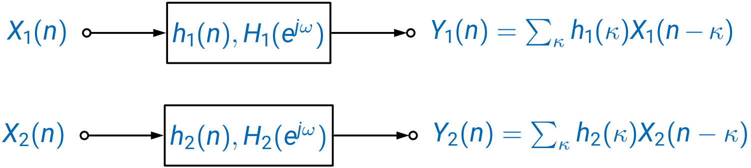
Cross-covariance function and cross-spectrum

- We wish to determine $C_{YY}(e^{j\omega})$ and $C_{YX}(e^{j\omega})$:

$$
\begin{aligned}
H(e^{j\omega}) &= \frac{1/3}{1-\frac{1}{4}\mathrm{e}^{-j\omega}} + \frac{2/3}{1+\frac{1}{2}\mathrm{e}^{-j\omega}} = \frac{1}{1+\frac{1}{4}\mathrm{e}^{-j\omega}-\frac{1}{8}\mathrm{e}^{-j2\omega}} \\
C_{YY}(e^{j\omega}) &= |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) \\
&= \frac{\sigma_X^2}{\left|1+\frac{1}{4}\mathrm{e}^{-j\omega}-\frac{1}{8}\mathrm{e}^{-j2\omega}\right|^2} \\
C_{YX}(e^{j\omega}) &= H(e^{j\omega}) C_{XX}(e^{j\omega}) \\
&= \frac{\sigma_X^2}{1+\frac{1}{4}\mathrm{e}^{-j\omega}-\frac{1}{8}\mathrm{e}^{-j2\omega}}
\end{aligned}
$$

Interaction of Two Linear Systems

- The previous result may be generalised to obtain the cross-covariance or cross-spectrum for two linear systems.
- Consider the two linear systems shown below:



Interaction of Two Linear Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

- If $X_1(n)$ and $X_2(n)$ are WSS and the systems are LTI, the output cross-covariance function is:

$$c_{Y_1Y_2}(\kappa) = \sum_k \sum_l h_1(k)h_2(l)^* c_{X_1X_2}(\kappa - k + l)$$

- It can be seen from the above equation that

$$c_{Y_1Y_2}(\kappa) = h_1(\kappa) * h_2(-\kappa)^* * c_{X_1X_2}(\kappa)$$

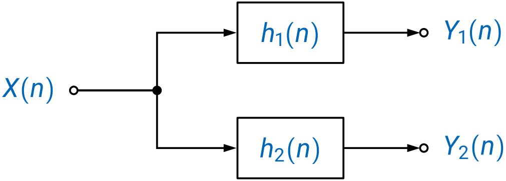
where $*$ denotes convolution.

- Taking the Fourier transform leads to

$$C_{Y_1Y_2}(e^{j\omega}) = H_1(e^{j\omega})H_2(e^{j\omega})^* C_{X_1X_2}(e^{j\omega})$$

Interaction of Two Linear Systems

- Example: Consider the following system



- $X(n)$ is stationary, zero-mean white noise with variance σ_X^2.
- Let $h_1(n)$ be the same system from the previous example and $h_2(n)$ given by

$$h_2(n) = \delta(n) + \frac{1}{2}\delta(n-1) - \frac{1}{2}\delta(n-2)$$

where $\delta(n)$ is Kronecker's delta.

Interaction of Two Linear Systems

TECHNISCHE UNIVERSITÄT DARMSTADT

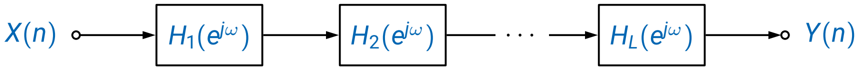
- We wish to find the cross-spectrum $C_{Y_1Y_2}(e^{j\omega})$:

$$
\begin{aligned}
H_2(e^{j\omega}) &= \sum_n h_2(n)\mathrm{e}^{-j\omega n} \\
&= 1 + \frac{1}{2}\mathrm{e}^{-j\omega} - \frac{1}{2}\mathrm{e}^{-j2\omega}
\end{aligned}
$$

$$
\begin{aligned}
C_{Y_1Y_2}(e^{j\omega}) &= H_1(e^{j\omega})H_2(e^{j\omega})^* C_{XX}(e^{j\omega}) \\
&= \left[\frac{1 + \frac{1}{2}e^{j\omega} - \frac{1}{2}\mathrm{e}^{j2\omega}}{1 + \frac{1}{4}\mathrm{e}^{-j\omega} - \frac{1}{8}\mathrm{e}^{-j2\omega}}\right]\sigma_X^2
\end{aligned}
$$

Cascade of linear systems

- We consider a cascade of L linear systems in serial connection:.



- The transfer function $H(e^{j\omega})$ is the product of the L individual system functions

$$H(e^{j\omega}) = H_1(e^{j\omega})H_2(e^{j\omega})\cdots H_L(e^{j\omega}).$$

Cascade of linear systems

- The spectrum of the output:

$$
\begin{aligned}
C_{YY}(e^{j\omega}) &= C_{XX}(e^{j\omega}) \left| \prod_{i=1}^{L} H_i(e^{j\omega}) \right|^2 \\
&= C_{XX}(e^{j\omega}) \prod_{i=1}^{L} \left| H_i(e^{j\omega}) \right|^2
\end{aligned}
$$

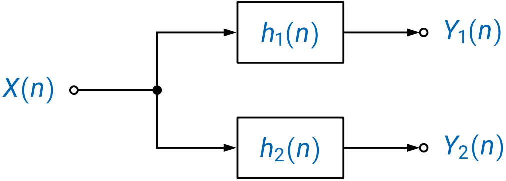
- The cross spectrum of the output with the input:

$$
C_{YX}(e^{j\omega}) = C_{XX}(e^{j\omega}) \prod_{i=1}^{L} H_i(e^{j\omega})
$$

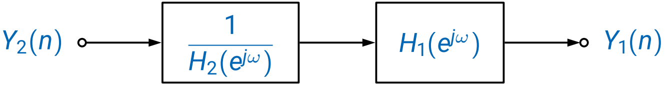
Cascade of linear systems

TECHNISCHE UNIVERSITÄT DARMSTADT

- Example:
  We wish to find the cross spectrum $C_{Y_1Y_2}(e^{j\omega})$ for the system



- We reformulate the problem as a cascade of two systems:



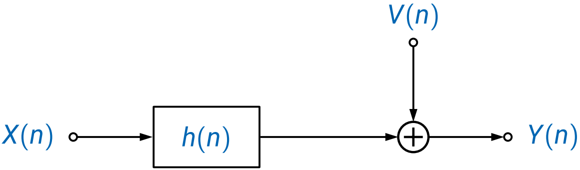
Cascade of linear systems

- Remembering that $C_{Y_2Y_2}(e^{j\omega}) = |H_2(e^{j\omega})|^2 C_{XX}(e^{j\omega})$ we obtain

$$
\begin{aligned}
C_{Y_1Y_2}(e^{j\omega}) &= \frac{H_1(e^{j\omega})}{H_2(e^{j\omega})} C_{Y_2Y_2}(e^{j\omega}) \\
&= \frac{H_1(e^{j\omega})}{H_2(e^{j\omega})} |H_2(e^{j\omega})|^2 C_{XX}(e^{j\omega}) \\
&= H_1(e^{j\omega}) H_2(e^{j\omega})^* C_{XX}(e^{j\omega})
\end{aligned}
$$

Linear filtered process plus noise

- Consider a linear filtered process plus noise:



  - $X(n)$ is stationary with $\mathrm{E}[X(n)] = 0$
  - $V(n)$ stationary with $\mathrm{E}[V(n)] = 0$
  - $c_{VX}(n) = 0$ ($X(n)$ and $V(n)$ uncorrelated)
  - $h(n)$, $X(n)$ and $V(n)$ are real-valued.

Linear filtered process plus noise

- The expected value of the output is derived below.

$$
\begin{aligned}
Y(n) &= \sum_{k=-\infty}^{\infty} h(k)X(n-k) + V(n) \\
\mathrm{E}\left[Y(n)\right] &= \sum_{k=-\infty}^{\infty} h(k)\mathrm{E}\left[X(n-k)\right] + \mathrm{E}\left[V(n)\right] = 0
\end{aligned}
$$

- The covariance function of the output is then ...

Linear filtered process plus noise

TECHNISCHE UNIVERSITÄT DARMSTADT

$$
\begin{aligned}
c_{YY}(\kappa) &= \mathrm{E}\left[Y(n+\kappa)Y(n)\right] \\
&= \mathrm{E}\Bigg[\Big(\sum_k h(k)X(n+\kappa-k) + V(n+\kappa)\Big) \\
&\quad \times \Big(\sum_l h(l)X(n-l) + V(n)\Big)\Bigg] \\
&= \sum_k \sum_l h(k)h(l)\mathrm{E}\left[X(n+\kappa-k)X(n-l)\right] \\
&+ \sum_k h(k)\mathrm{E}\left[X(n+\kappa-k)V(n)\right] \\
&+ \sum_l h(l)\mathrm{E}\left[V(n+\kappa)X(n-l)\right] + \mathrm{E}\left[V(n+\kappa)V(n)\right]
\end{aligned}
$$

Linear filtered process plus noise

TECHNISCHE UNIVERSITÄT DARMSTADT

- Since X(n) and V(n) are uncorrelated, we obtain

$$c_{yy}(\kappa) = \sum_{k} \sum_{l} h(k)h(l)c_{XX}(\kappa - k + l) + c_{VV}(\kappa)$$

- Applying the Fourier transform to the previous expression:

$$\begin{aligned} C_{YY}(e^{j\omega}) &= \mathcal{F}\{c_{YY}(\kappa)\} \\ &= |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega}) \end{aligned}$$

Linear filtered process plus noise

- The cross-covariance function of the output is:

$$
\begin{aligned}
c_{YX}(\kappa) &= \mathrm{E}\left[Y(n+\kappa)X(n)\right] \\
&= \mathrm{E}\left[\left(\sum_k h(k)X(n+\kappa-k)+V(n+\kappa)\right)X(n)\right] \\
&= \sum_k h(k)\mathrm{E}\left[X(n+\kappa-k)X(n)\right] = \sum_k h(k)c_{XX}(\kappa-k)
\end{aligned}
$$

- Applying the Fourier transform to the above expression:

$$
C_{YX}(e^{j\omega}) = H(e^{j\omega})C_{XX}(e^{j\omega})
$$

- When the additive output noise $V(n)$ is uncorrelated with the input process $X(n)$, it has no effect on the cross-spectrum of the input and output.

Linear filtered process plus noise

- Example: Multi-component signal in noise
  The process $X(n)$ is comprised of a sum of K complex-valued signals:

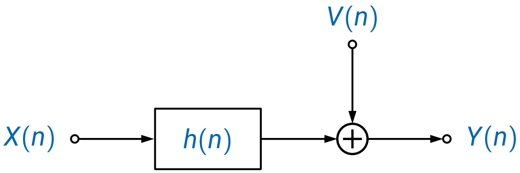
$$X(n) = \sum_{k=1}^{K} A_k e^{j(\omega_0 n + \phi_k)}$$

  - ω_0 is a constant
  - A_k is a zero-mean random amplitude of the kth signal with variance $\sigma_k^2,\ k = 1, 2, \ldots, K$.
  - The phase ϕ_k is uniformly distributed on $[-\pi, \pi)$.
  - The random variables A_k and ϕ_k are statistically pairwise independent for all $k = 1, \ldots, K$.

Linear filtered process plus noise

TECHNISCHE UNIVERSITÄT DARMSTADT

- $X(n)$ is the input to a stable LTI system with known impulse response $h(n)$.



- The output of the system is buried in zero-mean noise $V(n)$ of known covariance $c_{VV}(n)$, independent of $X(n)$.
- We wish to determine the cross-spectrum $C_{YX}(e^{j\omega})$ and the auto-spectrum $C_{YY}(e^{j\omega})$.

Linear filtered process plus noise

TECHNISCHE UNIVERSITÄT DARMSTADT

- We first calculate the expected value the input process $X(n)$

$$\mathrm{E}\left[X(n)\right] = \mathrm{E}\left[\sum_{k=1}^{K} A_k e^{j(\omega_0 n + \phi_k)}\right]$$

Since A_k and ϕ_k are independent for all $k = 1, \ldots, K$

$$\mathrm{E}\left[X(n)\right] = \sum_{k=1}^{K} \mathrm{E}\left[A_k\right] \mathrm{E}\left[e^{j(\omega_0 n + \phi_k)}\right] = 0$$

- The covariance function of the input process in then ...

Linear filtered process plus noise

$$
\begin{aligned}
c_{XX}(n+\kappa, n) &= \mathrm{E}\left[X(n+\kappa)X(n)^*\right] \\
&= \mathrm{E}\left[\sum_{k=1}^{K} A_k e^{j(\omega_0(n+\kappa)+\phi_k)} \sum_{l=1}^{K} A_l e^{-j(\omega_0 n+\phi_l)}\right] \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K} \mathrm{E}\left[A_k A_l\right] e^{j\omega_0\kappa}\,\mathrm{E}\left[e^{j(\phi_k-\phi_l)}\right] \\
&= \sum_{k=1}^{K} \mathrm{E}\left[A_k^2\right] e^{j\omega_0\kappa} \\
&= e^{j\omega_0\kappa}\sum_{k=1}^{K}\sigma_k^2 = c_{XX}(\kappa)
\end{aligned}
$$

- $X(n)$ is a wide-sense stationary process.

Linear filtered process plus noise

- The cross-spectrum between input and output is:

$$
\begin{aligned}
C_{YX}(e^{j\omega}) &= H(e^{j\omega})C_{XX}(e^{j\omega}) \\
&= H(e^{j\omega})\mathcal{F}\{c_{XX}(\kappa)\} \\
&= H(e^{j\omega})\left[\sum_{k=1}^{K}\sigma_k^2\right]2\pi\sum_{l}\delta(\omega-\omega_0+2\pi l) \\
&= 2\pi\left[\sum_{k=1}^{K}\sigma_k^2\right]\sum_{l}H(e^{j(\omega_0-2\pi l)})\delta(\omega-\omega_0+2\pi l)
\end{aligned}
$$

Linear filtered process plus noise

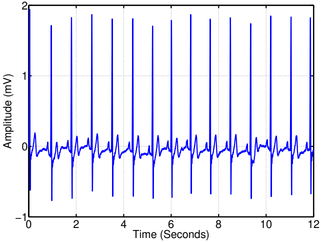
- The spectrum of the output is:

$$
\begin{aligned}
C_{YY}(e^{j\omega}) &= |H(e^{j\omega})|^2 C_{XX}(e^{j\omega}) + C_{VV}(e^{j\omega}) \\
&= 2\pi \left[\sum_{k=1}^{K} \sigma_k^2\right] \sum_{l} |H(e^{j(\omega_0 - 2\pi l)})|^2 \delta(\omega - \omega_0 + 2\pi l) \\
&\quad + C_{VV}(e^{j\omega})
\end{aligned}
$$

Outline

1. Random Variables and Stochastic Processes
2. **Brief Introduction to Elements of Estimation Theory**
   - Examples
   - Estimation, Estimators and Estimates
   - Properties of Estimators
   - Maximum Likelihood Estimation
   - Method of Least Squares
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation
8. Application: Airplane Tracking using Kalman Filter
9. Discrete Kalman Filter

- **Biomedicine:** heart rate, cardiac arrhythmia from heartbeat, etc.
- Radar system: range, speed, acceleration etc. of a target
- Speech recognition: value of each phoneme, direction of arrival, etc.



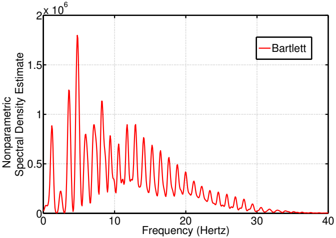


Figure: ECG signal in the time domain (left). Power Spectral Density Estimate of the EC
Signal (right).

- **Biomedicine:** heart rate, cardiac arrhythmia from heartbeat, etc.
- Radar system: range, speed, acceleration etc. of a target
- Speech recognition: value of each phoneme, direction of arrival, etc.

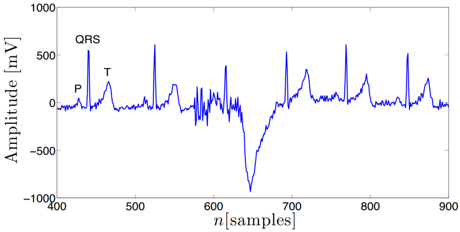


Figure: ECG with motion artifact

Introduction to Estimation Theory
Examples

- Biomedicine: heart rate, cardiac arrhythmia from heartbeat, etc.
- **Radar system:** range, speed, acceleration etc. of a target
- Speech recognition: value of each phoneme, direction of arrival, etc.

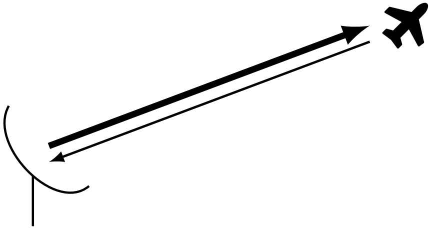


Figure: Radar system

- Biomedicine: heart rate, cardiac arrhythmia from heartbeat, etc.
- Radar system: range, speed, acceleration etc. of a target
- **Speech recognition:** value of each phoneme, direction of arrival, etc.

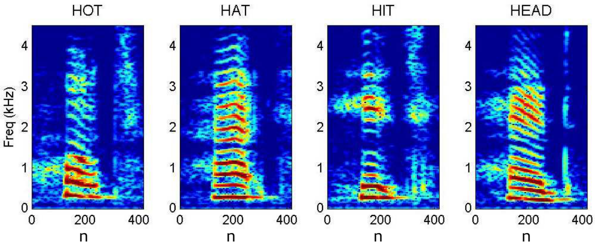
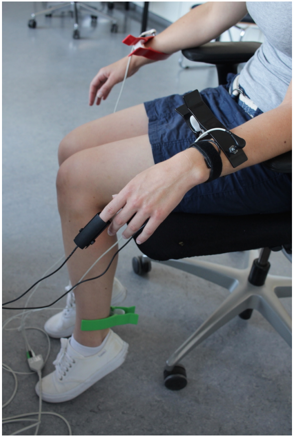


Figure: Spectrograms of different words (Source: Auditory Neuroscience)

Photoplethysmography (PPG) Example

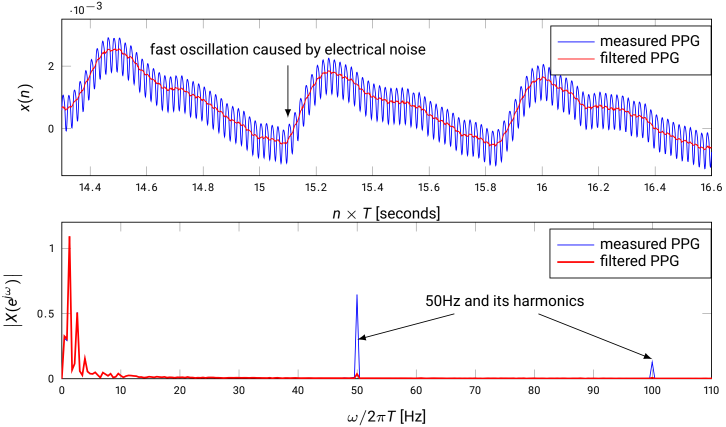
TECHNISCHE UNIVERSITÄT DARMSTADT



- The PPG signal can contribute information, e.g., about heart beat rate.
- Measurements are contaminated, e.g., by electrical power noise.
- Measuring PPG signal using a finger clip sensor in the SPG Biomedical Lab.

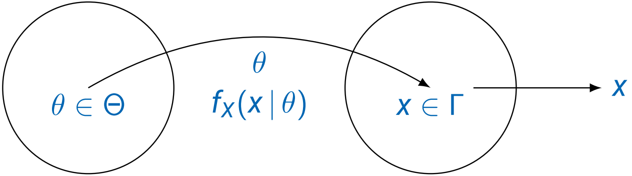
TECHNISCHE UNIVERSITÄT DARMSTADT

Band-stop filtering attenuates noise without distorting the PPG pulse.

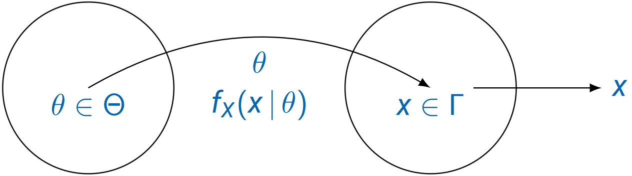


Estimation

- We have a parameter $\theta \in \Theta$ and a set of observations $x \in \Gamma$, which are linked through the pdf $f_X(x \mid \theta)$ (sometimes denoted $f_X(x; \theta)$).
- Goal of estimation theory is to find the value θ under which the observations x were generated.



Estimation



- The form of $f_X(x \mid \theta)$ may be
  1. Completely known.
  2. Known to a certain extent, i.e., the pdf may be symmetric and unimodal.
  3. Completely unknown.
- Two fundamental approaches to estimation exist
  1. Frequentist: θ is treated as an unknown deterministic constant.
  2. Bayesian: θ is treated as a RV with a known distribution.

Estimators and Estimates

TECHNISCHE UNIVERSITÄT DARMSTADT

- An *estimator* $\hat{\Theta}$ is a function of the RVs $X_1, \ldots, X_N$.
- An *estimate* $\hat{\theta}$ is a function of the observations $x_1, \ldots, x_N$.
- The difference?
  - x_t is a realisation of the RV X_t.
  - $\hat{\theta}$ is a realisation of the RV $\hat{\Theta}$.
- Say we are interested in estimating the expected value of X, $\mathrm{E}[X]$, from N i.i.d. observations.
  - An estimator for $\mathrm{E}[X]$ is the sample mean,

$$\hat{\Theta} = \frac{1}{N} \sum_{t=1}^{N} X_t.$$

  - An estimate of $\mathrm{E}[X]$ is given by the sample mean of a particular realisation of the observations,

$$\hat{\theta} = \frac{1}{N} \sum_{t=1}^{N} x_t.$$

Properties of Estimators
Bias and Variance

- The *bias* of an estimator is

$$\text{bias}[\hat{\Theta}] \triangleq \mathsf{E}[\hat{\Theta}] - \theta.$$

If the bias is zero, the estimator is *unbiased*.

- The *variance* of an estimator is

$$\text{Var}[\hat{\Theta}] = \mathsf{E}[(\hat{\Theta} - \mathsf{E}[\hat{\Theta}])^2] = \mathsf{E}[\hat{\Theta}^2] - \mathsf{E}[\hat{\Theta}]^2.$$

Properties of Estimators
Mean square error

- The *mean square error* (MSE) of an estimator is

$$\mathrm{MSE}[\hat{\Theta}] = \mathrm{E}[(\hat{\Theta} - \theta)^2].$$

- There is a relationship between the bias, variance and MSE of an estimator:

$$\begin{aligned}
\mathrm{MSE}[\hat{\Theta}] &= \mathrm{E}[((\hat{\Theta} - \mathrm{E}[\hat{\Theta}]) + (\mathrm{E}[\hat{\Theta}] - \theta))^2] \\
&= \mathrm{E}[(\hat{\Theta} - \mathrm{E}[\hat{\Theta}])^2] + \mathrm{E}[(\mathrm{E}[\hat{\Theta}] - \theta)^2] + 2\mathrm{E}[(\hat{\Theta} - \mathrm{E}[\hat{\Theta}])(\mathrm{E}[\hat{\Theta}] - \theta)] \\
&= \mathrm{Var}[\hat{\Theta}] + \mathrm{bias}[\hat{\Theta}]^2.
\end{aligned}$$

Properties of Estimators
Consistency

- An estimator is *consistent* if

$$\lim_{N\to\infty} \Pr\left[|\hat{\Theta} - \theta| \geq \epsilon\right] = 0 \quad \forall\, \epsilon > 0$$

where the estimator is based on N observations.

- An equivalent definition is

$$\lim_{N\to\infty} \Pr\left[|\hat{\Theta} - \theta| < \epsilon\right] = 1 \quad \forall\, \epsilon > 0$$

Why?

- Both these are equivalent to saying that the estimator *converges in probability* to its true value θ

$$\hat{\Theta} \underset{N\to\infty}{\longrightarrow} \theta.$$

Convergence is in the asymptotic sense, as the number of observations approaches ∞.

Example

TECHNISCHE UNIVERSITÄT DARMSTADT

Consider the sample mean as an estimator for the mean of a RV X with

- $\mathrm{E}[X] = \mu$
- $\mathrm{Var}[X] = \sigma^2$.

For N i.i.d. observations of X the estimator is $\bar{X} = \frac{1}{N}\sum_{t=1}^{N} X_t$.

1. Is the estimator linear?

   A linear estimator is of the form $\hat{\Theta} = \sum_{t=1}^{N} a_t X_t$. $\bar{X}$ is a linear estimator with $a_t = 1/N$.

Example

2. Is the estimator unbiased?

$$
\begin{aligned}
\text{bias}[\bar{X}] = \mathrm{E}[\bar{X}] - \mu = \mathrm{E}\left[\frac{1}{N}\sum_{t=1}^{N} X_t\right] - \mu \\
&= \frac{1}{N}\sum_{t=1}^{N} \mathrm{E}\left[X_t\right] - \mu \qquad (\mathrm{E}\left[\cdot\right] \text{ is a linear operation}) \\
&= \frac{1}{N}\sum_{t=1}^{N} \mu - \mu \qquad (X_t \text{ are i.i.d. with mean } \mu) \\
&= 0.
\end{aligned}
$$

So the estimator is unbiased.

Example

3. What is the variance of the estimator?

$$
\begin{aligned}
\text{Var}[\hat{\Theta}] &= \text{Var}\left[\frac{1}{N}\sum_{t=1}^{N} X_t\right] \\
&= \frac{1}{N^2}\sum_{t=1}^{N} \text{Var}\left[X_t\right] \qquad (X_t \text{ are i.i.d.}) \\
&= \frac{1}{N^2}\sum_{t=1}^{N} \sigma^2 = \frac{\sigma^2}{N}.
\end{aligned}
$$

4. What is the MSE of the estimator?

$$
\begin{aligned}
\text{MSE}[\hat{\Theta}] &= \text{Var}[\hat{\Theta}] + \text{bias}[\hat{\Theta}]^2 \\
&= \frac{\sigma^2}{N}.
\end{aligned}
$$

Example

TECHNISCHE UNIVERSITÄT DARMSTADT

5. Is the estimator consistent? Start with Chebyshev's inequality

$$\Pr\left[\left|\bar{X} - \mathrm{E}\left[\bar{X}\right]\right| \geq \delta\right] \leq \frac{\mathrm{Var}\left[\bar{X}\right]}{\delta^2}, \qquad \delta > 0$$

and substitute in $\mathrm{E}\left[\bar{X}\right] = \mu$ and $\mathrm{Var}\left[\bar{X}\right] = \sigma^2/N$,

$$\Pr\left[\left|\bar{X} - \mu\right| \geq \delta\right] \leq \frac{\sigma^2}{N\delta^2}.$$

Taking the limit as $N \to \infty$ gives

$$\lim_{N\to\infty} \Pr\left[\left|\bar{X} - \mu\right| \geq \delta\right] \leq 0$$

and since a probability, i.e., the LHS of the above equation, cannot be less than zero,

$$\lim_{N\to\infty} \Pr\left[\left|\bar{X} - \mu\right| \geq \delta\right] = 0, \qquad \delta > 0$$

which is the definition of a consistent estimator. This is also known as the weak law of large numbers (WLLN): If X_t are i.i.d. RVs with mean μ, then by the WLLN $\bar{X}$ converges in probability to μ.

Maximum Likelihood Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

- For a set of observations $x_1, \ldots, x_N$ the *likelihood function* (LF) is defined as

$$L(\theta; x_1, \ldots, x_N) = f_{X_1,\ldots,X_N}(x_1, \ldots, x_N \mid \theta)$$

- For i.i.d. observations the likelihood reduces to the product of the marginal pdfs

$$L(\theta; x_1, \ldots, x_N) = \prod_{t=1}^{N} f_{X_t}(x_t \mid \theta)$$

- $L(\theta; x_1, \ldots, x_N)$ is a measure of the probability that x was generated from $f_X(x \mid \theta)$ for a particular value of θ.

Maximum Likelihood Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

$$L(\theta; x_1, \ldots, x_N) = \prod_{t=1}^{N} f_{X_t}(x_t \mid \theta)$$

- Given a set of observations, L changes with θ.
- Some values of θ are then 'more likely' to have been responsible for generating x.
- Maximising $L(\theta; x_1, \ldots, x_N)$ w.r.t. θ gives the *maximum likelihood* (ML) estimate of θ.
- The MLE is usually consistent and for Gaussian observations is the MV estimator.

Method of Least Squares

The Sum of Squared Errors

- We assume an observable signal $Y(n)$, $n \in \mathbb{Z}$, e.g. an output of a system, as a linear superposition of K known inputs $x_{n1}, \ldots, x_{nK}$ for a given $n \in \mathbb{Z}$ and an unknown measurement error $Z(n)$, $n \in \mathbb{Z}$. For N observations, we define

$$y_n = \sum_{i=1}^{K} \theta_i x_{ni} + z_n, \qquad n = 1, \ldots, N.$$

- To estimate the unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\mathsf{T}$, we define

$$\mathcal{S}(\boldsymbol{\theta}) = \mathcal{S}(\theta_1, \ldots, \theta_K) = \sum_{n=1}^{N} z_n^2 = \sum_{n=1}^{N} (y_n - \sum_{i=1}^{K} \theta_i x_{ni})^2$$

as the **sum of the squared errors** (SSE).

Method of Least Squares

The Least Squares Estimate

- The **least squares estimate** (LSE) is the vector $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^\mathsf{T}$ that minimizes $\mathcal{S}(\boldsymbol{\theta})$.
- The solution is given by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y},$$

where

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NK} \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}.$$

Least Squares Estimation

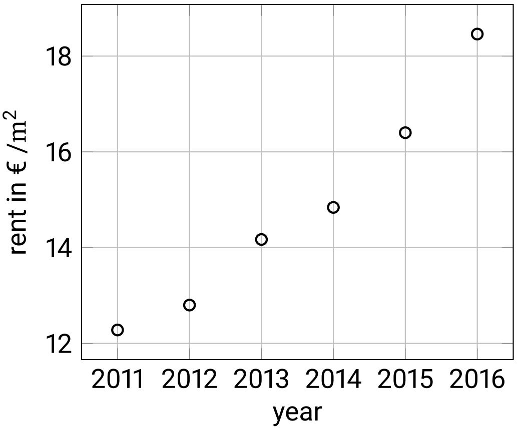
Rent in Rhine-Main Region

- Rent in Rhine-Main Region from 2011 to 2016

Year	Rent in € $/\mathrm{m}^2$
2011	12.28
2012	12.80
2013	14.17
2014	14.84
2015	16.40
2016	18.46

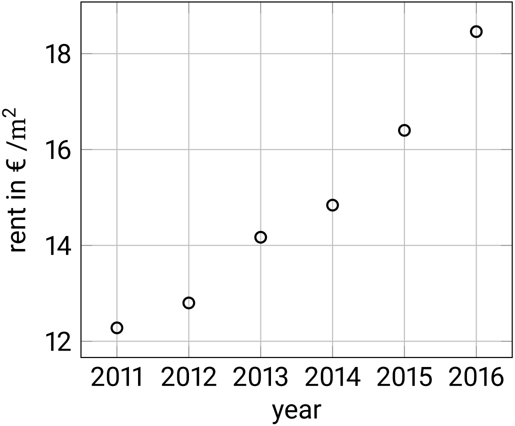
- Goal:
  - Model the evolution of rent over time
  - predict future values

Least Squares Estimation
Rent in Rhine-Main Region



Least Squares Estimation

Rent in Rhine-Main Region



$\Rightarrow$ use linear model to model the data

Least Squares Estimation

Rent in Rhine-Main Region

- Linear model relating rent y and year x

$$y = \theta_1 x + \theta_0$$

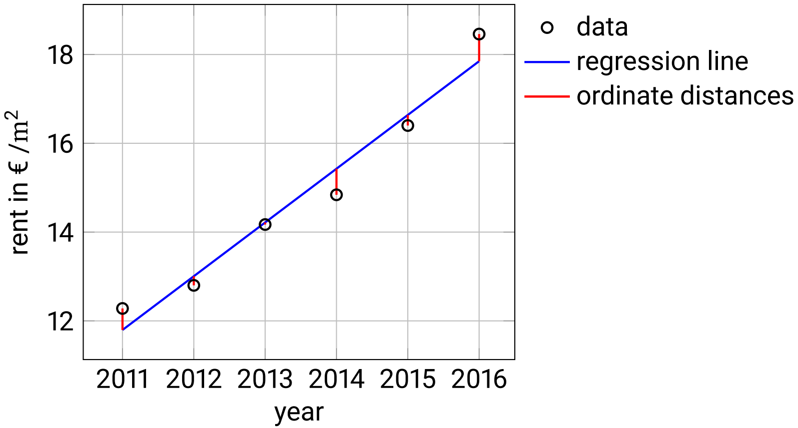
- Problem is also called *linear regression*
- Estimate the model parameters θ_0 and θ_1

$$\hat{y} = \hat{\theta}_1 x + \hat{\theta}_0$$

- Sum of squared errors $(\hat{y}_i - y_i)^2$ of each point pair (x_i, y_i) is minimized
- Method is called *least squares estimation*

Least Squares Estimation

Rent in Rhine-Main Region



Least Squares Estimation

Rent in Rhine-Main Region

TECHNISCHE UNIVERSITÄT DARMSTADT

- In this example:

$$x = \text{year} - 2005$$

- Estimated model

$$\hat{y} = 1.2106x + 4.535$$

- What happens in the future? What about the rent in 2020?

Least Squares Estimation

Rent in Rhine-Main Region

- In this example:

$$x = \text{year} - 2005$$

- Estimated model

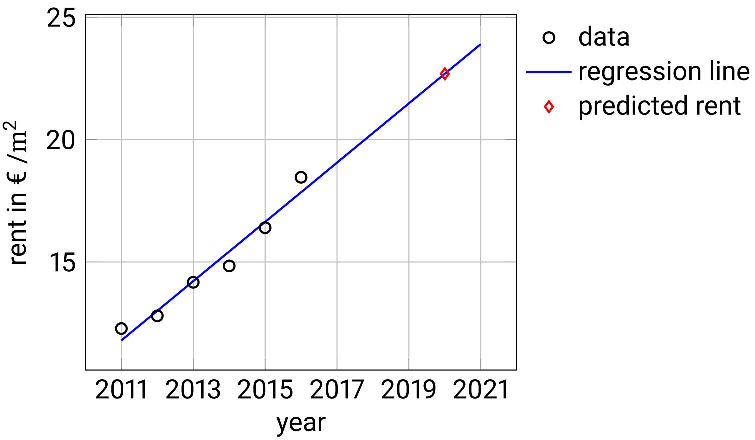
$$\hat{y} = 1.2106x + 4.535$$

- What happens in the future? What about the rent in 2020?
- Use the estimated model to predict the rent in 2020

$$\hat{y}(2020) = 1.2106(2020 - 2005) + 4.535$$

Least Squares Estimation

Rent in Rhine-Main Region



Outline I

1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. **The Finite Fourier Transform**
   - Definition and Properties of the Finite Fourier Transform
   - Statistical Properties of the Finite Fourier Transform
   - Relation to the Power Spectral Density
   - Relation to the Cross-Power Spectral Density
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation

Outline II

TECHNISCHE UNIVERSITÄT DARMSTADT

8. **Application: Airplane Tracking using Kalman Filter**
9. **Discrete Kalman Filter**

Definition of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

Let $x(0), \dots, x(N-1)$ be realisations of a stationary random process $X(n)$, $n = 0, 1, 2, \dots N-1$.

The Finite Fourier Transform

$$X_N(e^{j\omega}) = \sum_{n=0}^{N-1} X(n) e^{-j\omega n}, \quad -\infty < \omega < \infty$$

Inverse transform

$$X(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_N(e^{j\omega}) e^{j\omega n} \, d\omega, \quad n = 0, 1, \dots, N-1$$

Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

The properties are similar to those of the DTFT

- **Periodicity** $X_N(e^{j(\omega+k2\pi)}) = X_N(e^{j\omega}) \quad \forall\, k \in \mathbb{N}$
- **Symmetry** $X_N(e^{-j\omega}) = X_N(e^{j\omega})^* \quad \forall\, X(n) \in \mathbb{R}$
- **Linearity** $\mathcal{F}\{aX(n) + bY(n)\} = aX_N(e^{j\omega}) + bY_N(e^{j\omega})$

Properties of the Finite Fourier Transform

The finite Fourier transform can be taken at discrete frequencies $\omega_k = 2\pi k/N$, $k = 0, .., N-1$

$$X_N(e^{j\frac{2\pi k}{N}}) = \sum_{n=0}^{N-1} X(n)\mathrm{e}^{-j2\pi kn/N}, \quad k = 0, 1, \ldots, N-1$$

which can be computed by means of the Fast Fourier transform (FFT).

Statistical Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

Example: Finite Fourier Transform of a white Gaussian process

Preliminaries:

- $X(0), \ldots, X(N-1)$ are i.i.d. real-valued random variables
- $X(n) \sim \mathcal{N}(0, 1)$
- $N = 2^r$ with $r \in \mathbb{N}$

The real and imaginary part of the Finite Fourier Transform can be expressed as:

$$\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\} = \sum_{n=0}^{N-1} X(n) \cos(2\pi kn/N)$$

and

$$\mathrm{Im}\left\{X_N(e^{j2\pi k/N})\right\} = -\sum_{n=0}^{N-1} X(n) \sin(2\pi kn/N)$$

Statistical Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

- Real and imaginary parts are sums of independent normal random variables and so have a normal distribution
- Find mean and variance of them to specify their distribution
- The mean of the real part is given by

$$\begin{aligned} \mathrm{E}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}\right] &= \mathrm{E}\left[\sum_{n=0}^{N-1} X(n)\cos(2\pi kn/N)\right] \\ &= \sum_{n=0}^{N-1} \mathrm{E}\left[X(n)\right]\cos(2\pi kn/N) \\ &= 0 \end{aligned}$$

- The expected value of the imaginary part can also be determined to $\mathrm{E}\left[\mathrm{Im}\left\{X_N(e^{j2\pi k/N})\right\}\right] = 0$

Statistical Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

- The variance of the real part is given by

$$
\begin{aligned}
\mathrm{Var}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}\right] &= \mathrm{E}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}^2\right] - \mathrm{E}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}\right]^2 \\
&= \mathrm{E}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}^2\right] \\
&= \mathrm{E}\left[\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X(n)X(m)\cos(2\pi kn/N)\cos(2\pi km/N)\right] \\
&= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \mathrm{E}\left[X(n)X(m)\right]\cos(2\pi kn/N)\cos(2\pi km/N)
\end{aligned}
$$

Statistical Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

- Since $X(0), \ldots, X(N-1)$ are independent random variables, we have

$$\mathrm{E}\left[X(n)X(m)\right] = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases}$$

Using this and the trigonometric identity $\cos(\theta)^2 = \frac{1}{2}(1 + \cos(2\theta))$we get

$$\begin{aligned} \mathrm{Var}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}\right] &= \sum_{n=0}^{N-1} \cos(2\pi kn/N)^2 \\ &= \frac{1}{2}\sum_{n=0}^{N-1}(1 + \cos(4\pi kn/N)) \\ &= \frac{N}{2} + \frac{1}{2}\sum_{n=0}^{N-1} \cos(4\pi kn/N) \end{aligned}$$

Statistical Properties of the Finite Fourier Transform

- The last result can be simplified to

$$\text{Var}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}\right] = N/2 \quad \text{for } k = 1, \ldots, N/2 - 1$$

- Combining the results leads to

$$\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\} \sim \mathcal{N}\left(0, N/2\right) \quad \text{for } k = 1, \ldots, N/2 - 1$$

and

$$\mathrm{Im}\left\{X_N(e^{j2\pi k/N})\right\} \sim \mathcal{N}\left(0, N/2\right) \quad \text{for } k = 1, \ldots, N/2 - 1$$

Note: For $k = 0$ and $k = N/2$ the Finite Fourier Transform is purely real and $\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\} \sim \mathcal{N}\left(0, N\right)$

Statistical Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

- Correlation between real and imaginary parts of the Finite Fourier Transform

$$
\begin{aligned}
\mathrm{Cov}\left[\mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\}, \mathrm{Im}\left\{X_N(e^{j2\pi k/N})\right\}\right] \\
&= -\sum_{n=0]}^{N-1}\sum_{m=0}^{N-1} \mathrm{E}\left[X(n)X(m)\right]\cos(2\pi kn/N)\sin(2\pi km/N) \\
&= -\frac{1}{2}\sum_{n=0}^{N-1}\sin(4\pi kn/N) = 0, \qquad k = 0, \ldots, N/2
\end{aligned}
$$

- Real and imaginary parts of the Finite Fourier Transform are uncorrelated
- The distribution of the Finite Fourier Transform of a white Gaussian stationary process can be written as

$$
\begin{bmatrix} \mathrm{Re}\left\{X_N(e^{j2\pi k/N})\right\} \\ \mathrm{Im}\left\{X_N(e^{j2\pi k/N})\right\} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{N}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \qquad k = 1, \ldots, N/2-1
$$

Statistical Properties of the Finite Fourier Transform

- Result for the general case of a real-valued stationary random process with an arbitrary $\mu_x, c_{XX}(\kappa)$ and $C_{XX}(e^{j\omega})$
- The asymptotic results, i.e. when $N \to \infty$, are given without proof
  - The asymptotic distribution is

$$\begin{bmatrix} \mathrm{Re}\left\{X_N(e^{j\omega})\right\} \\ \mathrm{Im}\left\{X_N(e^{j\omega})\right\} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{N}{2} C_{XX}(e^{j\omega}) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right), \quad \omega \in (0, \pi)$$

  - For $\omega = 0$ and $\omega = \pi$ the Finite Fourier Transform is purely real and we have a different asymptotic distribution, which is given by

$$\mathrm{Re}\left\{X_N(e^{j\omega})\right\} \sim \mathcal{N}\left(N\mu_x\delta(\omega), NC_{XX}(e^{j\omega})\right)$$

  - For fixed frequencies $0 \leq \omega_{(1)} < \omega_{(2)} < \cdots < \omega_{(M)} \leq \pi$ the random variables $X_N(e^{j\omega_{(1)}}), X_N(e^{j\omega_{(2)}}), \ldots, X_N(e^{j\omega_{(M)}})$ are asymptotically for $N \to \infty$ independently distributed.

Statistical Properties of the Finite Fourier Transform

Two important modifications of random processes

- Division into segments
  - Divide random process $X(n), n = 0, \ldots, N-1$ into L segments of length M with $N = ML$
  - The Finite Fourier Transform of each segment l is given by

$$X_M(e^{j\omega}, l) = \sum_{n=0}^{M-1} X(n-(l-1)M)e^{-j\omega n}, \quad l = 1, \ldots, L$$

  - The asymptotic distribution of $X_M(e^{j\omega}, l)$ is the same as for the non-segmented process
  - The random variables $X_M(e^{j\omega}, l)$ are for $l = 1, \ldots, L$ asymptotically as $N \to \infty$ independently distributed.
  - Both $M, N \to \infty$ while the number of segments L is finite

Statistical Properties of the Finite Fourier Transform

TECHNISCHE UNIVERSITÄT DARMSTADT

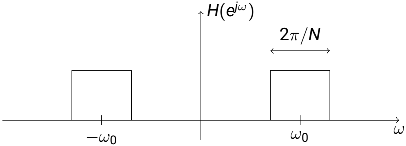
- Windowing the process
  - A window $w(n) \neq 0$ for $n = 0, \ldots, N-1$ is applied to the random process
  - The Finite Fourier Transform of the windowed process is given by

$$X_{N,w}(e^{j\omega}) = \sum_{n=0}^{N-1} w(n) X(n) e^{-j\omega n}$$

  - The asymptotic distribution of $X_{N,w}(e^{j\omega})$ is similar as for the non-windowed process
  - Except: The variance is multiplied by the factor $\sum_{n=0}^{N-1} w(n)^2$ and the mean at $\omega = 0$ is multiplied by the factor $\sum_{n=0}^{N-1} w(n)$

Statistical Properties of the Finite Fourier Transform

- Assume an ideal narrow bandpass filter with
  $H(e^{j\omega}) = \begin{cases} 1, & |\omega \pm \omega_o| < \pi/N \\ 0, & \text{else} \end{cases}$

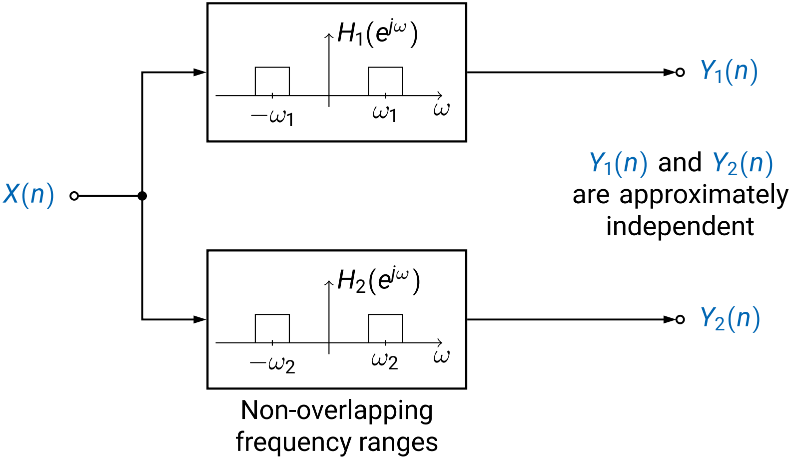


If $X(n)$ is filtered, then $Y(n) = \sum h(m)X(n-m)$. For $(\frac{2\pi}{N} \to 0)$

$$Y(n) \approx \frac{1}{N}X_N(e^{j\frac{2\pi s}{N}})e^{j2\pi sn/N} + \frac{1}{N}X_N(e^{-j\frac{2\pi s}{N}})e^{-j2\pi sn/N}$$

with $2\pi s/N \approx \omega_o$, then $Y(n)$ is approximately $\mathcal{N}(0, \frac{1}{N}C_{XX}(e^{j\omega_o}))$ distributed.

Statistical Properties of the Finite Fourier Transform



Relation to the Power Spectral Density

- Power Spectral Density was introduced as the Fourier Transform of the SOMF
- PSD is also related to the expectation of the magnitue squared of the Finite Fourier Transform when $N \to \infty$

Calculation of the Average Power

- $X(n, \zeta_i), n \in \mathbb{N}$ is a sample function of a real-valued random process X(n)
- Define the truncated version of this sample as

$$x_N(n) = \begin{cases} X(n, \zeta_i) & n < N, N \in \mathbb{N} \\ 0 & \text{elsewhere} \end{cases}$$

- The Finite Fourier Transform of $x_N(n)$is given by

$$X_N(e^{j\omega}) = \sum_{n=0}^{N-1} x_N(n) e^{-j\omega n}$$

Relation to the Power Spectral Density

TECHNISCHE UNIVERSITÄT DARMSTADT

- The normalized energy of the truncated signal can be calculated as

$$E_N = \sum_{n=0}^{N-1} x_N(n)^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X_N(e^{j\omega})|^2 d\omega$$

- Parseval's theorem has been used to obtain the above result
- Dividing E_N by the observation time N, we get the average Power P_N in $x_N(n)$

$$P_N = \frac{1}{N} \sum_{n=0}^{N-1} x_N(n)^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{N} |X_N(e^{j\omega})|^2 d\omega$$

Relation to the Power Spectral Density

- To obtain the power of the entire sample function, we must consider the limit as $N \to \infty$
- If we want the average power of the random process, we must take the expectation
- The average power of $X(n)$ is defined by

$$P_{XX} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{E}\left[X(n)^2\right] = \lim_{N\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{N} \mathrm{E}\left[\left|X_N(e^{j\omega})\right|^2\right] d\omega$$

Relation to the Power Spectral Density

TECHNISCHE UNIVERSITÄT DARMSTADT

- We can also determine P_{XX} by a frequency domain integration
- Let us define the PSD as

$$S_{XX}(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[|X_N(e^{j\omega})|^2\right]$$

- The average power is then given by

$$P_{XX} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(e^{j\omega})d\omega = r_{xx}(0)$$

Relation to the Power Spectral Density

TECHNISCHE UNIVERSITÄT DARMSTADT

Example:

- Consider the random process $X(n) = A\cos(\omega_0 n + \Phi)$
- A and ω_0 are real-valued constants and Φ is uniform distributed on $[0, \pi/2)$
- Determine P_{XX} of $X(n)$ using

$$P_{XX} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{E}\left[X(n)^2\right]$$

- The mean-squared value of $X(n)$ is

$$\begin{aligned}
\mathrm{E}\left[X(n)^2\right] &= \mathrm{E}\left[A^2\cos(\omega_0 n + \Phi)^2\right] \\
&= \mathrm{E}\left[A^2(1 + \cos(2\omega_0 n + 2\Phi))/2\right] \\
&= \frac{A^2}{2} + \frac{A^2}{2}\mathrm{E}\left[\cos(2\omega_0 n + 2\Phi)\right]
\end{aligned}$$

Relation to the Power Spectral Density

TECHNISCHE UNIVERSITÄT DARMSTADT

$$
\begin{aligned}
\mathrm{E}\left[X(n)^2\right] &= \frac{A^2}{2} + \frac{A^2}{2}\mathrm{E}\left[\cos(2\omega_0 n + 2\Phi)\right] \\
&= \frac{A^2}{2} + \frac{A^2}{2}\int_0^{\pi/2} \cos(2\omega_0 n + 2\varphi)\frac{2}{\pi}d\varphi \\
&= \frac{A^2}{2} + \frac{A^2}{2}\frac{2}{\pi}\left[\frac{\sin(2\omega_0 n + \pi) - \sin(2\omega_0 n)}{2}\right] \\
&= \frac{A^2}{2} - \frac{A^2}{\pi}\sin(2\omega_0 n)
\end{aligned}
$$

- $X(n)$ is non stationary because $\mathrm{E}\left[X(n)^2\right]$ is a function of time
- The time average of the above expression is

$$
\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left[\frac{A^2}{2} - \frac{A^2}{\pi}\sin(2\omega_0 n)\right] = \frac{A^2}{2}
$$

Relation to the Cross-Power Spectral Density

TECHNISCHE UNIVERSITÄT DARMSTADT

- $x_N(n)$ and $y_N(n)$ are observations of the random processes $X(n)$ and $Y(n)$
- Their Finite Fourier Transforms are given by

$$X_N(e^{j\omega}) = \sum_{n=0}^{N-1} x_N(n) e^{-j\omega n}$$

and,

$$Y_N(e^{j\omega}) = \sum_{n=0}^{N-1} y_N(n) e^{-j\omega n}$$

- The cross-power spectral density function is given as

$$S_{XY}(e^{j\omega}) = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[X_N(e^{j\omega}) Y_N(e^{j\omega})^*\right]$$

- The total average cross power is

$$P_{XY} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{XY}(e^{j\omega}) d\omega$$

Outline

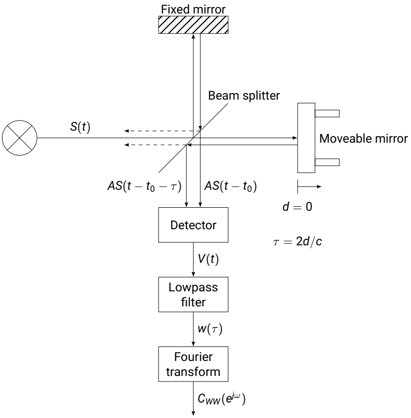
1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. **Introduction to Digital Spectral Analysis**
   - Why Spectral Analysis?
   - Applications
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation
8. Application: Airplane Tracking using Kalman Filter
9. Discrete Kalman Filter

Why Spectral Analysis?

TECHNISCHE UNIVERSITÄT DARMSTADT

1. To provide useful descriptive statistics.
2. As a diagnostic tool to indicate which further analysis method might be relevant.
3. To check postulated theoretical models.

Why Spectral Analysis?



A Michelson Interferometer

The objective is to get a good estimate of the spectrum without moving the mirror at large d.

Applications

Examples:

- **Physics** In spectroscopy we wish to investigate the distribution of energy in a radiation field as a function of frequency.

Example: Michelson Interferometer

$$X(t) = A\left(S(t-t_0) + S(t-t_0-\tau)\right)$$
$$V(t) = |X(t)|^2 = A^2|S(t-t_0) + S(t-t_0-\tau)|^2$$

$$\begin{aligned} w(\tau) = \mathrm{E}\left[V(t)\right] &= A^2\left[c_{SS}(0) + c_{SS}(\tau) + c_{SS}(-\tau) + c_{SS}(0)\right] \\ &= A^2\left[2c_{SS}(0) + 2\mathrm{Re}\{c_{SS}(\tau)\}\right] \end{aligned}$$

Performing a Fourier transform leads to

$$C_{WW}(e^{j\omega}) = \mathcal{F}\{c_{WW}(\tau)\} = A^2\left[2c_{SS}(0)\delta(\omega) + C_{SS}(e^{j\omega}) + C_{SS}(e^{-j\omega})\right]$$

Applications

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Speech Processing** Spectral analysis of speech signals allows automated speech recognition.

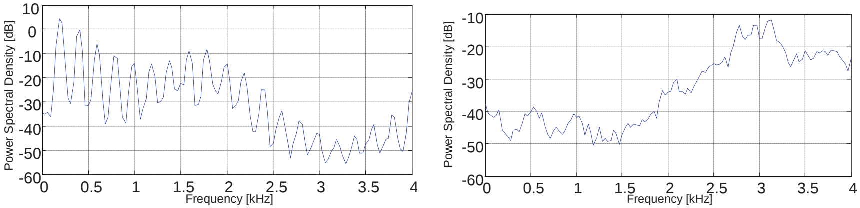


Figure: Voiced speech signal (vowel) and unvoiced speech signal (consonant)

Applications

- **Electrical Engineering** Measurement of the power in various frequency bands of some electromagnetic signal of interest. In radar, there is the problem of signal detection and interpretation.
- **Acoustics** The power spectrum plays the role of a descriptive statistic. We may use the spectrum to calculate the direction of arrival (DOA) of an incoming wave or to characterise the source.
- **Geophysics** Earthquakes give rise to a number of different waves such as pressure waves, shear waves and surface waves. These waves are received by an array of sensors.
- **Mechanical Engineering** Detection of abnormal combustion in a spark ignition engine. The power spectrum is a descriptive statistic.
- **Medicine** EEG (electroencephalogram) and ECG (electrocardiogram) analysis.

Spectrum Estimation using Bandpass Filters

Example: A simple device for measuring the spectrum.

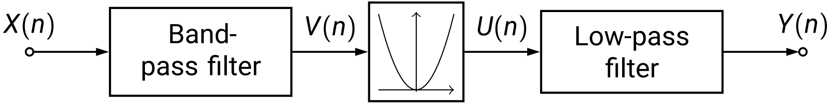
$X(n)$ is a stochastic process with covariance function $c_{XX}(n)$ and spectrum $C_{XX}(e^{j\omega})$. Assume $\mathrm{E}[X(n)] = 0$. If we transform $X(n)$ by

$$Y(n) = X(n)^2$$

then

$$\mathrm{E}[Y(n)] = \mathrm{E}\left[X(n)^2\right] = c_{XX}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{XX}(e^{j\omega})d\omega$$

Spectrum Estimation using Bandpass Filters



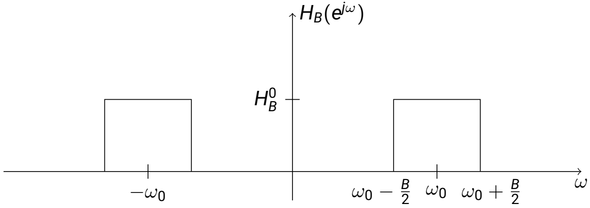
The bandpass filter has the transfer function

$$H_B(e^{j\omega}) = \begin{cases} H_B^0, & |\omega \pm \omega_0| \leq \frac{B}{2} \text{ for } |\omega| \leq \pi \\ 0, & \text{else} \end{cases}$$

H_B^0 will be specified later. The lowpass filter to calculate the mean is

$$h_l(n) = \begin{cases} \frac{1}{N}, & n = 0, 1, 2, \ldots, N-1 \\ 0, & \text{else} \end{cases} .$$

Spectrum Estimation using Bandpass Filters



$$H_L(e^{j\omega}) = \frac{\sin(\frac{\omega N}{2})}{N\sin(\omega/2)} e^{-j\omega\frac{N-1}{2}}$$

$$C_{VV}(e^{j\omega}) = |H_B(e^{j\omega})|^2 C_{XX}(e^{j\omega})$$

Spectrum Estimation using Bandpass Filters

The output process is given by

$$Y(n) = \sum_{m=0}^{N-1} \frac{1}{N} U(n-m) = \frac{1}{N} \sum_{m=0}^{N-1} V(n-m)^2.$$

Spectrum Estimation using Bandpass Filters

TECHNISCHE UNIVERSITÄT DARMSTADT

With the mean

$$
\begin{aligned}
\mu_Y &= \mathrm{E}\left[Y(n)\right] = \frac{1}{N}\sum_{m=0}^{N-1}\mathrm{E}\left[V(n-m)^2\right] = c_{VV}(0) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{VV}(e^{j\omega})\,d\omega \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} |H_B(e^{j\omega})|^2 C_{XX}(e^{j\omega})\,d\omega \\
&= \frac{(H_B^0)^2}{\pi}\int_{\omega_0-B/2}^{\omega_0+B/2} C_{XX}(e^{j\omega})\,d\omega \\
&\approx (H_B^0)^2\frac{B}{\pi}C_{XX}(e^{j\omega_0}).
\end{aligned}
$$

Outline

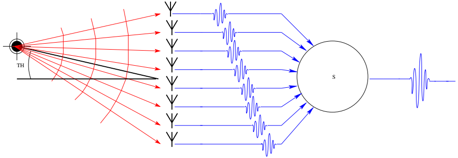
TECHNISCHE UNIVERSITÄT DARMSTADT

1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. **Application: Spatial Spectra Estimation - Direction-Finding**
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation
8. Application: Airplane Tracking using Kalman Filter
9. Discrete Kalman Filter

Application: Spatial Spectra Estimation

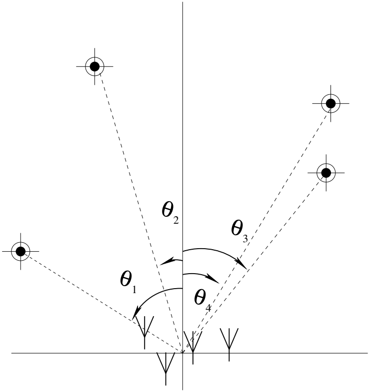
For a variety of applications (communications, radar, sonar, etc), exact knowledge of a signals location in space is desired.

- Principles of spectrum estimation can also be applied to *spatial spectra*.
- The spatial frequency of a signal translates to its *Direction-Of-Arrival (DOA)*.
- The *DOA* can be estimated using an array of sensors, extracting the direction-dependent phase delay of the signals.



Nonparametric approach - Beamformer

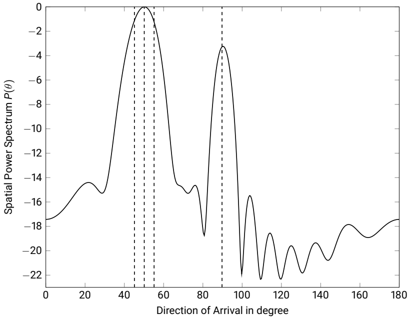
Assume there are p sources in space impinging on an M-element sensor array ($M > p$). Using a spatial periodogram (=*beamformer*), one can estimate the power from each direction without assumptions on the nature of the signals.



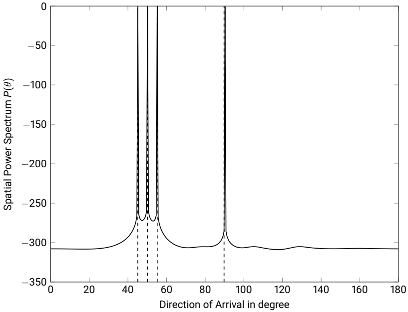
Performance Comparison

TECHNISCHE UNIVERSITÄT DARMSTADT

Experiment: 4 closely spaced sources, 20dB SNR, $M = 12$



Nonparametric Approach



Model-based Approach

Note: The beamformer can not resolve all sources!

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. **Non-Parametric Spectrum Estimation**
   - Consistency of Spectral Estimates
   - The Periodogram
   - Averaging Periodograms
   - Smoothing the Periodogram
   - A General Class of Spectral Estimators
   - Estimation of the Spectrum using the Sample Covariance
   - Cross-Spectrum Estimation
7. Parametric Spectrum Estimation

Consistency of Spectral Estimates

TECHNISCHE UNIVERSITÄT DARMSTADT

Example: Sample Mean

- Assume $X(0), \ldots, X(N-1)$ be real-valued independent variables each distributed as $\mathcal{N}(\mu_X, \sigma_X^2)$
- Determine the bias and the variance of the sample mean estimator

$$\hat{\mu}_X = \frac{1}{N} \sum_{n=0}^{N-1} X(n)$$

- Random variables are independent, thus the expected value is calculated as

$$\mathrm{E}\left[\hat{\mu}_X\right] = \mathrm{E}\left[\frac{1}{N} \sum_{n=0}^{N-1} X(n)\right] = \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{E}\left[X(n)\right] = \frac{1}{N} \sum_{n=0}^{N-1} \mu_X = \mu_X$$

Consistency of Spectral Estimates

TECHNISCHE UNIVERSITÄT DARMSTADT

- Bias: Difference of the expected value of the estimator and the true one
  In this case: $\mathrm{bias}\left[\hat{\mu}_X\right] = \mathsf{E}\left[\hat{\mu}_X\right] - \mu_X$
- The variance can be computed with the identity $\mathrm{Var}\left[aX\right] = a^2\mathrm{Var}\left[X\right]$ as

$$\mathrm{Var}\left[\hat{\mu}_X\right] = \mathrm{Var}\left[\frac{1}{N}\sum_{n=0}^{N-1} X(n)\right] = \frac{1}{N^2}\mathrm{Var}\left[\sum_{n=0}^{N-1} X(n)\right] = \frac{\sigma_X^2}{N}$$

- The mean square error (MSE) is defined as $\mathrm{MSE}(\hat{\mu}_X) = \mathsf{E}\left[(\hat{\mu}_X - \mu_X)^2\right]$
- The MSE can be rewritten in terms of bias and variance as follows

$$\mathrm{MSE}(\hat{\mu}_X) = \mathrm{Var}\left[\hat{\mu}_x\right] + \mathrm{bias}(\hat{\mu}_x)^2$$

- The estimator $\hat{\mu}_X$ is called consistent if it converges in probability as $N \to \infty$. A sufficient condition for consistency is:

$$\lim_{N\to\infty} \mathrm{MSE}(\hat{\mu}_X) = 0$$

Consistency of Spectral Estimates

TECHNISCHE UNIVERSITÄT DARMSTADT

- An estimator $\hat{C}_{XX}(e^{j\omega})$ of the spectrum $C_{XX}(e^{j\omega})$ is a random variable
- It is impossible to design an ideal estimator
- We wish to design an estimator for which bias and variance decrease when N increases
- A sufficient condition for consistency is

$$\lim_{N\to\infty} \mathrm{MSE}(\hat{C}_{XX}(e^{j\omega})) = 0$$

- This is equivalent to

$$\lim_{N\to\infty} \mathrm{Var}(\hat{C}_{XX}(e^{j\omega})) = 0 \quad \text{and} \quad \lim_{N\to\infty} \mathrm{bias}(\hat{C}_{XX}(e^{j\omega})) = 0$$

- If the estimator is biased it results in a systematic error
- If the variance is non zero, it results in a random error

The Periodogram

The Periodogram

$$I_{XX}^{N}(e^{j\omega}) \quad = \quad \frac{1}{N}|X_N(e^{j\omega})|^2 = \frac{1}{N}\left|\sum_{n=0}^{N-1} X(n)e^{-j\omega n}\right|^2$$

is a candidate estimator for $C_{XX}(e^{j\omega})$
It was introduced by Schuster (1898) as a tool for the identification of hidden periodicities. Indeed, in the case of

$$X(n) = \sum_{j=1}^{J} a_j \cos(\omega_j n + \phi_j), \quad n = 0, \ldots, N-1, \; N \text{ large}$$

$I_{XX}^{N}(e^{j\omega})$ has peaks at frequencies $\omega = \omega_j$, $j = 1, \ldots, J$. Note that $I_{XX}^{N}(e^{j\omega})$ has the same symmetry, non-negativity and periodicity properties as $C_{XX}(e^{j\omega})$.

The Periodogram

Distribution of the periodogram

The periodogram ordinates $I_{XX}^N(e^{j\omega})$ for data values $X(n)$ are independently distributed as

$$I_{XX}^N(e^{j\omega}) \sim \begin{cases} \frac{C_{XX}(e^{j\omega})}{2}\chi_2^2, & \omega \neq \pi k \\ C_{XX}(e^{j\omega})\chi_1^2, & \omega = \pi k \end{cases} \qquad k \in \mathbb{Z} \ .$$

The Periodogram

Mean of the Periodogram

$$
\begin{aligned}
&\mathrm{E}\left[I_{XX}^{N}(e^{j\omega})\right] \\
&= \mathrm{E}\left[\frac{1}{N}\sum_{n=0}^{N-1} X(n)e^{-j\omega n}\sum_{m=0}^{N-1} X(m)^{*}e^{j\omega m}\right] \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}\mathrm{E}\left[X(n)X(m)^{*}\right]e^{-j\omega(n-m)} \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}(c_{XX}(n-m)+\mu_X\mu_X^{*})e^{-j\omega(n-m)} \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}c_{XX}(n-m)e^{-j\omega(n-m)}+\frac{1}{N}|\Delta^{N}(e^{j\omega})|^{2}|\mu_X|^{2} \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}e^{-j\omega(n-m)}\frac{1}{2\pi}\int_{-\pi}^{\pi}C_{XX}(e^{j\lambda})e^{j\lambda(n-m)}\,d\lambda+\frac{1}{N}|\Delta^{N}(e^{j\omega})|^{2}|\mu_X|^{2}
\end{aligned}
$$

The Periodogram

$$
\begin{aligned}
\mathrm{E}\left[I_{XX}^{N}(e^{j\omega})\right] &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} e^{-j\omega(n-m)} \int_{-\pi}^{\pi} C_{XX}(e^{j\lambda}) e^{j\lambda(n-m)}\,\frac{d\lambda}{2\pi} + \frac{1}{N}|\Delta^{N}(e^{j\omega})|^2|\mu_X|^2 \\
&= \frac{1}{N}\int_{-\pi}^{\pi} |\Delta^{N}(e^{j(\omega-\lambda)})|^2 C_{XX}(e^{j\lambda})\,\frac{d\lambda}{2\pi} + \frac{1}{N}|\Delta^{N}(e^{j\omega})|^2|\mu_X|^2 \\
&= \frac{|\Delta^{N}(e^{j\omega})|^2}{N} \circledast C_{XX}(e^{j\omega}) + \frac{|\Delta^{N}(e^{j\omega})|^2}{N}\cdot|\mu_X|^2
\end{aligned}
$$

where

$$
\Delta^{N}(e^{j\omega}) = \sum_{n=0}^{N-1} e^{-j\omega n} = e^{-j\omega\frac{N-1}{2}} \cdot \frac{\sin(\frac{\omega N}{2})}{\sin(\frac{\omega}{2})}\,.
$$

The Periodogram

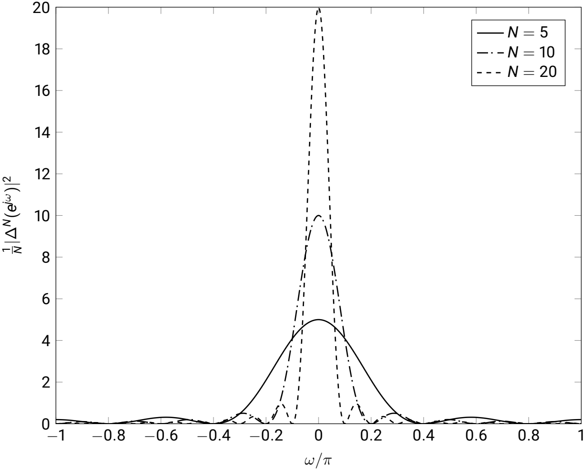
Thus, for a zero mean stationary process $X(n)$, $\mathrm{E}\left[I_{XX}^N(e^{j\omega})\right]$ is equal to $C_{XX}(e^{j\omega})$ convolved with the magnitude square of the Fourier transform of the rectangular window $|\Delta^N(e^{j\omega})|^2$.

- $\lim\limits_{N\to\infty} \mathrm{E}\left[I_{XX}^N(e^{j\omega})\right] = C_{XX}(e^{j\omega}) \quad \omega \neq 2k\pi$
- $\lim\limits_{N\to\infty} \mathrm{E}\left[I_{X-\mu_X,X-\mu_X}^N(e^{j\omega})\right] = C_{XX}(e^{j\omega}) \quad \forall\, \omega$

because

$$\lim_{N\to\infty} \frac{1}{N}|\Delta^N(e^{j\omega})|^2 = 2\pi \sum_k \delta(\omega + 2\pi k)$$

The Periodogram



The Periodogram

$\Rightarrow$ Advantages of using different window types are stated in the filter design theory part.

Mean of $X(n)$ multiplied by a window $w_N(n)$ of length N

$$
\begin{aligned}
\mathrm{E}\left[I^N_{WX,WX}(e^{j\omega})\right] &= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} w_N(n)w_N(n)^*\,\mathrm{E}\left[X(n)X(m)^*\right]e^{-j\omega(n-m)} \\
&= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} w_N(n)w_N(n)^* c_{XX}(n-m)e^{-j\omega(n-m)} \\
&\quad + \frac{1}{N}|W_N(e^{j\omega})|^2\mu_X\mu_X^* \\
&= \frac{1}{N}\int_{-\pi}^{\pi}|W_N(e^{j(\omega-\lambda)})|^2 C_{XX}(e^{j\lambda})\frac{d\lambda}{2\pi} + \frac{1}{N}|W_N(e^{j\omega})|^2|\mu_X|^2
\end{aligned}
$$

The Periodogram

Comparing this result with the one for rectangular windowed data, we observe the following:

If $C_{XX}(e^{j\alpha})$ has a significant peak for α in the neighborhood of ω, the expected value of $I_{XX}^{N}(e^{j\omega})$ and $I_{WX,WX}^{N}(e^{j\omega})$ can differ quite substantially from $C_{XX}(e^{j\omega})$. **The advantage of employing a taper is now apparent**. It can be considered as a shape to reduce the effect of neighboring peaks.

The Periodogram

Variance of the Periodogram

For simplicity, let us assume that $X(n)$ is a real-valued white Gaussian process with variance σ^2 and $\omega, \lambda \neq 0$, then

$$
\begin{aligned}
&\operatorname{Cov}\left[I_{XX}^N(e^{j\omega}), I_{XX}^N(e^{j\lambda})\right] \\
&\quad = \mathrm{E}\left[I_{XX}^N(e^{j\omega}) I_{XX}^N(e^{j\lambda})\right] - \mathrm{E}\left[I_{XX}^N(e^{j\omega})\right] \mathrm{E}\left[I_{XX}^N(e^{j\lambda})\right] \\
&\quad = \frac{1}{N^2}\left(|\Delta^N(e^{j(\omega+\lambda)})|^2 + |\Delta^N(e^{j(\omega-\lambda)})|^2\right)\sigma^4 .
\end{aligned}
$$

The Periodogram

If we don't assume a white and Gaussian process, we get [Brillinger 1981]

$$\begin{aligned}\mathrm{Cov}\left[I_{XX}^N(e^{j\omega}), I_{XX}^N(e^{j\lambda})\right] \approx \\ C_{XX}(e^{j\omega})^2 \cdot \frac{1}{N^2}\left(|\Delta^N(e^{j(\omega+\lambda)})|^2 + |\Delta^N(e^{j(\omega-\lambda)})|^2\right),\end{aligned}$$

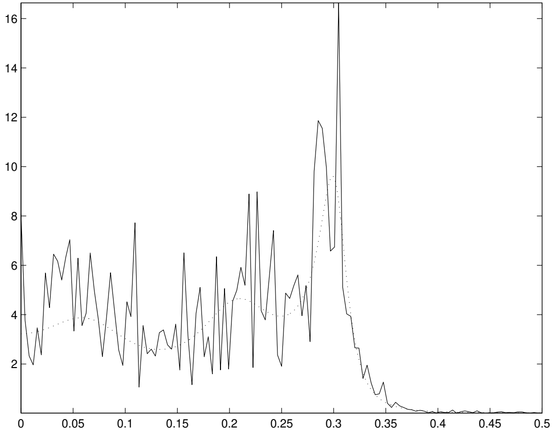
which reduces the variance of the periodogram to

$$\mathrm{Var}\left[I_{XX}^N(e^{j\omega})\right] \approx C_{XX}(e^{j\omega})^2 \cdot \frac{1}{N^2}\left(|\Delta^N(e^{j2\omega})|^2 + N^2\right).$$

This suggests that no matter how large N is, the variance of $I_{XX}^N(e^{j\omega}), \omega \neq \pi k$ will tend to remain at the level $C_{XX}(e^{j\omega})^2$.

- The periodogram is not a consistent estimator !!

The Periodogram



Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

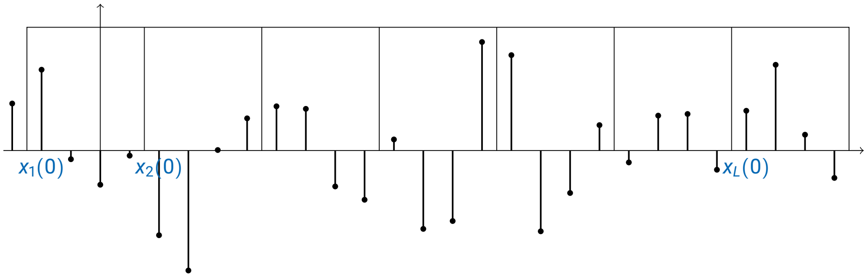
One way to reduce the variance of an estimator is to perfom multiple independent measurements and average the respective estimator.

If we assume that the covariance function $c_{XX}(k)$ of the real-valued process $X(n)$ is finite or can be viewed as finite with a small error and the length of $c_{XX}(k)$ is much smaller than the length of the data stretch N, then we can divide the measurement process $X(n)$ of length N into L parts with length M each.

Averaging Periodograms

With this assumption, we can calculate L asymptotic independent periodograms $I_{XX}^{M}(e^{j\omega}, l), l = 1, \ldots, L$ of sequences

$$X_l(n) = X(n + (l-1) \cdot M), \qquad n = 0, \ldots, M-1\,.$$



SPG

Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

The independent periodograms are written as

$$I_{XX}^{M}(e^{j\omega}, l) = \frac{1}{M} \left| \sum_{n=0}^{M-1} X_l(n) e^{-j\omega n} \right|^2 \qquad l = 1, \ldots, L$$

and the estimator as

$$\hat{C}_{XX}^{B}(e^{j\omega}) = \frac{1}{L} \sum_{l=1}^{L} I_{XX}^{M}(e^{j\omega}, l)\,.$$

This method of estimating the spectrum $C_{XX}(e^{j\omega})$ is called Bartlett's method because it is due to Bartlett in 1953.

Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

Mean of $\hat{C}^B_{XX}(e^{j\omega})$

$$
\begin{aligned}
\mathrm{E}\left[\hat{C}^B_{XX}(e^{j\omega})\right] &= \mathrm{E}\left[\frac{1}{L}\sum_{l=1}^{L} I^M_{XX}(e^{j\omega}, l)\right] \\
&= \frac{1}{L}\sum_{l=1}^{L} \mathrm{E}\left[I^M_{XX}(e^{j\omega}, l)\right] \\
&\approx \mathrm{E}\left[I^M_{XX}(e^{j\omega})\right]
\end{aligned}
$$

This is the same as the mean of the periodogram of a process $X(n)$, $n = 0, \ldots, M-1$. It can be seen that Bartlett's estimator is an unbiased estimator of the spectral density for $M \to \infty$ (That is when L is fixed and $N \to \infty$).

Averaging Periodograms

Variance of Bartlett's estimator

$$
\begin{aligned}
\mathrm{Var}\left[\hat{C}_{XX}^{B}(e^{j\omega})\right] &= \mathrm{Var}\left[\frac{1}{L}\sum_{l=1}^{L} I_{XX}^{M}(e^{j\omega}, l)\right] \\
&= \frac{1}{L^2}\sum_{l=1}^{L} \mathrm{Var}\left[I_{XX}^{M}(e^{j\omega}, l)\right] \\
&\approx \frac{1}{L}\mathrm{Var}\left[I_{XX}^{M}(e^{j\omega})\right]
\end{aligned}
$$

The variance of Bartlett's estimator is equivalent to the variance of the periodogram divided by the number of data segments L.

The variance decreases when L increases.

Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

Welch's method

Is indeed very similar to Bartlett's method except :

- the data segments are multiplied by a window $w_M(n)$ of length M and
- the segments may overlap.

For each periodogram l, $l = 1, \ldots, L$ the data sequence is now

$$X_l(n) = X(n + (l-1) \cdot D)\,,$$

where $(l-1) \cdot D$ is the starting point of the lth sequence.

Averaging Periodograms

Accordingly, the estimator is defined as

$$\hat{C}_{XX}^{W}(e^{j\omega}) = \frac{1}{L}\sum_{l=1}^{L} I_{WX,WX}^{M}(e^{j\omega}, l)$$

where A is a factor to get an asymptotic unbiased estimation and

$$I_{WX,WX}^{M}(e^{j\omega}, l) = \frac{1}{MA}\left|\sum_{n=0}^{M-1} w_M(n)\cdot X_l(n)\cdot e^{-j\omega n}\right|^2 .$$

For $A = 1$ and $w_M(n)$ as a rectangular window of length M and $D = M$, Welch's method is equivalent to Bartlett's method, so **Bartlett's method is a special case of Welch's method**.

Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

How should the factor A be chosen to get an asymptotic unbiased estimator?

With $D = 0$ it follows

$$
\begin{aligned}
\mathrm{E}\left[\hat{C}_{XX}^{W}(e^{j\omega})\right] &= \mathrm{E}\left[\frac{1}{L}\cdot\sum_{l=1}^{L} I_{WX,WX}^{M}(e^{j\omega},l)\right] \\
&= \frac{1}{L}\cdot\sum_{l=1}^{L}\mathrm{E}\left[I_{WX,WX}^{M}(e^{j\omega},l)\right] \\
&\approx \mathrm{E}\left[I_{WX,WX}^{M}(e^{j\omega})\right] \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{XX}(e^{j\lambda})\frac{1}{MA}|W_M(e^{j(\omega-\lambda)})|^2 d\,\lambda \\
&\quad +\frac{1}{MA}\cdot|W_M(e^{j\omega})|^2\cdot\mu_X^2\,.
\end{aligned}
$$

Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

Using Parseval's theorem, we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|W_M(e^{j\lambda})|^2 d\lambda = \sum_{n=0}^{M-1}|w_M(n)|^2 = MA$$

Assuming that M is large, i.e., the energy of $W_M(e^{j\omega})$ is concentrated around $\omega = 0$, we can state the approximation

$$\mathrm{E}\left[\hat{C}_{XX}^{W}(e^{j\omega})\right] \approx \frac{1}{MA}C_{XX}(e^{j\omega})\frac{1}{2\pi}\int_{-\pi}^{\pi}|W_M(e^{j\lambda})|^2 + \frac{1}{MA}|W_M(e^{j\omega})|^2\mu_X^2 \approx C_{XX}(e^{j\omega})\,.$$

we see that for $\omega \neq 0$ and $M \to \infty$ Welch's estimator is unbiased

Averaging Periodograms

For $M \to \infty$, the Variance of Welch's method for non-overlapping data segments with $D = M$ is asymptotically the same as that of Bartlett's method

$$\lim_{M\to\infty} \mathrm{Var}\left[\hat{C}^{W}_{XX}(e^{j\omega})\right] = \lim_{M\to\infty} \mathrm{Var}\left[\hat{C}^{B}_{XX}(e^{j\omega})\right] \approx \frac{1}{L} C_{XX}(e^{j\omega})^2 \,.$$

In the case of 50% overlap ($D = \frac{M}{2}$) Welch calculated the asymptotic variance of the estimator as

$$\lim_{M\to\infty} \mathrm{Var}\left[\hat{C}^{W}_{XX}(e^{j\omega})\right] \approx \frac{9}{8L} \cdot C_{XX}(e^{j\omega})^2 \,.$$

Averaging Periodograms

TECHNISCHE UNIVERSITÄT DARMSTADT

Remarks

The variance of Welch's estimator with 50% overlap is higher than the variance of Barlett's estimator, but the advantage of using Welch's estimator is that we can build more data segments L' from the same amount of data $X(n)$, $n = 0, \ldots, N-1$, or we can increase the length of the segments M to get a better bias behavior.

This advantage makes Welch's estimator one of the most used non-parametric spectrum estimators. This method is also implemented in the "pwelch" function of MATLAB.

Smoothing the Periodogram

TECHNISCHE UNIVERSITÄT DARMSTADT

- For $\omega_k \neq \omega_l$, $X_N(e^{j\omega_k})$ and $X_N(e^{j\omega_l})$ are asymptotically independent
- $I_{XX}^N(e^{j\omega_l})$ and $I_{XX}^N(e^{j\omega_k})$ are also independent
- This motivates the estimator

$$\hat{C}_{XX}^S(e^{j\omega}) = \frac{1}{2m+1} \sum_{k=-m}^{m} I_{XX}^N\left(e^{j2\pi(k(\omega)+k)/N}\right)$$

- $2\pi k(\omega)/N$ is the nearest frequency to ω, i.e. $2\pi k(\omega)/N \approx \omega$ for $N \to \infty$, at which we want to estimate the spectrum $C_{XX}(e^{j\omega})$

Smoothing the Periodogram

TECHNISCHE UNIVERSITÄT DARMSTADT

Mean of the Smoothed Periodogram

$$\mathrm{E}\left[\hat{C}_{XX}^{S}(e^{j\omega})\right] = \frac{1}{2m+1}\sum_{k=-m}^{m}\mathrm{E}\left[I_{XX}^{N}\left(e^{j\frac{2\pi}{N}(k(\omega)+k)}\right)\right]$$

with $\omega = 2\pi k(\omega)/N$, we obtain

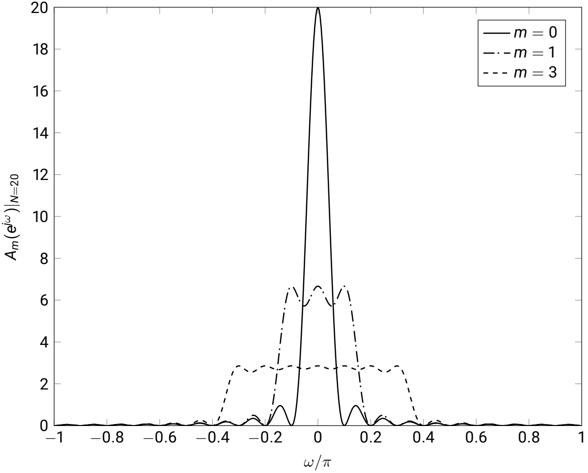
$$\begin{aligned}\mathrm{E}\left[\hat{C}_{XX}^{S}(e^{j\omega})\right] = \; & \frac{1}{2m+1}\sum_{k=-m}^{m}\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{N}|\Delta^{N}(e^{j(\omega+\frac{2\pi k}{N}-\lambda)})|^{2}C_{XX}\left(e^{j\lambda}\right)\,d\lambda \\ & +\frac{1}{N}|\Delta^{N}(e^{j\omega})|^{2}\mu_{X}^{2}\end{aligned}$$

Smoothing the Periodogram

$$
\begin{aligned}
\mathrm{E}\left[\hat{C}^{S}_{XX}(e^{j\omega})\right] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \underbrace{\frac{1}{2m+1}\cdot\sum_{k=-m}^{m}\frac{1}{N}|\Delta^{N}(e^{j(\omega+\frac{2\pi k}{N}-\lambda)})|^{2}}_{A_m(e^{j(\omega-\lambda)})} C_{XX}\left(e^{j\lambda}\right)\,d\lambda \\
&\quad +\frac{1}{N}|\Delta^{N}(e^{j\omega})|^{2}\mu_X^2 \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} A_m\left(e^{j(\omega-\lambda)}\right) C_{XX}\left(e^{j\lambda}\right)\,d\lambda + \frac{1}{N}|\Delta^{N}(e^{j\omega})|^{2}\mu_X^2\,.
\end{aligned}
$$

But how does this function $A_m\left(e^{j\omega}\right)$ behave?

Smoothing the Periodogram



Smoothing the Periodogram

TECHNISCHE UNIVERSITÄT DARMSTADT

We can describe $A_m(e^{j\omega})$ as having an approximately rectangular shape

- The bias of $\hat{C}^S_{XX}(e^{j\omega})$ will be greater as the bias of $I_{XX}(e^{j\omega})$, except when the spectrum $C_{XX}(e^{j\omega})$ is flat.

Asymptotically for $N \rightarrow \infty$, $\hat{C}^S_{XX}(e^{j\omega})$ is unbiased.

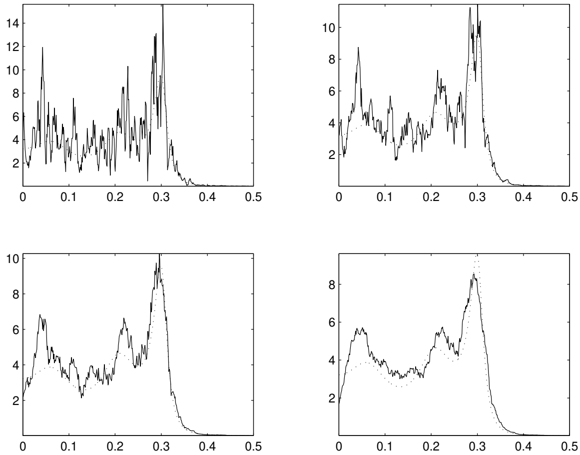
Smoothing the Periodogram

The calculation of the variance for $\hat{C}^S_{XX}(e^{j\omega})$ is similar to the calculation for the variance of Bartlett's estimator $\hat{C}^B_{XX}(e^{j\omega})$ and results in

$$\lim_{N\to\infty} \mathrm{Var}\left[\hat{C}^S_{XX}(e^{j\omega})\right] = \lim_{N\to\infty} \frac{1}{2m+1} \mathrm{Var}\left[I^N_{XX}(e^{j\omega})\right] = \frac{1}{2m+1} C_{XX}(e^{j\omega})^2$$

which is what we expected.

Smoothing the Periodogram



Smoothed periodograms with a moving average filter length of 5, 11, 21, 41.

A General Class of Spectral Estimators

TECHNISCHE UNIVERSITÄT DARMSTADT

The **smoothed periodogram** weighs all periodogram estimates equally.

- Reasonable when the spectrum is flat.
- Large Bias when the spectrum is not flat.

$\Rightarrow$ **It is more appropriate to weigh periodogram ordinates near ω more than those further away.**

A general class of estimators can be defined analogously to the smoothed periodogram as

$$\hat{C}_{XX}^{SW}(e^{j\omega}) = \frac{1}{(2m+1)A} \sum_{k=-m}^{m} W_k I_{XX}^{N}\left(e^{j2\pi(k(\omega)+k)/N}\right)$$

. With $A = \frac{1}{2m+1} \sum_{k=-m}^{m} W_k$

A General Class of Spectral Estimates

TECHNISCHE UNIVERSITÄT DARMSTADT

- For $\omega = 2\pi k(\omega)/N$, the mean of estimator $\hat{C}_{XX}^{SW}(e^{j\omega})$ is the same as for the smoothed periodogram
- But: Replacing the function $A_M(e^{j\omega})$ with

$$A_M^W(e^{j\omega}) = \frac{1}{(2m+1)A} \sum_{k=-m}^{m} W_k \frac{1}{N} |\Delta^N(e^{j(\omega+2\pi k/N+\lambda)})|^2$$

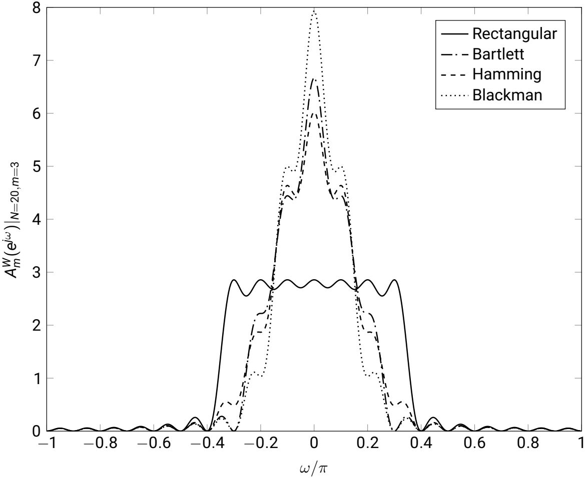
- If we set $W_k = 1$ for $k = -m, \ldots, m$, it follows that $A_m^W(e^{j\omega}) = A_m(e^{j\omega})$

$\Rightarrow$ **The smoothed periodogram is a special case of the general class of spectral estimators**

- The variance of $\hat{C}_{XX}^{SW}(e^{j\omega})$ is

$$\lim_{N\to\infty} \mathrm{Var}\left[\hat{C}_{XX}^{SW}(e^{j\omega})\right] = \frac{1}{(2m+1)^2 A^2} \left(\sum_{k=-m}^{m} W_k^2\right) C_{XX}(e^{j\omega})^2$$

A General Class of Spectral Estimates



The Log-Spectrum

TECHNISCHE UNIVERSITÄT DARMSTADT

- Estimating $10 \log C_{XX}(e^{j\omega})$ from $10 \log \hat{C}_{XX}(e^{j\omega})$ instead of the spectrum
- Log-Function is used in variance stabilization techniques
- Expectation and variance are given by

$$
\begin{aligned}
\mathrm{E}\left[10 \log \hat{C}_{XX}(e^{j\omega})\right] &\approx \quad 10 \log C_{XX}(e^{j\omega}) \\
\mathrm{Var}\left[10 \log \hat{C}_{XX}(e^{j\omega})\right] &\approx \quad \frac{(10 \log e)^2}{C}
\end{aligned}
$$

- C is a constant, which controls the variance ($C = L$ for Bartlett)
- Variance is constant and does not depend on $C_{XX}(e^{j\omega})$
- A good approach to constructing confidence intervals for $C_{XX}(e^{j\omega})$

Estimation of the Spectrum using the Sample Covariance

The covariance function $c_{XX}(\kappa)$ is defined as

$$c_{XX}(\kappa) = \mathrm{E}\left[(X(n+\kappa) - \mu_X)(X(n) - \mu_X)^*\right] .$$

Given $X(0), \ldots, X(N-1)$, we can estimate $c_{XX}(\kappa)$ by

$$\hat{c}_{XX}(\kappa) = \frac{1}{N} \sum_{n=0}^{N-1-\kappa} (X(n+\kappa) - \bar{X})(X(n) - \bar{X})^*, \quad 0 \leq \kappa \leq N-1$$

where $\bar{X} = \frac{1}{N} \sum\limits_{n=0}^{N-1} X(n)$ is an estimator of the mean $\mu_X = \mathrm{E}\left[X(n)\right]$.

To get $\hat{c}_{XX}(\kappa)$ for $\kappa < 0$ we use $\hat{c}_{XX}(\kappa) = \hat{c}_{XX}(-\kappa)^*$.

Estimation of the Spectrum using the Sample Covariance

TECHNISCHE UNIVERSITÄT DARMSTADT

Motivated by the fact that the spectrum $C_{XX}(e^{j\omega})$ is the Fourier transform of the covariance function $c_{XX}(\kappa)$, it is intuitive to suggest the estimator

$$\hat{C}_{XX}(e^{j\omega}) = \sum_{\kappa=-\infty}^{\infty} \hat{c}_{XX}(\kappa) e^{-j\omega\kappa}$$

Simple calculation (see tutorial) shows that this is equivalent to

$$\hat{C}_{XX}(e^{j\omega}) = I^N_{X-\bar{X},X-\bar{X}}(e^{j\omega}) \,.$$

This is not a consistent method for estimating the spectrum.

Estimation of the Spectrum using the Sample Covariance

TECHNISCHE UNIVERSITÄT DARMSTADT

The Blackman-Tukey Method

The basic idea of the method is to estimate the sample covariance function $\hat{c}_{XX}(\kappa)$ and window it with a function $w_{2M-1}(\kappa)$ of the following form

$$w_{2M-1}(\kappa) = \begin{cases} w_{2M-1}(-\kappa)\,, & \text{for} \quad |\kappa| \leq M-1 \quad \text{with} \quad M \ll N \\ 0\,, & \text{otherwise} \end{cases} .$$

$$\Rightarrow \quad \hat{C}_{XX}^{BT}(e^{j\omega}) = \sum_{\kappa=-M+1}^{M-1} \hat{c}_{XX}(\kappa) \cdot w_{2M-1}(\kappa) \cdot e^{-j\omega\kappa}$$

Estimation of the Spectrum using the Sample Covariance

Which is the same as

$$\hat{C}_{XX}^{BT}(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_{X-\bar{X},X-\bar{X}}^{N}(e^{j\lambda}) \cdot W_{2M-1}\left(e^{j(\omega-\lambda)}\right) d\,\lambda$$

which is the convolution of the Periodogram $I_{X-\bar{X},X-\bar{X}}^{N}(e^{j\omega})$ with the Fourier transform of the window $w_{2M-1}(n)$. From this, we get another constraint for the window

$$W_{2M-1}(e^{j\omega}) \geq 0 \qquad \forall\omega\,,$$

to ensure the estimate $\hat{C}_{XX}^{BT}(e^{j\omega})$ is non-negative for all ω.

Estimation of the Spectrum using the Sample Covariance

The mean of the estimator $\hat{C}^{BT}_{XX}(e^{j\omega})$ can be calculated as

$$\begin{aligned}
\mathrm{E}\left[\hat{C}^{BT}_{XX}(e^{j\omega})\right] &= \mathrm{E}\left[I^N_{X-\bar{X},X-\bar{X}}(e^{j\omega})\right] \circledast W_{2M-1}(e^{j\omega}) \\
&= \frac{1}{N}|\Delta^N(e^{j\omega})|^2 \circledast W_{2M-1}(e^{j\omega}) \circledast C_{XX}(e^{j\omega}) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} C_{XX}(e^{j\theta}) \\
&\quad \times \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{N}|\Delta^N(e^{j\vartheta})|^2 \cdot W_{2M-1}(e^{j(\omega-\theta-\vartheta)})d\,\vartheta d\,\theta\,.
\end{aligned}$$

Estimation of the Spectrum using the Sample Covariance

Since $\lim_{N\to\infty} \frac{1}{N} \left|\Delta^N(e^{j\omega})\right|^2 = 2\pi \sum_k \delta(\omega + 2\pi k)$, we find

$$\lim_{N\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{N} |\Delta^N(e^{j\vartheta})|^2 \cdot W_{2M-1}(e^{j(\omega-\theta-\vartheta)}) d\,\vartheta = W_{2M-1}(e^{j(\omega-\theta)})$$

which reduces the term for the mean of the estimator to a simple convolution

$$\mathrm{E}\left[\hat{C}_{XX}^{BT}(e^{j\omega})\right] = \frac{1}{2\pi} \int_{-\pi}^{\pi} C_{XX}(e^{j\theta}) \cdot W_{2M-1}\left(e^{j(\omega-\theta)}\right) d\,\theta\,.$$

Therefore the Blackman-Tukey estimator is asymptotically unbiased for $M \to \infty$.

Estimation of the Spectrum using the Sample Covariance

The variance of the estimator is given by [Proakis, Manolakis 1989] as

$$\begin{aligned} \mathrm{Var}\left[\hat{C}_{XX}^{BT}(e^{j\omega})\right] &\approx C_{XX}(e^{j\omega})^2 \cdot \frac{1}{N} \int_{-\pi}^{\pi} W_{2M-1}(e^{j\omega})^2 d\,\omega \\ &\approx C_{XX}(e^{j\omega})^2 \cdot \frac{1}{N} \sum_{\kappa=-M+1}^{M-1} w_{2M-1}(\kappa)^2 \,. \end{aligned}$$

Based on this approximation, the variance of the Blackman Tukey method tends to zero if the window function $w_{2M-1}(\kappa)$ has finite energy.

Cross-Spectrum Estimation

If $X(n)$ and $Y(n)$ are jointly stationary and real-valued processes, the cross-covariance function is defined as

$$
\begin{aligned}
c_{YX}(\kappa) &= \mathrm{E}\left[(Y(n+\kappa) - \mathrm{E}\left[Y(n)\right])(X(n) - \mathrm{E}\left[X(n)\right])\right] \\
c_{YX}(\kappa) &= \mathrm{E}\left[Y(n+\kappa)X(n)\right] - \mathrm{E}\left[Y(n)\right]\mathrm{E}\left[X(n)\right]
\end{aligned}
$$

The cross-spectrum is defined by

$$
C_{YX}(e^{j\omega}) = \sum_{\kappa=-\infty}^{\infty} c_{YX}(\kappa) e^{-j\omega\kappa} = C_{XY}(e^{j\omega})^{*} = C_{XY}(e^{-j\omega})\,.
$$

Cross-Spectrum Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

Given $X(0), \ldots, X(N-1), Y(0), \ldots, Y(N-1)$, the objective is to estimate the cross-spectrum $C_{YX}(e^{j\omega})$ for $-\pi < \omega < \pi$. The cross-periodogram is defined as

$$I_{YX}^{N}(e^{j\omega}) = \frac{1}{N} Y_N(e^{j\omega}) X_N(e^{j\omega})^*$$

where the Finite Fourier Transform of $X(n)$ and $Y(n)$ are given by

$$X_N(e^{j\omega}) = \sum_{n=0}^{N-1} X(n) e^{-j\omega n} \quad \text{and} \quad Y_N(e^{j\omega}) = \sum_{n=0}^{N-1} Y(n) e^{-j\omega n}$$

As for the periodogram, the expectation and the variance of the cross-periodogram are given by

$$\lim_{N\to\infty} \mathrm{E}\left[I_{YX}^{N}(e^{j\omega})\right] = C_{YX}(e^{j\omega}) \quad \text{and} \quad \lim_{N\to\infty} \mathrm{Var}\left[I_{YX}^{N}(e^{j\omega})\right] = C_{YY}(e^{j\omega}) C_{XX}(e^{j\omega})$$

Cross-Spectrum Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Because of the identity $|C_{YX}(e^{j\omega})|^2 \leq C_{YY}(e^{j\omega})C_{XX}(e^{j\omega})$, the variance of $\hat{C}^B_{YX}(e^{j\omega})$ is never smaller than $|C_{YX}(e^{j\omega})|^2$ for all N
- To construct estimators with lower variance, we have to use the independence properties between segments

$$\hat{C}^B_{YX}(e^{j\omega}) = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{M}Y_M(e^{j\omega},l)X_M(e^{j\omega},l)^*$$

- The finite Fourier transforms of segments $l = 1, \dots, L$ are given by

$$Y_M(e^{j\omega},l) = \sum_{n=0}^{M-1} Y(n+(l-1)M)e^{-j\omega n}$$

$$X_M(e^{j\omega},l) = \sum_{n=0}^{M-1} X(n+(l-1)M)e^{-j\omega n}$$

Cross-Spectrum Estimation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Asymptotically for large segment length M (that is when L is fixed and $N \to \infty$), we have

$$
\begin{aligned}
\mathrm{E}\left[\hat{C}_{YX}(e^{j\omega})\right] &\approx C_{YX}(e^{j\omega}) \\
\mathrm{Var}\left[\hat{C}_{YX}(e^{j\omega})\right] &\approx \frac{1}{L} C_{YY}(e^{j\omega}) C_{XX}(e^{j\omega})
\end{aligned}
$$

- Smoothing of neighboring frequencies is an alternative and leads to similar results
- Take care with smoothing when one is interested in the phase!

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

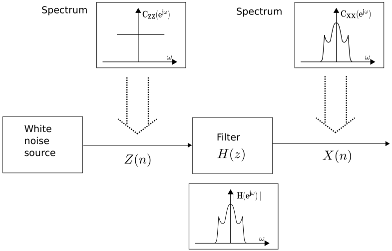
1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. **Parametric Spectrum Estimation**
   - Autoregressive Process
   - Moving Average Process
   - Model Order Selection
8. Application: Airplane Tracking using Kalman Filter
9. Discrete Kalman Filter

Parametric Spectrum Estimation Introduction

- Non-parametric spectrum estimation does not incorporate *a priori* knowledge of the spectrum
- Using this information may lead to more accurate and higher resolution estimates of the spectrum
- Parametric spectrum estimation consists of the following steps:
  1. Select an appropriate model of the random process
  2. Estimate the model parameters from the observations
  3. Incorporate the estimated parameters into the spectrum model

Parametric Spectrum Estimation Introduction

TECHNISCHE UNIVERSITÄT DARMSTADT



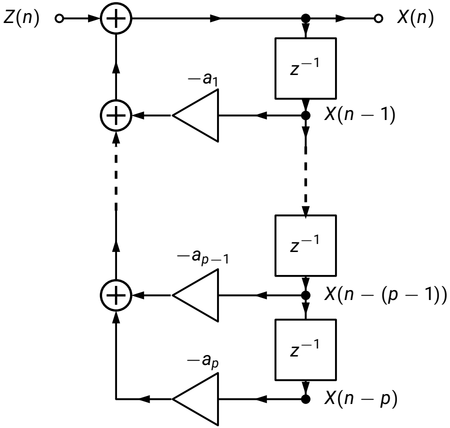
Signal generation model of parametric spectral analysis

Main idea of parametric spectrum estimation:

- White noise as input process (unobservable)
- Filter shapes the spectrum of the signal
- Estimate the coefficients of the filter and the power of the noise to get an estimate of the spectrum

Auto-regressive Process

Consider the following system with output $X(n)$



The difference equation and the spectrum are given by:

$$X(n)+\sum_{k=1}^{p} a_k X(n-k) = Z(n)\,.$$

$$\Rightarrow \quad H(e^{j\omega}) = \frac{1}{1+\sum_{k=1}^{p} a_k e^{-j\omega k}}$$

Auto-regressive Process

TECHNISCHE UNIVERSITÄT DARMSTADT

Because we have assumed the filter to be stable, the roots of the characteristic equation

$$z^p + \sum_{k=1}^{p} a_k z^{p-k} = 0, \quad z = e^{j\omega}$$

lie in the unit disc, i.e. $|z_l| < 1$ for all p roots of the above equation.

Auto-regressive Process

The spectrum is given as

$$C_{XX}(e^{j\omega}) = |H(e^{j\omega})|^2 C_{ZZ}(e^{j\omega})$$

and thus,

$$C_{XX}(e^{j\omega}) = \frac{\sigma_Z^2}{|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}|^2}$$

The objective is to estimate $C_{XX}(e^{j\omega})$ based on observations of the process $X(n)$ for $n = 0, \ldots, N-1$.

Auto-regressive Process

We will devise a method to estimate the parameters $a_1, \dots, a_p$, which will be used to estimate the spectrum $C_{XX}(e^{j\omega})$.
Multiplying

$$X(n) + \sum_{k=1}^{p} a_k X(n-k) = Z(n)$$

by $X(n-l)$ (from the right), for $l \geq 0$ and taking expectation, leads to

$$\mathrm{E}\left[X(n)X(n-l)\right] + \sum_{k=1}^{p} a_k \mathrm{E}\left[X(n-k)X(n-l)\right] = \mathrm{E}\left[Z(n)X(n-l)\right], \quad l \geq 0$$

$$c_{XX}(l) + \sum_{k=1}^{p} a_k \mathrm{E}\left[X(n-k)X(n-l)\right] = \mathrm{E}\left[Z(n)X(n-l)\right], \quad l \geq 0$$

Auto-regressive Process

Assuming the recursive filter to be causal, i.e. $h(n) = 0$ for $n < 0$ leads to the relationship

$$\begin{aligned} X(n) &= \sum_{k=0}^{\infty} h(k)Z(n-k) \\ &= h(0)Z(n) + \sum_{k=1}^{\infty} h(k)Z(n-k) \end{aligned}$$

Thus, shifting $X(n)$ by $l \geq 0$ gives

$$X(n-l) = h(0)Z(n-l) + \sum_{k=1}^{\infty} h(k)Z(n-l-k), \quad l \geq 0$$

Auto-regressive Process

Taking the expected value of the product $Z(n) \cdot X(n-l)$ leads to

$$\begin{aligned} \mathrm{E}\left[Z(n)X(n-l)\right] &= h(0)c_{ZZ}(l) + \underbrace{\sum_{k=1}^{\infty} h(k)c_{ZZ}(l+k)}_{=0}, \quad l \geq 0 \\ \mathrm{E}\left[Z(n)X(n-l)\right] &= h(0)c_{ZZ}(l) = h(0)\sigma_Z^2\delta(l) \end{aligned}$$

Assuming $h(0) = 1$, leads to the so called *Yule-Walker equations*

$$c_{XX}(l) + \sum_{k=1}^{p} a_k c_{XX}(l-k) = \begin{cases} \sigma_Z^2, & l = 0 \\ 0, & l = 1, \ldots, p \end{cases}$$

Auto-regressive Process

TECHNISCHE UNIVERSITÄT DARMSTADT

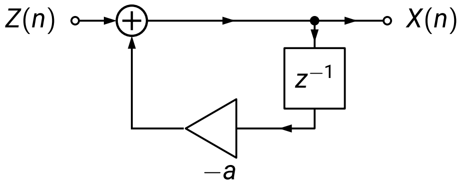
Using $\hat{c}_{XX}(0), \ldots, \hat{c}_{XX}(p)$, one would get a unique solution for $a_1, \ldots, a_p$. It is unique if the inverse matrix with the elements $c_{XX}(l-k)$ exists which is guaranteed if the filter is stable. Having observed $X(n)$, one can estimate $c_{XX}(n)$. Given estimates $\hat{a}_1, \ldots, \hat{a}_p$ of $a_1, \ldots, a_p$, one can use the Yule-Walker equations for $l = 0$ to get $\hat{\sigma}_Z^2$. Then it is easy to find an estimate for $C_{XX}(e^{j\omega})$ through

$$\hat{C}_{XX}(e^{j\omega}) = \frac{\hat{\sigma}_Z^2}{|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j\omega k}|^2}$$

Auto-regressive Process

TECHNISCHE UNIVERSITÄT DARMSTADT

Example:



- The difference equation of the system is given by

$$X(n) + a \cdot X(n-1) = Z(n)$$

- Multiplying by $X(n-l)$ from the right for $l \geq 0$ and taking the expectation leads to

$$c_{XX}(l) + ac_{XX}(l-1) = c_{ZX}(l), \quad l \geq 0$$

Auto-regressive Process

TECHNISCHE UNIVERSITÄT DARMSTADT

- With $c_{ZX}(l) = h(0)c_{ZZ}(l)$ and $h(0) = 1$, we get

$$c_{XX}(0) + ac_{XX}(1) = \sigma_Z^2$$
$$c_{XX}(1) + ac_{XX}(0) = 0$$

- Solving the equation yields to

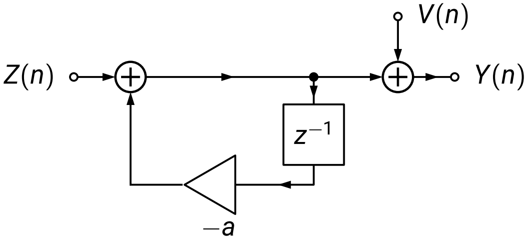
$$a = -\frac{c_{XX}(1)}{c_{XX}(0)} \quad \text{and} \quad \sigma_Z^2 = c_{XX}(0) - \frac{c_{XX}(1)^2}{c_{XX}(0)}$$

- By replacing the parameters a and σ_Z^2 by their estimates, we get the estimator of the spectrum:

$$\hat{C}_{XX}(e^{j\omega}) = \frac{\hat{\sigma}_Z^2}{|1 + \hat{a}e^{-j\omega}|^2}$$

Auto-regressive Process plus Noise

Example:



- $Z(n)$ and $V(n)$ are uncorrelated white noise processes with variances σ_Z^2 and σ_V^2, respectively.
- The difference equation of the system is given by

$$Y(n) + aY(n-1) = Z(n) + V(n) + aV(n-1)$$

Auto-regressive Process plus Noise

- Multiplying by $Y(n-l)$ from the right for $l \geq 0$ and taking the expectation leads to

$$c_{YY}(l) + a c_{YY}(l-1) = c_{ZY}(l) + c_{VY}(l) + a c_{VY}(l-1)$$

- Because $l \geq 0$ the cross-covariance function $c_{ZY}(l)$ can be found to be

$$c_{ZY}(l) = c_{ZX}(l) = h(0) c_{ZZ}(l).$$

- The cross-covariance function $c_{VY}(l)$ is given by

$$c_{VY}(l) = \mathrm{E}[V(n)Y(n-l)] = \mathrm{E}[V(n)\sum_{m=0}^{\infty} h(m)Z(n-l-m) + V(n-l)] = c_{VV}(l)$$

Auto-regressive Process plus Noise

TECHNISCHE UNIVERSITÄT DARMSTADT

- Finally, assuming $h(0) = 1$, we get an extended set of Yule-Walker equations

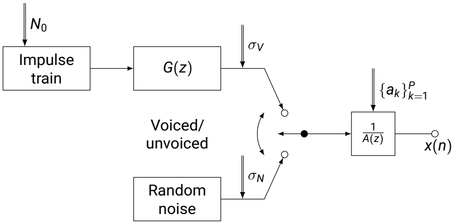
$$c_{YY}(l) + a c_{YY}(l-1) = \begin{cases} \sigma_Z^2 + \sigma_V^2, & l = 0 \\ a\sigma_V^2, & l = 1 \\ 0, & l > 1. \end{cases}$$

- The general form (for $p > l$) of the extended Yule-Walker equations is given by

$$c_{YY}(l) + \sum_{k=1}^{p} a_k c_{YY}(l-k) = \begin{cases} \sigma_Z^2 + \sigma_V^2, & l = 0 \\ a_l\sigma_V^2, & l = 1, \ldots, p \\ 0, & l > p. \end{cases}$$

Auto-regressive Process Application

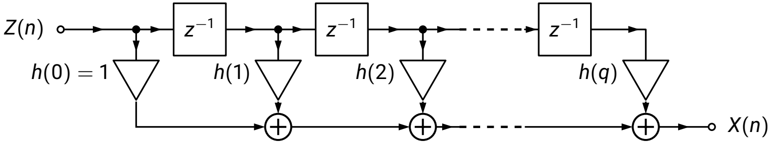
Application: Speech Error Coding



Signal generation model of a speech signal

- Different models for voiced and unvoiced speech is used
- Speech is modeled as AR process with polynomial $A(z)$
- Inverting this generation model, leads to a system which predicts the signals
- Prediction Coding is used in practice and only the error of the predicted signal and the observed signal is transmitted/stored

Moving Average Process



Consider a transversal filter $h(n)$ with input real-valued $Z(n)$ and real-valued output $X(n)$:

$$X(n) = \sum_{k=0}^{q} h(k) Z(n-k) \,.$$

If $Z(n)$ is a white noise process, i.e., $\mathrm{E}[Z(n)] = 0$ and $c_{ZZ}(l) = \sigma_Z^2 \delta(l)$, $X(n)$ is a

Moving Average (MA) process.

with

$$C_{XX}(e^{j\omega}) = \left|H(e^{j\omega})\right|^2 C_{ZZ}(e^{j\omega}) = \left|H(e^{j\omega})\right|^2 \sigma_Z^2$$

Moving Average Process

TECHNISCHE UNIVERSITÄT DARMSTADT

- The covariance function $c_{XX}(l) = \mathrm{E}\left[X(n+l)X(n)\right]$ is given by

$$c_{XX}(l) = \mathrm{E}\left[\left(\sum_{k=0}^{q} h(k)Z(n+l-k)\right)\left(\sum_{m=0}^{q} h(m)Z(n-m)\right)\right]$$
$$= \begin{cases} \sigma_Z^2 \sum_{m=0}^{q-|l|} h(m)h(m+|l|), & 0 \leq |l| \leq q \\ 0, & \text{otherwise} \end{cases}$$

- If we have estimates of $c_{XX}(0), \ldots, c_{XX}(q)$, we have to solve a non-linear system of equations and thus not find a unique solution
- For $b(m) = \sigma_Z h(m)$ with $m = 0, \ldots, q$ and $h(0) = 1$, we get

$$c_{XX}(l) = \sum_{m=0}^{q-|l|} b(m)b(m+|l|), \quad \text{for} \quad 0 \leq |l| \leq q$$

Moving Average Process

- We can find the parameters of a moving average process by applying a spectral factorization
- The Fourier transform of the unit sample response $b(m)$ is

$$B(e^{j\omega}) = \sum_{m=0}^{q} b(m) e^{-j\omega m}$$

- We know that the relationship between the spectra of input and output signals can be expressed as

$$C_{XX}(e^{j\omega}) = \sum_{l=-q}^{q} c_{XX}(l) e^{-j\omega l} = B(e^{j\omega}) B(e^{-j\omega})$$

Moving Average Process

- By substituting $e^{j\omega} = z$, we obtain $C_{XX}(z) = B(z)B(z^{-1})$, where $B(z)$ is the z-transform of $b(m)$. We then:
  1. Calculate $P(z) = z^q C_{XX}(z)$
  2. Assign all roots z_i of $P(z)$ inside the unit circle to $B(z)$

     $B(z) = b(0)\prod_{i=1}^{q}(1 - z_i z^{-1})$ with $b(0) > 0$ and $|z_i| \leq 1$
  3. Calculate $h(m) = b(m)/b(0)$ for $m = 1, \ldots, q$.
- To get the coefficients $b(m)$, we can apply the method of equating coefficients

$$b(0) + b(1)z^{-1} + \ldots + b(q)z^{-q} = \sigma_Z \left(1 + h(1)z^{-1} + \ldots + h(q)z^{-q}\right)$$

and we obtain

$$\begin{aligned} b(0) &= \sigma_Z \\ b(1) &= \sigma_Z h(1) \Longleftrightarrow h(1) = b(1)/b(0) \\ &\vdots \end{aligned}$$

Moving Average Process

- For the power σ_Z^2 of $Z(n)$, we can solve

$$c_{XX}(l) = \sum_{m=0}^{q-l} b(m)b(m+l)$$

for $l = 0$ and get
$$b(0)^2 = \sigma_Z^2 = \frac{c_{XX}(0)}{1 + \sum_{m=1}^{q} h(m)^2}$$

- By replacing $c_{XX}(l)$ by its estimate $\hat{c}_{XX}(l)$, we obtain the estimates of the parameters of the MA process
- The spectrum can be obtained via the relationship

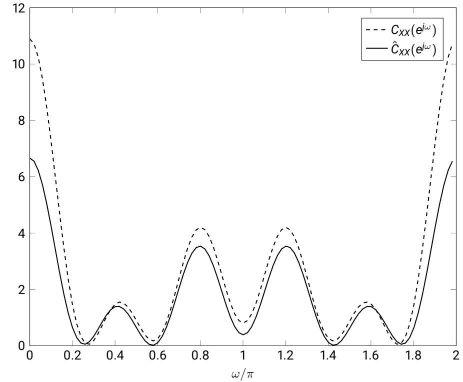
$$\hat{C}_{XX}(e^{j\omega}) = \left| \sum_{n=0}^{q} \hat{b}(n) e^{-j\omega n} \right|^2 .$$

Moving Average Process

Example:
Consider a receiver input signal $X(n)$ from a wireless transmission. To find the estimate of the spectrum $C_{XX}(e^{j\omega})$, we first calculate the sample covariance $\hat{c}_{XX}(n)$ and take the z-transform $\hat{C}_{XX}(z)$. By taking $P(z) = z^{N-1}\hat{C}_{XX}(z)$ we find $N-1$ roots z_i, $i = 1, \ldots, N-1$ inside the unit circle and assign them to $\hat{B}(z)$. The estimator $\hat{C}_{XX}(e^{j\omega})$ can then be found according to

$$\hat{C}_{XX}(e^{j\omega}) = \left|\sum_{n=0}^{q} \hat{b}(n)e^{-j\omega n}\right|^2 .$$

Moving Average Process



Estimate of an MA process with $N = 100$ and $q = 5$

Auto-regressive Moving Average Process

TECHNISCHE UNIVERSITÄT DARMSTADT

- Both, Auto-regressive and Moving Average Processes can be combined to an Auto-regressive Moving Average (ARMA) Process
- The difference equation of an ARMA process is given by

$$X(n) + \sum_{k=1}^{p} a_k X(n-k) = Z(n) + \sum_{k=1}^{q} b_k Z(n-k)$$

- The process is called ARMA(p,q)
- Modified Yule-Walker equations can be obtained similar to AR and MA processes

Parametric Spectrum Estimation Models

TECHNISCHE UNIVERSITÄT DARMSTADT

- AR, MA and ARMA modeling may not always be adequate.
- A priori information can be used to model the observed process.
- An example could be

$$X(n) = A\cos(\omega_0 n + \phi), \quad \phi \sim \mathcal{U}(-\pi, \pi)$$

for which the true spectrum is

$$C_{XX}(e^{j\omega}) = \frac{A^2}{2}\pi\left[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\right], \quad -\pi < \omega \leq \pi.$$

One can estimate A and ω_0, then $\hat{C}_{XX}(e^{j\omega})$ is obtained.

Parametric Spectrum Estimation Models

TECHNISCHE UNIVERSITÄT DARMSTADT

- Alternatively, one may model the spectrum itself, e.g.

$$C_{XX}(e^{j\omega}) = \sum_{i=1}^{P} A_i p(\omega; \omega_i, \sigma_i),$$

  where, e.g. could be $p(\omega; \omega_i, \sigma_i) = e^{-\frac{2}{\sigma_i}(\omega-\omega_i)^2}$
- The spectrum is observed directly
- The method of non-linear least-squares leads to

$$\hat{C}_{XX}(e^{j\omega}) = \sum_{i=1}^{P} \hat{A}_i p(\omega; \hat{\omega}_i, \hat{\sigma}_i),$$

  obtained from fitting the model to the observed spectrum.

Model order selection

TECHNISCHE UNIVERSITÄT DARMSTADT

- **Problem:** How to estimate the order of the system?
- A parsimonious model of the process is desirable
- Variance of the input process always decreases when the order increases
- There is no meaningful minimum order
- Criteria must be found to estimate the order and avoid overfitting
- A penalty term is added to the variance

Model order selection

TECHNISCHE UNIVERSITÄT DARMSTADT

Several techniques exists:

- **Akaike's information criterion (AIC)**
  Minimize the following function with respect to m.

$$\mathrm{AIC}(m) = \log \hat{\sigma}^2_{Z,m} + m\frac{2}{N}, \quad m = 1, \ldots, M$$

- More accurate: **minimum description length (MDL)**
  Take the smallest m that minimizes

$$\mathrm{MDL}(m) = \log \hat{\sigma}^2_{Z,m} + m\frac{\log N}{N}, \quad m = 1, \ldots, M\,.$$

Model order selection

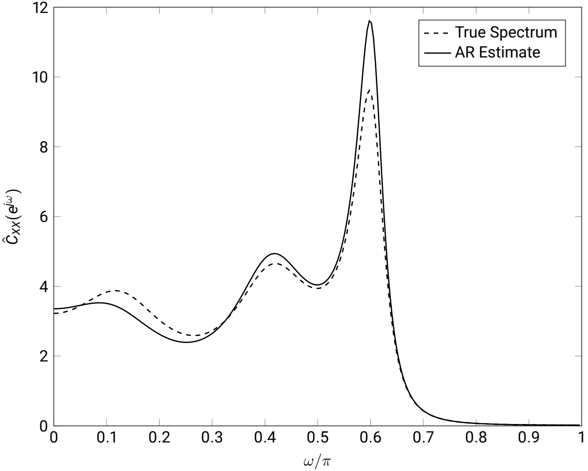
TECHNISCHE UNIVERSITÄT DARMSTADT

Example:
We generate realizations from the $\mathcal{N}(0,1)$ distribution. We filter the data recursively to get 1024 values. The filter has poles

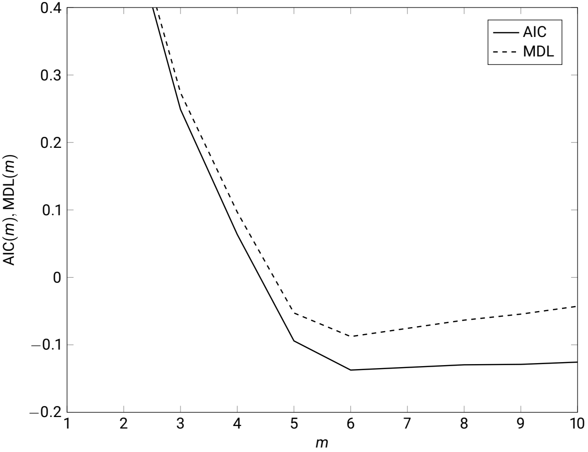
$$z_{1,2} = 0.7e^{\pm j0.4}, z_{3,4} = 0.75e^{\pm j1.3}, z_{5,6} = 0.9e^{\pm j1.9}$$

It is a recursive filter of order 6. The process is an AR(6) process.

Model order selection



Model order selection



Outline

1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation
8. **Application: Airplane Tracking using Kalman Filter**
9. Discrete Kalman Filter

Airplane Tracking using Kalman Filter

Motivation

- Huge increase in air traffic in recent years has led to crowded airspaces.
- Deutsche Flugsicherung (DFS) provides information about air traffic and air lanes and is responsible for aerial surveillance.
- Collision avoidance of airplanes is an essential task for the operators at the ground.
- Information about the position, velocity and acceleration of an airplane is required.

Problem

- Airplanes are tracked using a radar station which measures the distance X between the radar station and an airplane (Time-delay estimation).
- Knowing the distance at time instant n_0, operator is interested to predict position, velocity and acceleration of the airplane.
- Kalman Filter may be used for this task where we have three different state variables:
  1. distance, $X_1(n)$
  2. velocity, $X_2(n)$
  3. acceleration, $X_3(n)$

State Space Model

TECHNISCHE UNIVERSITÄT DARMSTADT

- Three spatial dimensions (x, y, z) lead to a state vector X with nine dimensions, but for simplicity only one spatial dimension is considered here.

We have the following relationships with T being the sample period:

$$
\begin{aligned}
X_2(n+1) &= \frac{1}{T}(X_1(n+1) - X_1(n)) \\
X_3(n+1) &= \frac{1}{T}(X_2(n+1) - X_2(n)),
\end{aligned}
$$

- Acceleration of the airplane equals zero and is only influenced by gusts of wind, i.e. $X_3(n+1) = U(n)$, where $U(n)$ is a zero-mean, stationary random process with variance σ_U^2.

State Space Model

After some simple algebra we obtain the system matrices

$$\mathbf{A} = \begin{pmatrix} 1 & T & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} T^2 \\ T \\ 1 \end{pmatrix}.$$

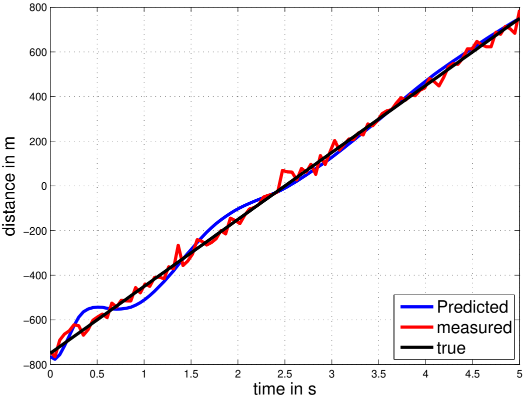
Since we only measure the distance to the airplane, we have the measured signal

$$Z(n) = \mathbf{H}\mathbf{X}(n) + V(n),$$

with $\mathbf{H} = (1,\ 0,\ 0)$ and $V(n)$ being a white Gaussian noise process.

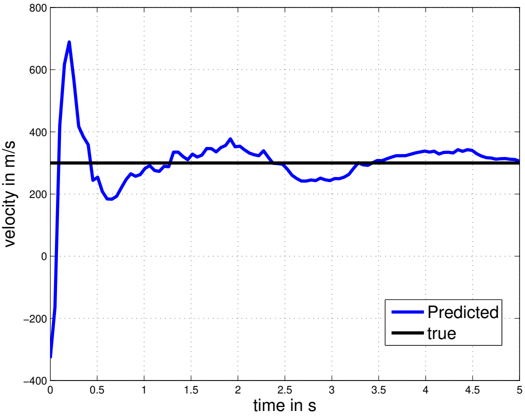
- Using the Kalman algorithm we are able to predict the distance and the velocity of the airplane.

Prediction of the position



Prediction of the velocity

TECHNISCHE UNIVERSITÄT DARMSTADT



Conclusions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Kalman Filter is a useful tool to track the position, velocity and acceleration of an airplane.
- Velocity and acceleration are not measurable but still can be estimated using Kalman Filter.
- Many applications of Kalman filter appear in biomedical signal processing, navigation systems, communication technology and others.

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Random Variables and Stochastic Processes
2. Brief Introduction to Elements of Estimation Theory
3. The Finite Fourier Transform
4. Introduction to Digital Spectral Analysis
5. Application: Spatial Spectra Estimation - Direction-Finding
6. Non-Parametric Spectrum Estimation
7. Parametric Spectrum Estimation
8. Application: Airplane Tracking using Kalman Filter
9. **Discrete Kalman Filter**
   - State Space Model
   - Kalman Filter

State Space Model

TECHNISCHE UNIVERSITÄT DARMSTADT

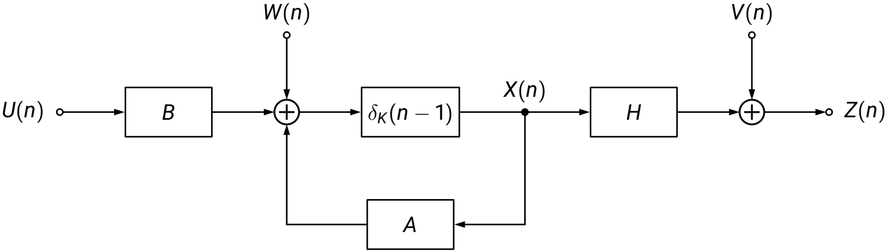
The state space model is the base of the Kalman filter!

- System equation: $X(n) = A \cdot X(n-1) + B \cdot U(n-1) + W(n-1)$
- Measurement equation: $Z(n) = H \cdot X(n) + V(n)$

Assumptions

- $W(n)$, $V(n)$ are independent, white and Gaussian with $W(n) \sim \mathcal{N}(0, Q)$ and $V(n) \sim \mathcal{N}(0, R)$
- A priori known : A, B, H, $U(n)$, Q, R

State Space Model



Kalman Filter

Algorithm seeks to estimate the unknown state $X(n)$

Based on two steps

- **Prediction:** Using a priori knowledge predict

$$\begin{aligned}\hat{X}^{-}(n) &= A \cdot \hat{X}(n-1) + B \cdot U(n-1) \\ \hat{Z}^{-}(n) &= H \cdot \hat{X}^{-}(n)\end{aligned}$$

- **Correction:** Correct the estimate based on the measurement

$$\hat{X}(n) = \hat{X}^{-}(n) + K(n)\left[Z(n) - \hat{Z}^{-}(n)\right]$$

But how to choose the Kalman gain $K(n)$?

Kalman Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

Choosing the Kalman gain $K(n)$ according to the MMSE-Criterion

$$\min_{K(n)} \mathrm{E}\left[\|X(n)-\hat{X}(n)\|^2\right]$$

which results in the following equations:

$$\begin{aligned} P^-(n) &= AP(n-1)A^T + Q \\ K(n) &= P^-(n)H^T(HP^-(n)H+R)^{-1} \\ P(n) &= (I-K(n)H)P^-(n) \end{aligned}$$

where $P(n) = \mathrm{E}\left[\left(X(n)-\hat{X}(n)\right)\left(X(n)-\hat{X}(n)\right)^T\right]$
and $P^-(n)$ its a priori estimate.

Kalman Filter

After initialising, using for each time index n the following algorithm:

1. *Predict* the current state and covariance using the previous estimate.
   1.1 $\hat{X}^{-}(n) = A\hat{X}(n-1) + BU(n-1)$
   1.2 $P^{-}(n) = AP(n-1)A^{T} + Q$
2. *Correct* the predictions using the current measurement.
   2.1 $K(n) = P^{-}(n)H^{T}(HP^{-}(n)H + R)^{-1}$
   2.2 $\hat{X}(n) = \hat{X}^{-}(n) + K(n)(Z(n) - \hat{Z}^{-}(n))$
   2.3 $P(n) = (I - K(n)H)P^{-}(n)$

Kalman Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

Example: Measuring a constant with additive noise

State space equations:

$$\begin{aligned} X(n) &= AX(n-1) + BU(n-1) + W(n-1) \\ \Rightarrow X(n) &= X(n-1) + W(n-1) \end{aligned}$$

$$\begin{aligned} Z(n) &= HX(n) + V(n) \\ \Rightarrow Z(n) &= X(n) + V(n) \end{aligned}$$

$$\Rightarrow A = 1 \quad B = 0 \quad H = 1$$

Kalman Filter

TECHNISCHE UNIVERSITÄT DARMSTADT

Kalman algorithm:

$$
\begin{array}{lcll}
\hat{X}^-(n) & = & A\hat{X}(n-1) + BU(n-1) & = \quad \hat{X}(n-1) \\
P^-(n) & = & AP(n-1)A^T + Q & = \quad P(n-1) + Q \\
\hline
K(n) & = & P^-(n)H^T(HP^-(n)H + R)^{-1} & = \quad P^-(n)(P^-(n) + R)^{-1} \\
\hat{X}(n) & = & \hat{X}^-(n) + K(n)(Z(n) - \hat{Z}^-(n)) & \\
 & = & \hat{X}^-(n) + K(n)(Z(n) - \hat{X}^-(n)) & \\
P(n) & = & (I - K(n)H)P^-(n) & = \quad (1 - K(n)H)P^-(n)
\end{array}
$$

Using the following values: $Q = 10^{-5}$, $R = 0.5$, $X(n) = 5$

Kalman Filter

TECHNISCHE UNIVERSITÄT DARMSTADT