TA ed.

FORSCHUNG - AUSBILDUNG - WEITERBILDUNG

BERICHT NR. 13

FUNDAMENTALS OF IDENTIFICATION OF TIME SERIES

W. KRÜGER, H. NEUNZERT, M. SCHULZ-REESE

(Workshop on Mathematics, Trieste, 13-24. May 1985)

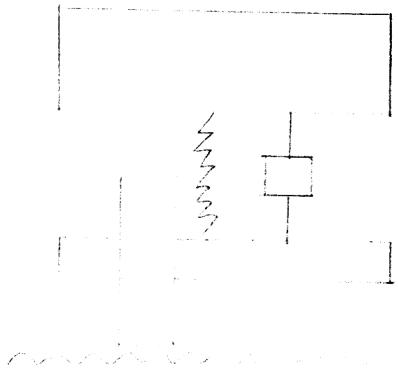
UNIVERSITÄT KAISERSLAUTERN
FACHBEREICH MATHEMATIK
ERWIN-SCHRÖDINGER-STRASSE
6750 KAISERSLAUTERN

MAI 1985

Content

IDENTIFICATION

The road-vehicle system may be considered as an "abstract system", i.e. a mapping from an inputspace into an output space.



The loadfunction y(t) depends on

the road	⟷	input
the driver the car	⟷	system
y(t)	⟷	output

Then

We are now assuming that the "driver is fixed", the road can be considered as time-depending input u(t).

If we have good informations about the physical system, we could use them for constructing a mathematical model of that system. If we have only few informations (tube!), we have to be very general.

Identification is defined by Zadehs [1] as follows:

"Identification is the determination, on the basis of input and output, of a system within a specified class of systems, to which the system under test is equivalent".

What means "system", "equivalent", "specified class"?

Generally, a system is a mapping from an inputspace into an outputspace.

"Identification" means then the following task:

Given $(u_j, y_j)_{j \in J}$, a set of pairs of input-output, determine a mapping T out of a certain class $\mathcal{T}$ of mappings, such that

$$y_j = T(u_j),\ j \in J.$$

If inputspace = outputspace = $\mathbb{R}$, $\mathcal{T}$ the set of linear mappings from $\mathbb{R}$ to $\mathbb{R}$ and (3,4) is given, then $y = T(x) = \alpha x$ is given: $\alpha = \frac{4}{3}$. If (3,4), (1,2) is given, the problem has no solution. But we can look for a mapping T, such that for example

$$(4-3a)^2 + (2-1a)^2 \quad \text{minimal!}$$

(this is the least square fit of a straight line to the given data).

But in reality, things are very much more complicated, inputs and outputs are functions or even stochastic processes.

Typical problems of identification of a system occur in

- ecology and meteorology (input: data from the climate; output: flow in the rivers)
- Natural gas storages (input: injection/projection history; output: pressure, ...) (see [2])
- Stockmarket (input: white noise; output: stockhistories) (see [3])

The last example shows that sometimes the input is not very obvious.

This first lecture on identification is used to clear up the different concepts of identification. In literature many different things are denoted by the same name. Therefore in order to use the literature you have to distinguish between the different concepts.

A general (abstract) introduction to systemtheory is given in the book of Kalman, Arbib, Falb [4] and in some articles by J. Willems [5].

We make it shorter and less rigorous: Input- and outputspaces ($\mathcal{U}$ and $\mathcal{O}$) are spaces of vectorvalued functions or generalized functions on $\mathbb{R}$ or $\mathbb{Z}$ with an algebraic and topological structure. Since we are from now on only considering linear problems, we assume that $\mathcal{U}$, $\mathcal{O}$ are linear spaces and that $\mathcal{T}$ is a subset of the set of all linear mappings from $\mathcal{U}$ to $\mathcal{O}$.

A system $T \in \mathcal{T}$ is called time-invariant, if

$$(Tu_{t_0})(t) = (Tu)_{t_0}(t), \quad t, t_0 \in \mathbb{R} \text{ or } \mathbb{Z}$$

where $v_{t_0}(t) := v(t-t_0)$

$T \in \mathcal{T}$ is called causal, if $u(t) = 0$ for $t \leq t_0$ implies $(Tu)(t) = 0$ for $t \leq t_0$.

T is called stable, if it is bounded.

If you know almost nothing, then you take a model which is very general, has a lot of degrees of freedom. Therefore the identification, i.e. the adaption of the model to the data will be rather good, but its use for prediction is less!

(A) The Impulse-Response-Model (IRM) (also called the general linear model)

$\mathcal{U} = L_\infty^m$ and $(Tu)(t) := \int_{-\infty}^{+\infty} S(t,\tau)u(\tau)d\tau$.

Since $u(t) \in \mathbb{R}^m$ (m-dim. input) and if we assume that $y(t) \in \mathbb{R}^n$ (n-dim. output), then

$$S(t,\tau) = (S_{ij}(t,\tau)/1 \leq i \leq n,\ 1 \leq j \leq m)$$

is a $(n \times m)$-matrix.

It is called Impulse-response-model, for you may interprete

$$S(t,\tau) \quad \text{as} \quad T[\delta_\tau](t),$$

using the "δ-function δ_τ as input. But this is quite formal.
If $\mathcal{U} \in \ell_\infty^m(\mathbb{Z})$ (i.e. U consists of bounded sequences $(u_k)_{k\in\mathbb{Z}}$) and $\mathcal{O} = \ell_\infty^n(\mathbb{Z})$, then the IRM is defined by

$$y_k = \sum_{\nu=-\infty}^{+\infty} S_{k,\nu} u_\nu$$

The IRM is

time-invariant, if $S(t,\tau) = \tilde{S}(t-\ )$ or $S_{k,\nu} = \tilde{S}_{k-\nu}$

causal, if $S(t,\tau) = 0$ for $t < \tau$ or $S_{k,\nu} = 0$ for $k < \nu$

stable, if $S(t,\cdot) \in L^1$ or $S_{k,\cdot} \in \ell_1$

A time-invariant causal IRM has therefore the shape

$$y(t) = \int_{-\infty}^{+\infty} \tilde{S}(t-\tau)u(\tau)d\tau \quad \text{or} \quad y_k = \sum_{\nu=-\infty}^{k} \tilde{S}_{k-\nu} u_\nu$$

and we are considering only that kind of IRM in the future.

Identificationproblem: Given u, y, determine $\tilde{S}$.
The problem is easy, if $u,y \in L^1$ (ℓ^1): Take the Fouriertransform

$$F[u](\omega) := \frac{1}{2\pi} \int_{-\infty}^{+\infty} u(t)e^{-i\omega t}dt \qquad \text{(transform each component)}$$

and use the convolution theorem, which tells you that

$$F[y](\omega) = 2\pi F[\tilde{S}](\omega) \cdot F[u](\omega)$$

from which you may calculate $F[\tilde{S}]$ and finally $\tilde{S}$.
But this doesn't work, if u, y are not in L^1 - which will be the case for stationary stochastic processes.

(B) The Dynamical System Model (DSM)

The input-output mapping is given by the dynamical system

$$(*) \qquad \begin{aligned} \dot{x} &= F(t)x + G(t)u \\ y &= H(t)x \end{aligned}$$

where $x(t) \in \mathbb{R}^p$ is called a "state of the system", $F(t) \in \mathbb{R}^{p\times p}$ the "system matrix", $G(t) \in \mathbb{R}^{p\times m}$ the "inputmatrix" and $H(t) \in \mathbb{R}^{n\times p}$ the "outputmatrix".

(*) together with an initial value $x(\overset{o}{t})$ defines the system (if F, G, H are const.).

We will assume that $x(-\infty) = 0$ for all systems. Then (F,G, H) defines the system uniquely.

Each DSM is an IRM:

$$x(t) = X(t,t_o)\,x\,(t) + \int_{t_o}^{t} X(t,\tau)G(\tau)u(\tau)d\tau$$

$$\Rightarrow\ y(t) = \int_{-\infty}^{t} H(t)X(t,\tau)G(\tau)u(\tau)d\tau\ .$$

The system is causal with $S(t,\tau) = H(t)X(t,\tau)G(\tau)$, where $X(t,\tau)$ is the transition matrix of $\dot{x} = F(t)x$.

Since $X(t,\tau) = X(t,0)\cdot X(0,\tau)$, we have

$$S(t,\tau) = H(t)X(t,0)\cdot X(0,\tau)G(\tau) = P(t)Q(\tau),$$

i.e. $S(t,\tau)$ is separated with respect to the variables t, τ.

This is also sufficient for an IRM to be an DSM:

An IRM is a DSM, if and only if $S(t,\tau) = P(t)Q(\tau)$, [6].

Time-invariance means here: F, G, H are independent of t. Then

$$X(t,\tau) = e^{(t-\tau)F} = \tilde{X}(t-\tau)$$

and an IRM is DSM if

$$\tilde{S}(t-\tau) = H(t)e^{tF}\cdot e^{-\tau F}G(\tau) = P(t)Q(\tau)$$

which can only be true, if and only if

$\tilde{S}_{ij}(t)$ is a linear combination of terms of the form

$$t^r e^{\lambda_s t} .$$

We are talking now on time-invariant DSM.

Identification means here: Determine a suitable dimension p of the state space and then the matrices F, G, H. The dimension p of the state space is important - in a Finite Element model of a tube it may go up to several thousands. Moreover, one may have a lot of redundant parameters, which do not influence the input-output behaviour.

There arise two questions:

1) If p is given, which (F, G, H) define the same input-output behaviour?
2) Can one eventually decrease p without influencing the input-output behaviour?

Answer to question 1: p given, let $(F,G,H) \in \Sigma = \mathbb{R}^{p\times p} \times \mathbb{R}^{p\times m} \times \mathbb{R}^{n\times p}$.

If $x(t) = A\tilde{x}(t)$, A nonsingular p×p-matrix, then

$$\dot{\tilde{x}} = A^{-1}FA\tilde{x} + A^{-1}Gu$$

$$y = HA\tilde{x}$$

It follows that $(A^{-1}FA, A^{-1}G, HA)$ gives the same input-output behaviour; we call therefore $(A^{-1}FA, A^{-1}G, HA)$ equivalent to (F,G,H) and call $\Sigma_o := \Sigma/\simeq$ the set of equivalence classes.

The topological structure of Σ_o is not trivial - consider for these questions [7].

Answer to question 2:

Here we need two notations from the theory of control: (F,G,H) is called *completely controllable* ($(F,G,H) \in \Sigma^{co}$), if one of the following three equivalent conditions hold:

(α) Given $t_o < t_1, x_o, x_1$ arbitrarily, there exists a function $u(t)$, $t_o \leq t \leq t_1$, such that the solution of $\dot{x} = Fx + Gu$ with $x(t_o) = x_o$ fulfills $x(t_1) = x_1$.

(β) $\dot{x} = Fx + Gu$ is <u>not</u> equivalent to a system

$$\dot{\xi}_1 = F_{11}\xi_1 + F_{12}\xi_2 + G_1 u$$
$$\dot{\xi}_2 = F_{22}\xi_2$$

where $x = (\xi_1, \xi_2)$, $1 \leq \dim \xi_1 \leq p-1$ (i.e. the state space cannot be separated into two parts, where the second one is neither influenced by the first nor by the input).

(γ) rank $(G|FG|\ldots|F^{p-1}G) = p$

(F,G,H) is called *completely observable* ($(F,G,H) \in \Sigma^{ob}$), if one of the following *equivalent* conditions hold:

(α') $x(t)$ is known in $[t_o,t_1]$ if $u(t)$, $y(t)$ are known in $[t_o,t_1]$.

(β') $\dot{x} = Fx+Gu$, $y = Hx$ is <u>not</u> equivalent to a system

$$\dot{\xi}_1 = F_{11}\xi_1 + G_1 u$$
$$\dot{\xi}_2 = F_{21}\xi_1 + F_{22}\xi_2 + G_2 u$$
$$y = H_1\xi_1$$

(i.e. the output is influenced only by ξ_1, which itself is not influenced by ξ_2).

(γ') rank $(H^T|F^TH^T|\ldots|(F^T)^{p-1}H^T) = p$.

There is now a canonical structure theorem, proved by Kalman, which says: Every DSM is equivalent to another DSM (F,G,H) of same state space dimension p, where

$$x = \Big(\underbrace{x_A}_{\substack{\text{controll.}\\ \text{not observ.}}} \Big| \underbrace{x_B}_{\substack{\text{controll.}\\ \text{observ.}}} \Big| \underbrace{x_C}_{\substack{\text{not controll.}\\ \text{not observ.}}} \Big| \underbrace{x_D}_{\substack{\text{not controll.}\\ \text{observ.}}} \Big)$$

(Dimensions $p_A + p_B + p_C + p_D = p$.

$(p_A,\ldots,p_D)$ "characteristic numbers"), such that

$$F = \begin{pmatrix} * & * & * & * \\ 0 & F_{BB} & 0 & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}, \quad G = \begin{pmatrix} * \\ G_B \\ 0 \\ 0 \end{pmatrix}, \quad H = (0 \quad H_B \quad 0 \quad *)$$

($*$ denotes a non-empty matrix, as well as F_{BB}, G_B, H_B).

There exists an algorithm to produce this canonical representation.

Now the main result for question 2 is, that p can be reduced to p_B:

The input-output behaviour of (F,G,H) is the same as that of (F_{BB},G_B,H_B). This is shown by the fact that the corresponding IRM is given by

$$\tilde{S}(t) = He^{Ft}G = H_B e^{F_{BB}t} G_B .$$

The same IRM cannot be represented by a DSM with $p < p_B$, the controllable and observable part is "irreducible".

The real object to look for is therefore

$$\Sigma_o^{co,ob}$$

the equivalence class of completely observ., completely controll. systems!

We just mention that everything can be done in the discrete case as well.

Then

$$x_{k+1} = Fx_k + Gu_k$$
$$y_k = Hu_k \quad , \qquad k \in \mathbb{Z} \qquad (u_{-\infty} = 0)$$

Again here we may define Σ, Σ_o, $\Sigma_o^{co,ob}$.

But there is a third kind of models around, at least in the discrete case:

(C) Autoregressive moving Average Models (ARMA)

We consider only the discrete case. For a one-dimensional series $u \in \ell_\infty$ we define the forward and backward shift by $(Zu)_k = u_{k+1}$, $(Bu)_k = u_{k-1}$; then for example $(Z^n u)_k = u_{k+n}$.

Let P, Q now be (u×n) and (u×m) matrices respectively, whose coefficients are polynomials in z and consider the "ARMA" model given by

$$P(z)y = Q(z)u.$$

If for example $m = n = 1$, $P(z) = z+a$, $Q(t) = 1$, we have the model

$$y_{k+1} + ay_k = u_k,$$

so that, if y_o is given, we have

$$y_k = \sum_{\nu=1}^{k} (-a)^{k-\nu} u_\nu + (-a)^k y_o \ .$$

If degree (det(P(z))) =: $p \geq 1$, but Q(z) is constant, we call the model an "autoregressive" model: AR(p). If P is constant and only degree (det Q(z)) =: $q \geq 1$, then we called it a "moving average" model: MA(q). Generally, we have ARMA (p,q).

Here obviously, the input-output behaviour is equal for systems (P,Q) and (P',Q'), if $P' = MP$, $Q' = MQ$ with a nonsingular, constant matrix - this defines again an equivalence. If we denote the pairs (P,Q) for fixed p, q by S, we again have an equivalence class $S_o = S/\cong$.

Guidorzi [2] has shown that Σ_o^{ob} is isomorphic to S_o, i.e. to each initial value for a dynamical system of Σ_o^{ob} you find initial values and an ARMA-model with the same input-output and vice versa.

In each equivalence class of Σ_o, S_o one may find a simple representative, a so-called canonical element. The structure and the parameters of these canonical elements have to be identified; there is a nice algorithm by Guidorzi [2], which is efficient and can be used on small computers.

The example treated in the paper is the gas storage prediction mentioned above.

Summary: There are many identification algorithms in the literature - look carefully which model they identify. Put into your model as much physical information as you have. The state space dimension p can also be estimated by the "information" content of the input-output (see [8]).

STOCHASTIC PROCESSES

(This chapter is mainly a selection from the two volumes of Priestley [9].)

We assume the participant to be familiar with the elementary definitions of probability theory, for example random variable, expectation-value etc. We may therefore begin by defining the notion of a stochastic process:

A family of realvalued random variables $(X_t)_{t\in T}$, indexed by a real parameter $t\in T$ (T will usually be again $\mathbb{R}$ or $\mathbb{Z}$), is called a stochastic process. If T is $\mathbb{R}$ or an intervall in $\mathbb{R}$, we call (X_t) a continuous parameter process, if $T = \mathbb{Z}$ or $\mathbb{N}$ etc., we call it a discrete parameter process.

For our purposes it is sufficient to assume that the stochastic process is given by the joint probability distribution of $(X_{t_1},\ldots,X_{t_m})$, where $\{t_1,\ldots,t_m\}$ is an arbitrary finite set in T:

$$F_{(t_1,\ldots,t_m)}(x_1,\ldots,x_m) = \text{Prob}((X_{t_1} < x_1) \cap \ldots \cap (X_{t_m} < x_m))$$

$F_{(t_1,\ldots,t_m)}$ may be defined through a density

$$f_{(t_1,\ldots,t_m)} = \frac{\partial^m}{\partial x_1 \ldots \partial x_m} F_{(t_1,\ldots,t_m)} .$$

Expectation value and autocovariance of the process are then defined by

$$\mu_t = E(X_t) = \int_{-\infty}^{+\infty} x \, dF_t(x) = \int_{-\infty}^{+\infty} x \, f_t(x)dx$$

$$\text{Cov}(X_t,X_s) = E[(X_t-\mu_t)(X_s-\mu_s)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_t)(y-\mu_s)dF_{(t,s)}(x,y)$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_t)(y-\mu_s)f_{(t,s)}(x,y)dxdy =: R(t,s)$$

Especially $R(t,t) = \text{Cov}(X_t,X_t) = \text{Var}(X_t) = \sigma_t^2$ is called the variance of (X_t).

A stochastic process is called stationary, if for all finite sets $\{t_1,\ldots,t_m\}\subset T$ and all $t\in T$

$$F_{(t_1+t,\ldots,t_m+t)}(x_1,\ldots,x_m) = F_{(t_1,\ldots,t_m)}(x_1,\ldots,x_m) \ .$$

It follows that

$$E(X_t) = \mu \qquad \text{is independent of } t$$

and $R(t,s) = \mathrm{Cov}(X_t,X_s) = \mathrm{Cov}(X_{t-s},X_0) = \tilde{R}(t-s)$ is a function of t-s only (if the Cov. exists at all, i.e. if we have a L^2-process).

Especially $\sigma_t = \sigma$ is independent of t!

A process is called <u>wide sense stationary</u>, if $E(X_t)$ is independent of t and $R(t,s)$ depends only on t-s. Since this kind of stationarity is enough for our purposes, we shall omit the "wide sense" in the future. So, (X_t) will be a stationary process with mean value μ and <u>autocovariance function</u> $R(\tau) = \mathrm{Cov}(X_t,X_{t+\tau})$.

$$\rho(\tau) := \frac{R(\tau)}{R(0)} = \frac{R(\tau)}{\sigma^2}$$

is called <u>autocorrelation function</u>.

Properties of R and ρ respectively

(a) $R(0) = \sigma^2$, $\rho(0) = 1$

(b) $|R(\tau)| \leq R(0)$, $|\rho(\tau)| \leq 1$ for all τ

(c) $R(\tau)$, $\rho(\tau)$ are even functions (since (X_t) is realvectorvalued)

(d) $R(\tau)$, $\rho(\tau)$ are positive semidefinite functions, i.e.

$$\sum_{r=1}^{n}\sum_{s=1}^{n} R(t_r-t_s)x_r x_s \geq 0 \quad \text{for arbitrary finite sets}$$

$\{t_1,\ldots,t_n\}$ and $\{x_1,\ldots,x_n\}$.

<u>Remark:</u> Stationarity does not mean that a single trajectory does not really depend on time; only the stochastic properties are time independent. Stationary processes occur in electric circuits, vibration, turbulence, economic time series, description of the road etc. Nonstationary processes describe time evolutions and occur in growth and decay of population, queuing systems, chain reactions, seismic observations etc..

Some special stationary discrete parameter processes:

Let $T = \mathbb{Z}$.

(A) White noise process (or "purely random process")

is defined by

$$R(k) = \begin{cases} 0 & \text{for } k \neq 0 \\ \sigma^2 & \text{for } k = 0 \end{cases} \qquad \left(\rho(k) = \begin{cases} 0, & k \neq 0 \\ 1, & k = 0 \end{cases}\right)$$

If additionally $\mu = 0$, we denote such a process by ε_t. Please remark that we assume nothing about the single distribution $f_k(x)$ (besides the fact that $E(X_k) = 0$), but we demand that X_k, X_ℓ are uncorrelated (not independent!) for $k \neq \ell$.

Remark: The definition of a white noise continuous parameter process is much more complicated and technical.

(B) Autoregressive process (AR)

(X_t) is called a first order autoregressive process (AR(1)) if it satisfies the difference equation

$$X_t - aX_{t-1} = \varepsilon_t \ .$$

(X_t) is called an autoregressive process of order k (AR(k)) if it satisfies the difference equation

$$X_t + a_1 X_{t-1} + \ldots + a_k X_{t-k} = \varepsilon_t \ .$$

Let us first consider an AR(1) - process.

$$X_k - aX_{k-1} = \varepsilon_k \Rightarrow X_k = \sum_{j=0}^{k-1} a^j \varepsilon_{k-j} + a^k X_0 \ .$$

We assume here that X_0 is a *given randomvariable*. We get

$$E(X_k) = a^k E(X_0)$$

(since $E(\varepsilon_k) = 0$) and in order (X_k) to be stationary, $E(X_0) = 0$ is necessary.

To calculate the autocovariance, one calculates

$$\mathrm{Cov}(X_k, X_{k+r}) = E(X_k \cdot X_{k+r}) = E[(\sum_{j=0}^{k-1} a^j \varepsilon_{k-j} + a^k X_0) \cdot (\sum_{\ell=0}^{k+r-1} a^\ell \varepsilon_{k+r-\ell} + a^{k+r} X_0)]$$

$$= \begin{cases} a^{2k+r}\mathrm{Var}(X_0) + \sigma_\varepsilon^2 a^{|r|} \dfrac{1-a^{2k}}{1-a^2} & \text{for } |a| \neq 1 \\ \mathrm{Var}(X_0) + |r|\sigma_\varepsilon^2 & \text{for } |a| = 1 \end{cases}$$

(σ_ε^2 is the variance of ε_k!)

One recognizes that the stationarity of (X_k) depends on the distribution of X_0:

(α) If X_0 is deterministic, $E(X_0) = \mathrm{Var}(X_0) = 0$, we get

$$R(k,k+r) = \sigma_\varepsilon^2 \, a^{|r|} \frac{1-a^{2k}}{1-a^2} \qquad (\text{for } |a| \neq 1),$$

which depends on k; the process is not stationary.

For $|a| < 1$, we have atleast

$$\lim_{k\to\infty} R(k,k+r) = \sigma_\varepsilon^2 \frac{a^{|r|}}{1-a^2} \; ;$$

for large k, the process can be considered as stationary and is therefore called *"asymptotically stationary"* (for large k the process forgets the "violence" due to the deterministic choice of X_0).

(β) If we choose $E(X_0) = 0$, $\mathrm{Var}(X_0) = \dfrac{\sigma_\varepsilon^2}{1-a^2}$, then the k-depending part in $R(k,k+r)$ cancels out and we get

$$R(r) = \frac{\sigma_\varepsilon^2 \, a^{|r|}}{1-a^2} \; , \quad \rho(r) = a^{|r|}.$$

(γ) We choose as initial value $X_{-n} = 0$ deterministic and let n tend to infinity. We get

$$X_k = \sum_{j=-\infty}^{k} a^{k-j} \varepsilon_j \qquad (\text{the convergence is taken in the mean square sense, i.e. } E(|X_k - \sum_{j=-N}^{k} a^{k-j}\varepsilon_j|^2) \to 0 \text{ with } N \to \infty.)$$

Naturally the variance is then

$$\mathrm{Var}(X_k) = \mathrm{Var}\Big(\sum_{\ell=0}^{\infty} a^{\ell}\varepsilon_{k-\ell}\Big) = \sigma_\varepsilon^2 \sum_{\ell=0}^{\infty} a^{2\ell}$$

which converges, if $|a| < 1$.

Next, one may use AR(2)

$$X_k + a_1 X_{k-1} + a_2 X_{k-2} = \varepsilon_k \qquad \text{or}$$

$$(**) \qquad (1 + a_1 B + a_2 B^2) X_k = \varepsilon_k$$

Denoting by ζ_1, ζ_2 the zeros of the polynomial $\zeta^2 + a_1\zeta + a_2$ then $(1+a_1B+a_2B^2) = (1-\zeta_1 B)(1-\zeta_2 B)$, such that

$$X_k = \frac{1}{(1-\zeta_1 B)(1-\zeta_2 B)}\,\varepsilon_k = \frac{1}{\zeta_1 - \zeta_2}\Big(\frac{\zeta_1}{1-\zeta_1 B} - \frac{\zeta_2}{1-\zeta_2 B}\Big)\varepsilon_k \,.$$

In this way we get a special solution of the inhomogeneous equation (**) as

$$X_k = \sum_{j=0}^{\infty} \Big(\frac{\zeta_1^{j+1} - \zeta_2^{j+1}}{\zeta_1 - \zeta_2}\Big)\varepsilon_{k-j} \,;$$

the general solution of the homogeneous equation is $A_1\zeta_1^k + A_2\zeta_2^k$, so that the general solution of (**) is

$$X_k = \sum_{j=0}^{\infty} \Big(\frac{\zeta_1^{j+1} - \zeta_2^{j+1}}{\zeta_1 - \zeta_2}\Big)\varepsilon_{k-j} + A_1\zeta_1^k + A_2\zeta_2^k \,. \qquad (\zeta_1 \neq \zeta_2)$$

This solution is asymptotically stationary, if $|\zeta_1|, |\zeta_2| < 1$. It may even be stationary for appropriate initial distributions X_0, X_1.

Assuming that (X_k) is stationary, one can easily calculate $\rho(\tau)$:

From $X_k + a_1 X_{k-1} + a_2 X_{k-2} = \varepsilon_k$ follows

$$E(X_k X_{k-r}) + a_1 E(X_{k-1} X_{k-r}) + a_2 E(X_{k-2} X_{k-r}) = E(\varepsilon_k X_{k-r}) \,.$$

For $r = 0$ we get $\sigma_X^2 + a_1 R(1) + a_2 R(2) = \sigma_\varepsilon^2$,

for $r = 1$ $\qquad R(1) + a_1 R(0) + a_2 R(1) = 0$,

for $r = 2$ $\qquad R(2) + a_1 R(1) + a_2 R(0) = 0$.

Solving these three equations gives $\rho(0)$, $\rho(1)$, $\rho(2)$ and analog the others

$$\rho(r) = \frac{(1-\zeta_2^2)\zeta_1^{r+1} - (1-\zeta_1^2)\zeta_2^{r+1}}{(\zeta_1-\zeta_2)(1+\zeta_1\zeta_2)} \quad .$$

One may now distinguish between the cases, where ζ_1, ζ_2 are real or complex etc. Similar calculations can be done for AR(k).

(C) Moving average process (MA)

(X_t) is called a moving average process of order j (MA(j)) if it satisfies the equation

$$X_t = b_0\varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_j\varepsilon_{t-j} \quad .$$

One has $\sigma_X^2 = \sigma_\varepsilon^2 \sum_{r=0}^{j} b_r^2$, $E(X_t) = 0$ and

$$\rho(r) = \begin{cases} \dfrac{b_r b_0 + \ldots + b_j b_{j-|r|}}{\sigma_X^2}\,\sigma_\varepsilon^2 & \text{for } 0 \leq |r| \leq j \\ 0 & \text{else} \end{cases}$$

$$= \begin{cases} \dfrac{\sum_{\ell=r}^{j} b_\ell b_{\ell-r}}{\sum_{\ell=0}^{j} b_\ell^2} & \text{if } 0 \leq |r| \leq j \\ 0 & \text{if } |r| > j \end{cases}$$

$\rho(r)$ has a finite cut at $r \pm j$!

The process is always stationary.

(D) Autoregressive/moving average process (ARMA)

(X_t) is called an autoregressive/moving average process of order (j,ℓ) (ARMA(j,ℓ)) if it satisfies an equation of the form

$$X_t + a_1X_{t-1} + \ldots + a_jX_{t-j} = b_0\varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_\ell\varepsilon_{t-\ell}$$

ARMA(j,ℓ) may be written as

$$\alpha(B)X_k = \beta(B)\varepsilon_k$$

with

$$\alpha(\zeta) = 1 + a_1\zeta + \ldots + a_j\zeta^j, \quad \beta(\zeta) = b_0 + b_1\zeta + \ldots + b_\ell\zeta^\ell.$$

With $f(\zeta) = \zeta^j\alpha(\frac{1}{\zeta}) = \prod_{i=1}^{j}(\zeta-\mu_i)$ we get again an asymptotically stationary solution, if $|\mu_i| < 1$, $i = 1,\ldots,j$.

(E) The general linear process

(X_t) is said to be a general linear process if it can expressed in the form

$$X_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u}$$

with

$$\sum_{u=0}^{\infty} g_u^2 < \infty.$$

(X_t can be observed as the reaction of a linear time-invariant causal system - defined by the impulse response (g_u) - when the input is white noise.)

Computation of $R(\tau)$ and $\rho(\tau)$:

$$X_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u} = \sum_{u=-\infty}^{\infty} g_u \varepsilon_{t-u} \qquad \text{if we define } g_u=0,\ u<0$$

$E(\varepsilon_t) = 0$ then $E(X_t) = 0$ and

$$R(\tau) = E(X_t X_{t+\tau}) = \sigma_\varepsilon^2 \left(\sum_{u=-\infty}^{\infty} g_u g_{u-\tau} \right)$$

and

$$\rho(\tau) = \frac{R(\tau)}{\sigma_X^2} = \frac{\sum_{u=-\infty}^{\infty} g_u g_{u-\tau}}{\sum_{u=0}^{\infty} g_u^2}$$

For $\sum_{u=0}^{\infty} g_u^2 < \infty$, (X_t) is always stationary.

(F) Harmonic processes are defined by

$$X_t = \sum_{j=1}^{s} A_j \cos(\omega_j t + \phi_j)$$

where A_j, ω_j are fixed numbers, but the phases ϕ_j are independent random variables with uniform distribution on $(-\pi,\pi)$, i.e.

$f_\phi(\varphi) = \frac{1}{2\pi}$ for $-\pi \leq \varphi \leq \pi$.

The trajectories are all sin-curves with arbitrary shifted argument.

Due to the independence of $(\phi_j)_{j \in \{1,\ldots,s\}}$, one gets easily

$$E(X_t) = 0,\ R(\tau) = \sum_{j=1}^{s} (\tfrac{1}{2} A_j^2)\cos \omega_j \tau,\ \sigma_x^2 = \sum_{j=1}^{s} \tfrac{1}{2} A_j^2.$$

Therefore, the process is stationary.

Fourier Analysis of stochastic processes

For a stationary process $(X_t)_{t\in\mathbb{R}}$ (or better: a trajectory of such a process), the Fourier transform will not exist in general. This trajectory will neither be L^1 nor periodic in general. One has therefore to change the method a little bit.

Let $(X_t)_{t\in\mathbb{R}}$ be a stationary process with $E(|X_t - X_{t_o}|^2) \to 0$ for $t \to t_o$ and $E(X_t) = 0$. We cut the trajectory $X(t)$ in multiplying with a "window" $\chi_{[-T,T]}(t) = \begin{cases} 1 & \text{for } |t| \leq T \\ 0 & \text{else} \end{cases}$ and get $X_T(t) = X(t)\chi_{[-T,T]}(t)$. X_T has compact support and may be written as a Fouriertransform

$$X_T(t) = \int_{-\infty}^{+\infty} G_T(\omega) e^{i\omega t} d\omega$$

with $$G_T(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X_T(t) e^{-i\omega t} dt \ .$$

(These formulas are correct, since $X_T \in L^1$ and "stochastically" continuous.)
$|G_T(\omega)|^2$ may be considered as the spectral energy density: $\int_{\omega_1}^{\omega_2} |G_T(\omega)|^2 d\omega$ is the energy contained in the frequencies between ω_1 and ω_2. But we cannot open the window widely, i.e. the limit $T \to \infty$ does not exist for $|G_T(\omega)|^2$. (The total energy in the whole trajectory is not finite.)
But if we consider instead of the energy the power, the limit of $\frac{|G_T(\omega)|^2}{2T}$, $T \to \infty$ may exist. Don't forget that this expression depends on the trajectory. We define therefore

$$h(\omega) := \lim_{T\to\infty} E\left(\frac{|G_T(\omega)|^2}{2T}\right)$$

if it exists and call h the spectral power density.

The most important result in this domain consists in the fact that this $h(\omega)$ is nothing else but the Fouriertransform of the autocovariance $R(\tau)$.

Theorem:

If $R \in L^1$, then

$$h(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega\tau} R(\tau) d\tau = \hat{R}(\omega).$$

The idea of the proof consists in using the convolution theorem:

If $\hat{f}$, $\hat{g}$ are Fouriertransforms of f, g, then

$$\hat{f} \cdot \bar{\hat{g}} = \hat{k} \quad \text{with} \quad k(t) = \int_{-\infty}^{+\infty} f(u) g(u-t) du .$$

Therefore $|G_T(\omega)|^2 = \hat{X}_T(\omega)\overline{\hat{X}_T(\omega)}$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega t} \left(\int_{-\infty}^{+\infty} X_T(u) X_T(u-t) du \right) dt \quad \text{and}$$

$$\frac{|G_T(\omega)|^2}{2T} = \hat{S}_T(\omega) \quad \text{with} \quad S_T(t) = \frac{1}{2T} \int_{-\infty}^{+\infty} X_T(u) X_T(u-t) du$$

$$\Rightarrow \quad h(\omega) = \lim_{T \to \infty} \widehat{E(S_T)}(\omega).$$

Now

$$S_T(t) = \begin{cases} \frac{1}{2T} \int_{-(T-|t|)}^{T} X(u) X(u-|t|)\, du & \text{for } |t| \leq 2T \\ 0 & \text{for } |t| > 2T \end{cases}$$

and therefore

$$E(S_T(t)) = \begin{cases} \frac{1}{2T} \int_{-(T-|t|)}^{T} R(t) du = R(t)\left(1 - \frac{|t|}{2T}\right) & \text{for } |t| \leq 2T \\ 0 & \text{for } |t| > 2T \end{cases}$$

We end up with

$$h(\omega) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-2T}^{2T} (1 - \frac{|t|}{2T}) e^{-i\omega t} R(t)dt$$

$$= \hat{R}(\omega), \text{ if } R \in L^1.$$

This was a sketch of the proof.

Now, if $R \in L^1$ and continuous, we have also

$$R(t) = \int_{-\infty}^{+\infty} e^{i\omega t} h(\omega)d\omega \quad \text{or, since R is an even function}$$

$$R(t) = \int_{-\infty}^{+\infty} \cos \omega t \; h(\omega)d\omega \quad .$$

We realize that $R(0) = \sigma_x^2 = \int_{-\infty}^{+\infty} h(\omega)d\omega < \infty$ and from the definition of h follows that $h(\omega) \geq 0$ for all ω. $f(\omega) = \frac{h(\omega)}{\sigma_x^2}$ is called the <u>normalized spectral power density</u>:

$$\int_{-\infty}^{+\infty} f(\omega)d\omega = 1.$$

It is clear that if $\rho \in L^1$ and ρ continuous, then

$$\rho(t) = \int_{-\infty}^{+\infty} e^{i\omega t} f(\omega)d\omega \quad \text{and} \quad f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega t} \rho(t)dt.$$

If ρ is not in L^1, f may not exist. But there is the theorem of *Wiener-Khintchine*: Let ρ be continuous. Then ρ is an autocorrelation function of a stationary stochastic process, if and only if there exists a distribution function $F(\omega)$ such that

$$\rho(t) = \int_{-\infty}^{+\infty} e^{i\omega t} dF(\omega)$$

The proof relies on the fact that ρ is positive semi definit and on the theorem of Bochner.

For discrete parameter processes similar relations are true: A function $\rho(r)$, $r \in \mathbb{Z}$ is an autocorrelation of a stationary, discrete parameter process $(X_k)_{k\in\mathbb{Z}}$ if and only if there exists a distribution function $F(\omega)$ for

$-\pi \leq \omega \leq \pi$, such that

$$\rho(r) = \int_{-\pi}^{+\pi} e^{i\omega r} dF(\omega), \qquad r \in \mathbb{Z}.$$

(Realize that F is now only defined in $[-\pi,\pi]$; this is due to the fact that for $r \in \mathbb{Z}$ the number $e^{i\omega r}$ and $e^{i(\omega+2\pi)r}$ are not distinguishable.)

Moreover: If $\rho \in \ell_1$, then F is differentiable a.e., $F' = f$ and

$$\rho(r) = \int_{-\pi}^{+\pi} e^{i\omega r} f(\omega) d\omega.$$

To invert this relation, we need a Fourier series instead of an integral

$$f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \rho(r) e^{-i\omega r} .$$

Finally $h(\omega) = \sigma_x^2 f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} R(r) e^{-i\omega r}$

Altogether constitutes a *theorem of Wold*.

What happens, if ρ is not in L_1 or in ℓ_1 respectively? If for example ρ has periodic ingredients? There is the Lebesgue decomposition theorem, telling us that the corresponding distribution F can be decomposed as

$$F = a_1 F_1 + a_2 F_2 + a_3 F_3 \quad \text{with} \quad a_i \geq 0,\ i = 1,2,3,\ \sum_{i=1}^{3} a_i = 1,$$

where all F_i are distribution functions (growing from 0 to 1) and

(α) F_1 is absolutely continuous, i.e. $\frac{dF_1}{d\omega} = f_1$ exists a.e.

(β) F_2 is a stepfunction with steps at ω_r of height p_r, $r \in \mathbb{N}$, $\sum_r p_r = 1$.

(γ) F_3 is a socalled singular function, which can be ignored for our reasons.

(See for example the very nice book by Natanson: Theory of functions of a real variable)

Therefore we get

$$\rho(\tau) = \int_{-\infty}^{+\infty} e^{i\tau\omega} d(a_1 F_1(\omega) + a_2 F_2(\omega))$$

$$= a_1 \int_{-\infty}^{+\infty} e^{i\tau\omega} f_1(\omega) d\omega + a_2 \sum_{r \in \mathbf{N}} e^{i\tau\omega_r} \cdot p_r =: \rho_1(\tau) + \rho_2(\tau) \quad .$$

Remark that the second part ρ_2 is not necessarily periodic, since $(\omega_r)_{r \in \mathbf{N}}$ might be incommensurable. If $\omega_r = k_r \omega$ with $k_r \in \mathbb{Z}$, then ρ_2 is periodic.

If f_1 is continuous, then $f_1(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\tau\omega} \rho_1(\tau) d\tau$; p_s may be rediscovered from ρ_2 by

$$p_s = \lim_{T \to \infty} \{ \frac{1}{2T} \int_{-T}^{T} \rho_2(\tau) e^{-i\tau\omega} \, d\tau \} \quad .$$

We now consider examples:

(A) White noise: $f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{+\infty} \rho(r) e^{-i\omega r} = \frac{1}{2\pi}$, $\omega \in [-\pi, \pi]$

(since $\rho(0) = 1$, $\rho(r) = 0$ else). Therefore we have only an absolutely continuous part: $a_1 = 1$, $a_2 = 0$. (For continuous parameter processes there arise some problems, since $\sigma_x^2 = \infty$. One gets $f(\omega) = 0$ $h(\omega) = \frac{\sigma_w^2}{2\pi}$, where σ_w^2 is the variance of the underlying Wiener process.)

(B) AR(1) gives $\rho(r) = a^{|r|}$ $(|a| < 1)$. Therefore

$$f(\omega) = \frac{1}{2\pi} \sum_{r=-\infty}^{+\infty} a^{|r|} \cos \omega r = \frac{1}{2\pi} \cdot \text{real part of } [1 + 2 \sum_{r=1}^{\infty} (a e^{i\omega})^r]$$

$$= \frac{1-a^2}{2\pi(1-2a \cos \omega + a^2)} \quad , \quad \omega \in (-\pi, \pi). \text{ Therefore again } a_1 = 1.$$

Similarly, for all ARMA-models we get $a_1 = 1$, $a_2 = 0$.

(C) The exception is a harmonic process.

Let for example be $X(t) = A_1 \cos(\omega_1 t + \phi)$, where ϕ is uniformly distributed on $[-\pi, \pi]$. Then $\rho(t) = \cos \omega_1 t = \frac{e^{i\omega_1 t} + e^{-i\omega_1 t}}{2}$.

This gives for f a distribution $F = \frac{1}{2}(\delta_{\omega_1} + \delta_{-\omega_1})$, i.e.

$$F(\omega) = \begin{cases} 0 & \text{for } -\infty < \omega < -\omega_1 \\ \frac{1}{2} & \text{for } -\omega_1 < \omega < \omega_1 \\ 1 & \text{for } \omega > \omega_1 \end{cases}$$

Here we have $a_1 = 0$, $a_2 = 1$.

For a general harmonic process we get jumps in F at $\pm\omega_i$ with a height $\frac{1}{2} \frac{A_i^2}{\sum_{j=1}^{S} A_j^2}$.

Linear systems

We will now analyze what happens if we consider a stationary process as an input of a linear system. We will consider linear systems which are described by socalled convolution integrals

$$y(t) = \int_{-\infty}^{\infty} x(t-u)g(u)du = \int_{-\infty}^{\infty} x(u)g(t-u)du ,$$

with $x(t)$ the input function, $g(u)$ the socalled impulse response, which characterizes the system and $y(t)$ the output function. The above integral describes a linear, time-invariant system.

Let $(X_t)_{t\in I}$ be a stationary process, then $(Y_t)_{t\in I}$ with

$$Y_t = \int_{-\infty}^{\infty} X_u g(t-u)du$$

is a stationary process ,too.

The relation between the autocorrelation functions of the two processes is given by

$$\rho_y(\tau) = \frac{\sigma_x^2}{\sigma_y^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)g(v)\rho_x(\tau+u-v)dudv.$$

If we assume that the spectral density function $f_X(\omega)$ of the input process exists and taking Fouriertransform on each side we get with

$$\Gamma(\omega) = \int_{-\infty}^{\infty} g(t)\, e^{-i\omega t}dt \quad \text{(Transfer function)}$$

$$f_Y(\omega) = \frac{\sigma_x^2}{\sigma_y^2} |\Gamma(\omega)|^2 f_X(\omega) \quad \text{or}$$

$$h_Y(\omega) = |\Gamma(\omega)|^2 h_X(\omega) .$$

In the discrete parameter case we get similar relations. The linear, time-invariant system is given by

$$Y_t = \sum_{n=-\infty}^{\infty} g_n X_{t-n} \qquad t\in T .$$

We get

$$f_Y(\omega) = \frac{\sigma_x^2}{\sigma_y^2} \, |\Gamma(\omega)|^2 \, f_X(\omega) \qquad \text{for } \omega \in [-\pi, \pi]$$

or

$$h_Y(\omega) = |\Gamma(\omega)|^2 \, h_X(\omega) \qquad \text{for } \omega \in [-\pi, \pi]$$

with

$$\Gamma(\omega) = \sum_{n=-\infty}^{\infty} g_n e^{-i\omega n} \, .$$

If we write

$$Y_t = \sum_{n=-\infty}^{\infty} g_n X_{t-n}$$

in the form

$$Y_t = \left(\sum_{n=-\infty}^{\infty} g_n B^n \right) X_t = G(B) X_t \, ,$$

with the backward shift operator B and

$$G(z) = \sum_{n=-\infty}^{\infty} g_n z^n \, .$$

Then

$$\Gamma(\omega) = \sum_{n=-\infty}^{\infty} g_n e^{-i\omega n} = G(e^{-i\omega})$$

and therefore

$$f_Y(\omega) = \frac{\sigma_x^2}{\sigma_y^2} \, |G(e^{-i\omega})|^2 \, f_X(\omega) \, .$$

We get the general result: Are the two stationary processes $(X_t)_{t \in I}$ and $(Y_t)_{t \in I}$ related by an equation

$$Y_t = \Phi(B) X_t$$

then

$$f_y(\omega) = \frac{\sigma_x^2}{\sigma_y^2} \, |\Phi(e^{-i\omega})|^2 \, f_X(\omega) \, .$$

Examples:

(A) For AR(1) we get $(1-aB)X = \varepsilon \Rightarrow X = \frac{1}{1-aB}\varepsilon$, $\phi(B) = \frac{1}{1-aB} \Rightarrow$

$$h_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|1-ae^{-i\omega}|^2} = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{1-2a\cos\omega+a^2}$$

(B) AR(k) $\quad (1+a_1B+\ldots+a_kB^k)X = \varepsilon \Rightarrow$

$$h_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{|1+a_1e^{-i\omega}+\ldots+a_ke^{-ik\omega}|^2}$$

(C) MA(ℓ) $\quad X = (b_0+b_1B+\ldots+b_\ell B^\ell)\varepsilon \Rightarrow$

$$h_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} |(b_0+b_1e^{-i\omega}+\ldots+b_\ell e^{-i\ell\omega})|^2$$

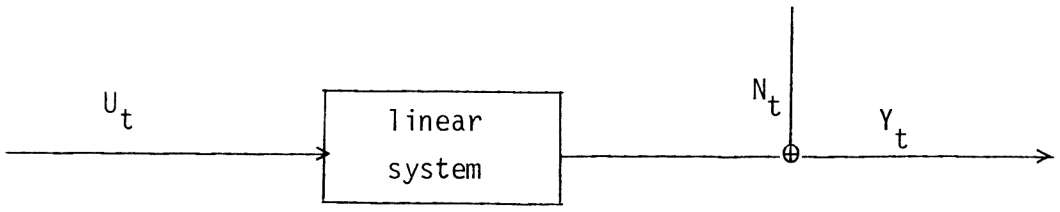
(D) ARMA(k,ℓ) $\quad (1+a_1B+\ldots+a_kB^k)X = (b_0+\ldots+b_\ell B^\ell)\varepsilon$

$$h_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\beta(e^{-i\omega})|^2}{|\alpha(e^{-i\omega})|^2} \quad \text{with} \quad \alpha(\zeta) = \sum_{j=0}^{k} a_j\zeta^j \quad (a_0 = 1)$$

and $\beta(\zeta)$ correspondingly. h is therefore a rational function of $e^{-i\omega}$.

Linear relationships with added noise

Let us consider a single input/output system in which the output is corrupted by a 'noise' disturbance N_t.



In place of the former equation

$$Y_t = \sum_{n=-\infty}^{\infty} g_n U_{t-n}$$

we have now to consider

$$Y_t = \sum_{n=-\infty}^{\infty} g_n U_{t-n} + N_t.$$

It is obvious that the simple relationship between the power spectral densities of the input and output is no longer valid since $h_y(\omega)$ now depends on Y_t and N_t. Consequently we cannot determine the transfer function from a knowledge of $h_y(\omega)$ and $h_u(\omega)$ only.

Till now we have considered specific quantities such as autocorrelation, autocovariance etc. which described an univariate process. We will now consider quantities which describe certain interrelationships between two or more processes.

In practice it often occurs that instead of observing just a single process X_t we observe (simultaneously) several processes $X_{1,t},\ldots,X_{p,t}$.

For example in engineering context we may wish to study the simultaneous variations over time of current and voltage, or pressure, temperature and volume, or seismic records taken at a number of different geographical locations. In economics f.e. we may be interested in studying inflation rates and money supply, unemployment and interest rates.

Therefore we will consider now correlation and spectral properties of multivariate stationary processes.

Suppose we are given two stochastic processes $\{X_t\}$, $\{Y_t\}$. We say $\{X_t;Y_t\}$ is a stationary bivariate process (or $\{X_t\}$, $\{Y_t\}$ are jointly stationary), if

(a) $\{X_t\}$, $\{Y_t\}$ are each stationary processes, and

(b) $\text{cov}(X_t;Y_s) = E[(X_t-\mu_x)(Y_s-\mu_y)]$ is a function of (s-t) only.

We will denote the autocorrelation or covariance functions of X_t respectively Y_t with

$$R_{xx}(\tau),\ R_{yy}(\tau) \quad \text{and} \quad \rho_{xx}(\tau),\ \rho_{yy}(\tau).$$

The above functions describe the correlation structure within each process. We will now introduce a new function which describes the correlation structure between the processes.

The cross-covariance function is defined by

$$R_{yx}(\tau) = \text{cov}\{X_t, Y_{t+\tau}\}$$

and the cross-correlation function is then given by

$$\rho_{yx}(\tau) = \frac{R_{yx}(\tau)}{\sqrt{R_{xx}(0)R_{yy}(0)}} \quad .$$

The complete covariance properties are then summarized by the covariance matrix

$$R(\tau) = \begin{pmatrix} R_{xx}(\tau) & R_{yx}(\tau) \\ R_{xy}(\tau) & R_{yy}(\tau) \end{pmatrix}$$

Some properties:

(1) The cross-covariance and cross-correlation functions are not even function. Instead we have

$$R_{yx}(\tau) = R_{xy}(-\tau) \ ,$$

this means that R_{yx} and R_{xy} contain the same informations.

(2) Analog we get

$$\rho_{yx}(\tau) = \rho_{xy}(-\tau) \quad \text{and}$$

$$|\rho_{yx}(\tau)| \leq 1$$

$\rho_{yx}(\tau)$ attains in general its maximum value not at $\tau = 0$.

Suppose $h_{xx}(\omega)$ and $h_{yy}(\omega)$ are the (non-normalized) spectral densities of the two processes $\{X_t\}$ and $\{Y_t\}$. We define the cross-spectral density function or simply the cross-spectrum by the Fourier transform of the cross-covariance

$$h_{yx}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} R_{yx}(\tau) e^{-i\tau\omega}$$

provided R_{yx} is absolutely summable $\Sigma|R_{yx}(\tau)| < \infty$. The spectral matrix is then given by

$$h(\omega) = \begin{pmatrix} h_{xx}(\omega) & h_{xy}(\omega) \\ h_{yx}(\omega) & h_{yy}(\omega) \end{pmatrix}$$

Analog we defined the normalized cross-spectrum

$$f_{yx}(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho_{yx}(\tau) e^{-i\tau\omega}$$

as the Fourier transform of the cross-correlation function.

Examples

1. The simplest example of a bivariate stationary process occurs when X_t and Y_t are uncorrelated processes, i.e. when $\text{cov}\{X_t, Y_s\} = 0$ for all s, t.

   $\Rightarrow \quad R_{xy}(\tau) = 0$ for all τ

   and consequently

   $$h_{xy}(\omega) = 0 \quad \text{for all } \omega.$$

2. Suppose X_t and Y_t satisfy a linear relationship of the form

   $$X_t = aY_t + \varepsilon_t \qquad (\varepsilon_t \text{ white noise , uncorrelated with } Y_t$$

   and without loss of generality we assume $E[Y_t] = 0 \Rightarrow E(X_t) = 0$)

$$\Rightarrow \quad R_{xy}(\tau) = E(Y_t \cdot X_{t+\tau}) = a\, E(Y_t \cdot Y_{t+\tau}) + E(\varepsilon_{t+\tau} \cdot Y_t)$$
$$= a\, R_{yy}(\tau)$$
$$\Rightarrow \quad h_{xy}(\tau) = a\, h_{yy}(\tau) \ .$$

Also we get

$$R_{xx}(\tau) = \mathrm{cov}(X_t, X_{t+\tau}) = \mathrm{cov}(aY_t + \varepsilon_t, aY_{t+\tau} + \varepsilon_{t+\tau})$$
$$= a^2 R_{yy}(\tau) + R_{\varepsilon}(\tau)$$
$$\Rightarrow \quad h_{xx}(\omega) = a^2 h_{yy}(\omega) + h_{\varepsilon}(\omega) \ .$$

A generalization of this example leads us back to our starting point:

The linear system with added noise:

$$Y_t = \sum_{n=-\infty}^{\infty} g_n U_{t-n} + N_t$$

Let us assume that U_t and N_t are uncorrelated and without loss of generality have zero mean.

Then we compute

$$R_{yu}(\tau) = \mathrm{cov}(U_t \cdot Y_{t+\tau}) = E[U_t\, Y_{t+\tau}]$$
$$= \sum_{n=-\infty}^{\infty} g_n\, E[U_t \cdot U_{t+\tau-n}]$$
$$= \sum_{n=-\infty}^{\infty} g_n [R_{uu}(\tau-n)]$$

therefore U_t and Y_t are jointly stationary.

Taking Fourier transforms of both sides we obtain

$$h_{yu}(\omega) = \sum_{\tau=-\infty}^{\infty} e^{-i\omega\tau} \Big(\sum_{n=-\infty}^{\infty} g_n\, R_{uu}(\tau-n) \Big)$$
$$= \sum_{n=-\infty}^{\infty} g_n e^{-i\omega n} \Big(\sum_{\tau=-\infty}^{\infty} e^{-i\omega(\tau-n)} R_{uu}(\tau-n) \Big)$$
$$= \Gamma(\omega)\, h_{yy}(\omega)$$
$$\Rightarrow \quad \Gamma(\omega) = \frac{h_{yu}(\omega)}{h_{yy}(\omega)} \qquad , h_{yy}(\omega) \neq 0$$

i.e. the transfer function $\Gamma(\omega)$ is the ratio of the cross-spectral density function between the input and output to the spectral density function of the input.

The relationship between input and output spectral densities we got by the following way:

We multiply both sides of

$$Y_t = \sum_{n=-\infty}^{\infty} g_n U_{t-n} + N_t$$

with $Y_{t-\tau}$, we get

$$Y_t \cdot Y_{t-\tau} = \left(\sum_{n=-\infty}^{\infty} g_n U_{t-n} + N_t \right)\left(\sum_{n=-\infty}^{\infty} g_n U_{t+\tau-n} + N_{t+\tau} \right)$$

$$\Rightarrow R_{yy}(\tau) = E[Y_t \cdot Y_{t-\tau}] = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g_n g_m R_{xx}(\tau-n+m) + R_{nn}(\tau) \ .$$

Taking Fourier transforms of both sides of this equation we find

$$h_{yy}(\omega) = |\Gamma(\omega)|^2 h_{uu}(\omega) + h_{nn}(\omega).$$

IDENTIFICATION OF STOCHASTIC PROCESSES

Estimation in the Time Domain

For practical problems the identification of stochastic processes is an important task. In the most cases we do not know really the model of the process. On the basis of observations we have to win some informations about the process. The most important characteristica of stochastic processes in the time domain are the expected value and the autocovariance or autocorrelation function. We have to determine these characteristica approximately out of the existing data.

For this task we need some notions from the statistics.

We know that random variables or stochastic processes can be completely described by the distribution function F or by systems of distribution functions. We will assume now that we know in some sense the mathematical form of the distribution function but we have a dependence on an unknown parameter $\theta \in \Theta$, which we have to estimate.

Definition

An *estimatable parameter* is a function $\gamma: \Theta \to \mathbf{R}$, i.e. a certain number $\gamma(\theta)$, dependent on θ.

An *estimator* of γ is a function

$$T_N: \mathbb{R}^N \to \mathbf{R}.$$

Are $x_{t_1},\ldots,x_{t_N}$ observations of the process, then the value $T_N(x_{t_1},\ldots,x_{t_N})$ is called *estimate* or *statistic* of $\gamma(\theta)$.

The term $T_N(X_{t_1},\ldots,X_{t_N})$ with the corresponding random variables $X_{t_1},\ldots,X_{t_N}$ is called *estimator* for $\gamma(\theta)$.

The estimator $T_N(X_{t_1},\ldots,X_{t_N})$ is called *unbiased* for γ, if

$$E_\theta(T_N(X_{t_1},\ldots,X_{t_N})) = \gamma(\theta)$$

for all $\theta \in \Theta$.

$T_N(X_{t_1},\ldots,X_{t_N})$ is called *consistent* for γ if for all $\varepsilon > 0$ and all $\theta \in \Theta$:

$$\lim_{N\to\infty} P_\theta(|T_N(X_{t_1},\ldots,X_{t_N})-\gamma(\theta)| > \varepsilon) = 0.$$

An estimator T_N which is not unbiased is called *biased* and the difference

$$E_\theta(T_N)-\gamma(\theta)$$

is called *bias*.

Examples

Let X be a random variable with finite variance and $X_1,\ldots,X_N$ independent random variables with the same distribution as X. Then

$$T_N: \mathbb{R}^N \to \mathbb{R} \text{ with } T_N(X_1,\ldots,X_N) = \frac{1}{N}\sum_{k=1}^{N} X_k$$

is a consistent, unbiased estimator for E(X).

Let $(X_t)_{t\in I}$ be a stationary process with $E(X_t) = m$. Then

$$\bar{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_{t_k}$$

is an unbiased estimator of m.

The estimator is consistent, iff the spectral distribution function of $(X_t)_{t\in I}$ is continuous in 0.

From now on we will consider only discrete stationary processes $(X_k)_{k\in\mathbb{Z}}$.

Estimation of the autocovariance function $R(r)$, $r \in \mathbb{Z}$

As before, let

$$\bar{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k$$

be an estimator for $E(X_k)$.

Then

$$R_N^*(r) = \frac{1}{N-r} \sum_{k=1}^{N-r} (X_k - \bar{X}_N)(X_{k+r} - \bar{X}_N),$$

and

$$\hat{R}_N(r) = \frac{1}{N} \sum_{k=1}^{N-r} (X_k - \bar{X}_N)(X_{k+r} - \bar{X}_N)$$

are two estimators for $R(r)$.

Both estimators are not unbiased. We have

$$E(R_N^*(r)) = R(r) + \mathrm{Var}(\bar{X}_N) + O(\frac{1}{N^2})$$

$$E(\hat{R}_N(r)) = (1 - \frac{|r|}{N})R(r) + (1 - \frac{|r|}{N})\mathrm{Var}(\bar{X}_N) + O(\frac{1}{N^2}).$$

If the spectral density exists for the process $(X_k)_{k\in\mathbb{Z}}$, i.e.

$$f(\omega) = \sum_{r=-\infty}^{\infty} \rho(r) e^{-i\omega r},$$

then

$$\mathrm{Var}(\bar{X}_N) \underset{N\to\infty}{\sim} R(0) \frac{f(0)}{N}$$

and both estimators are *asymptotically unbiased*, i.e.

$$\lim_{N\to\infty} E(R_N^*(r)) - R(r) = 0 \quad \text{and} \quad \lim_{N\to\infty} E(\hat{R}_N(r)) - R(r) = 0$$

for all $r \in \mathbb{Z}$.

Estimation of the autocorrelation function $\rho(r)$, $r \in \mathbb{Z}$

An estimator of the autocorrelation function $\rho(r)$ is

$$\hat{\rho}_N(r) = \frac{\hat{R}_N(r)}{\hat{R}_N(0)} \qquad r \in \{-(N-1), \ldots, N-1\}.$$

It is

$$|\hat{\rho}_N(r)| \leq 1 \qquad r \in \{-(N-1), \ldots, N-1\}.$$

If $(X_k)_{k\in\mathbb{N}}$ is a Gaussian process with a purely continuous spectrum, we get

$$E(\hat{\rho}_N(r)) \underset{N\to\infty}{\sim} (1 - \frac{|r|}{N})\rho(r)$$

and

$$\mathrm{Var}(\hat{\rho}_N(r)) \underset{N\to\infty}{\sim} \frac{1}{N} \sum_{m=-\infty}^{\infty} \rho^2(m).$$

Identification of ARMA-models

The process of fitting an ARMA-model to given data involves two separate stages, namely

(1) the estimation of the *parameters* of the model,

(2) the determination of the *order* of the model.

For both problems we find a lot of methods and procedures in literature. In some examples we will introduce the basic ideas of the methods.

Example (1)

Estimation of the parameters of an AR(1)-model.

Let $x_1,\ldots,x_N$ be the realizations of the random variables $X_1,\ldots,X_N$ with

$$X_{k+1} = aX_k + \varepsilon_{k+1} \qquad k = 1,2,\ldots,N-1.$$

Our task is to estimate $a \in (-1,1)$.

It is

$$\rho(r) = a^{|r|}, \qquad r \in \mathbb{Z}$$

and especially

$$\rho(1) = a.$$

We get an estimator

$$\hat{a} = \hat{\rho}(1) = \frac{\hat{R}_N(1)}{\hat{R}_N(0)}.$$

Example (2)

Estimation of the parameters of an AR(p)-model.

The above estimation-procedure can be easily generalized. We use the socalled *Yule-Walker-Equations* for the autocorrelation function $\rho(r)$ of an AR(p)-process given by

$$X_{n+1} = a_1X_n + \ldots + a_pX_{n-p+1} + \varepsilon_{n+1}$$

then $P \cdot \mathfrak{A} = \rho$ with

$\mathfrak{A} = (a_1,a_2,\ldots,a_p)$, $\rho = (\rho(1),\rho(2),\ldots,\rho(p))$

and

$$P = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(p-2) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \dots & 1 \end{pmatrix}$$

Replacing $\rho(1),\dots,\rho(p)$ by $\hat{\rho}(1),\dots,\hat{\rho}(p)$ and solving the linear equation system yields estimates $\hat{a}_1,\dots,\hat{a}_p$ for $a_1,\dots,a_p$.

Let us consider again the example of an AR(1)-model given by

$$X_n = aX_{n-1} + \varepsilon_n \quad , \qquad n \in \mathbb{Z} \, ,$$

with $|a| < 1$ and $(\varepsilon_n)_{n\in\mathbb{Z}}$ a sequence of pairwise independent $N(0,\sigma^2)$ distributed random variables. Then $(X_n)_{n\in\mathbb{N}}$ is a Gaussian process, i.e. for all $n_1,\dots,n_k$ is $(X_{n_1},\dots,X_{n_k})$ a k-dimensional normal distribution which parameters are determined by a and σ^2. Conversely, if the parameters of the distribution are known then a can be determined. The estimation of a is therefore equivalent to the estimation of the parameters of a multivariate normal distribution. The joint density of $X_1,\dots,X_N$ is given by

$$f_{X_1,\dots,X_N}(x_1,\dots,x_N) = \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\Big(-\frac{1}{2} \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{\sigma^2} \Big).$$

The probability of an observation $X_i \in \Delta x_i$, with

$$\Delta x_i = (x_i - \tfrac{1}{2}\Delta_i,\ x_i + \tfrac{1}{2}\Delta_i), \quad 0 < \Delta_i << \sigma^2,$$

is approximatively given by

$$P(X_1 \in \Delta x_1,\dots,X_N \in \Delta x_N) \approx \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\Big(-\frac{1}{2} \sum_{i=1}^{N} \frac{(x_i-\mu)^2}{\sigma^2} \Delta_1 \dots \Delta_N \Big).$$

Maximizing $P(X_1 \in \Delta x_1,\dots,X_N \in \Delta x_N)$ with respect to μ and σ^2 means to look for the "most suitable" μ and σ^2.

If we put

$$L(X_1,\ldots,X_N) = \log \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\Big(-\frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^2}\Big)$$

$$= -N \log\sqrt{2\pi} - \frac{N}{2}\log\sigma^2 - \frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^2},$$

then we have to maximize L, i.e.

$$\frac{\partial L}{\partial \mu} = \sum_{i=1}^{N}\frac{x_i-\mu}{\sigma^2} = 0$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^4} = 0.$$

We get

$$\hat{\mu}(x_1,\ldots,x_N) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and

$$\hat{\sigma}^2(x_1,\ldots,x_N) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{\mu})^2.$$

This procedure is called *maximum likelihood estimation.*

In a general case we define the *likelihood function* of a realization $(x_1,\ldots,x_N)$ of a random variable X:

$$L(\theta;x_1,..,x_N) = \begin{cases} \prod_{i=1}^{N} f_\theta(x_i) & X \text{ continuously distributed with density } f_\theta,\ \theta\in\Theta \\ \prod_{i=1}^{N} P_\theta(X=x_i) & X \text{ discrete with distribution } P_\theta,\ \theta\in\Theta. \end{cases}$$

Sometimes $L(\theta;x_1,\ldots,x_N)$ is replaced by $\log L(\theta;x_1,\ldots,x_N)$, the socalled *log-likelihood function.*

A parameter $\hat{\theta} = \hat{\theta}(x_1,\ldots,x_N)$ with

$$L(\hat{\theta};x_1,\ldots,x_N) \geq L(\theta;x_1,\ldots,x_N) \quad \text{for all } \theta\in\Theta$$

is called *maximum likelihood estimator* for $\theta\in\Theta$.

The advantage of the maximum likelihood estimation will be clear from the following results:

(A) If there is a fully efficient estimation T for the parameter $\theta \in \Theta$ (i.e. T is an unbiased estimator with uniformly minimal variance) then this estimation procedure is identical with the maximum likelihood estimation.

(B) If there is no fully efficient estimation for the parameter then the maximum likelihood estimation is after all asymptotically fully efficient.

Example (3)

Let $(X_t)_{t\in\mathbb{Z}}$ be an ARMA(k,0) process with $E(X_t) = \mu \neq 0$ i.e.

$$(X_t-\mu) + a_1(X_{t-1}-\mu) + \ldots + a_k(X_{t-k}-\mu) = \varepsilon_t .$$

The joint distribution of $X_{k+1},\ldots,X_N$, given the observation $X_1 = x_1,\ldots,X_k = x_k$ is then

$$f_{X_{k+1},\ldots,X_N | X_1 = x_1,\ldots,X_k = x_k}(x_1,\ldots,x_N)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{N-k} \exp\left(-\frac{1}{2\sigma^2}\sum_{l=k+1}^{N}(x_l-\mu+a_1(x_{l-1}-\mu)+\ldots+a_k(x_{l-k}-\mu))^2\right).$$

If this function is used as maximum likelihood function for the estimation of $a_1,\ldots,a_k$, we get the same result as in example (2) with the Yale-Walker-Equations.

Determination of the order of the model.

As for the estimation of the parameters of an ARMA-model there are many methods for the determination of the order. A frequently applied method is the socalled *Akaike's information criterion* (AIC). We consider an ARMA(k,l)-model given by

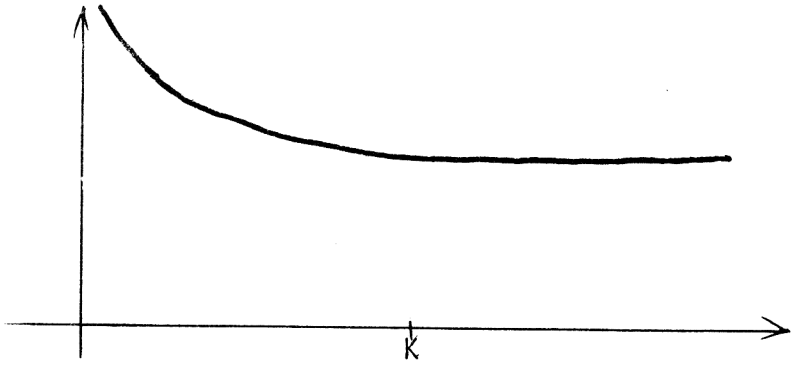
$$X_t + a_1X_{t-1} + \ldots + a_kX_k = \varepsilon_t + b_1\varepsilon_{t-1} + \ldots + b_l\varepsilon_{t-l},$$

$t \in \mathbb{Z}$ and ε_t approximately $N(0,\sigma^2)$-distributed.

Even if we know the parameters $a_1,\ldots,a_k,b_1,\ldots,b_l$ exactly, the variance σ^2 is a measure for the "uncertainty" still being in the observation. If we consider $\sigma^2(\nu)$ as a function of the order ν of the model, the following assumptions are useful:

(1) $\sigma^2(\nu)$ is a monotonic decreasing function left of the true order.

(2) $\sigma^2(\nu)$ is nearly constant right of the true order.



On this considerations bases the AIC-index which is defined for an ARMA(k,l)-model by

$$AIC(k,l) = N\cdot\log \hat{\sigma}^2 + 2(k+l)$$

with N the number of observations and

$$\hat{\sigma}^2 = \frac{1}{N} Q(\hat{a}_1,\ldots,\hat{a}_k,\hat{b}_1,\ldots,\hat{b}_l)$$

and

$$Q(a_1,\ldots,a_k,b_1,\ldots,b_l) = \sum_{\nu=k+1}^{N} \varepsilon_\nu^2 ,$$

where the ε_ν are determined recursively from the difference equation ($\varepsilon_{-(l-k-1)} = \ldots = \varepsilon_{-1} = \varepsilon_o = \varepsilon_1 = \ldots = \varepsilon_k = 0$):

$$\varepsilon_{k+1} = X_{k+1} + a_1 X_k + \ldots + a_k X_1$$

$$\varepsilon_{k+2} = X_{k+2} + a_1 X_{k+1} + \ldots + a_k X_2 - b_1 \varepsilon_{k+1}$$

$$\vdots$$

$$\varepsilon_{k+l} = X_{k+l} + a_1 X_{k+l-1} + \ldots + a_k X_l - b_1 \varepsilon_{k+l-1} - \ldots - b_{l-1} \varepsilon_{k+1}$$

$$\varepsilon_{k+l+1} = X_{k+l+1} + a_1 X_{k+l} + \ldots + a_k X_{l+1} - b_1 \varepsilon_{k+l} - \ldots - b_l \varepsilon_{k+1}$$

$$\vdots$$

etc.

By minimizing the function AIC(k,l) with respect to k and l we get an estimation for the order of an ARMA-model.

Estimation in the frequency domain

The most important characteristica of stationary processes in the frequency domain are the spectral density resp. the spectral distribution function. For the estimation of these functions we have to distinguish the three cases: $(X_t)_{t \in I}$ has a

a) purely discrete spectrum,

b) purely continuous spectrum,

c) mixed spectrum.

a) The general model of a stationary process with a purely discrete spectrum is the harmonic process

$$X_t = \sum_{i=1}^{K} A_i \cos(\omega_i t + \phi_i)$$

where K, A_i, ω_i, $i = 1,\ldots,K$ are constants and the ϕ_i are independent random variables with uniform distribution on $(-\pi,\pi)$.

The spectral distribution function of such a process is given by

$$F(\omega) = \frac{1}{2} \sum_{i=1}^{N} \frac{A_i^2}{A} \left(\delta(\omega-\omega_i) + \delta(\omega+\omega_i)\right)$$

with

$$A = \sum_{i=1}^{N} A_i^2 ,$$

i.e. the spectral distribution function is a step function with jumps at $\omega = \omega_i$, $i = 1,\ldots,K$ and the magnitude of the jumps are given by $\frac{1}{2} A_i^2/A$. Therefore the estimation of the spectral distribution function consists in estimating the location and magnitude of the jumps.

Let us rewrite the process in the form

$$X_t = \sum_{i=1}^{K} A_i \cos(\omega_i t + \phi_i)$$

$$= \sum_{i=1}^{K} (A_i^* \cos \omega_i t + B_i^* \sin \omega_i t)$$

with $A_i^* = A_i \cos \phi_i$, $B_i^* = -A_i \sin \phi_i$

such that

$$A_i = \sqrt{A_i^{*2} + B_i^{*2}} , \quad \phi_i = \tan^{-1}(B_i^*/A_i^*).$$

The basic idea of the famous method called *periodogram analysis* may be explained as follows: Suppose we have an estimate $\hat{\omega}_1$ of ω_1, then we can compute the estimates

$$\hat{A}_1^* = \frac{1}{N} \sum_{j=1}^{N} X_j \cos \hat{\omega}_1 t$$

$$\hat{B}_1^* = \frac{1}{N} \sum_{j=1}^{N} X_j \sin \hat{\omega}_1 t,$$

where $X_1,\ldots,X_N$ are the given observations.

If the guessed value $\hat{\omega}_1$ is close to ω_1 then $\hat{A}_1^*$, $\hat{B}_1^*$ will be close to A_1^*, B_1^* and the squared amplitude $(A_1^{*2} + B_1^{*2})$ will be non-zero. On the other hand, if $\hat{\omega}_1$ is substantially removed from ω_1 (or any of the other frequencies present in the model) then we are estimating the coefficients of a term which does not exist in the model and therefore $(A_1^{*2} + B_1^{*2})$ will be close to zero. If we now choose a sufficiently fine set of trial frequencies $\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3, \ldots$ and plot the squared amplitudes $(\hat{A}_p^{*2} + \hat{B}_p^{*2})$ against $\hat{\omega}_p$, the ordinates will be non-zero if $\hat{\omega}_p$ is close to one of the $\{\omega_i\}$ but will be close to zero otherwise. We have therefore now a method to locate the values of the $\{\omega_i\}$. We will select those whose values are appreciable greater than zero. This is the basic idea of periodogram analysis. In our procedure we will plot the normalized amplitudes $N(A_p^{*2} + B_p^{*2})$ to magnify the differences between large and small ordinates.

Let us now give a formal description of periodogram analysis:

Given N observations $X_1, \ldots, X_N$ the function

$$I_{X,N}(\omega) = A(\omega)^2 + \mathcal{B}(\omega)^2, \qquad \omega \in [-\pi, \pi],$$

with

$$A(\omega) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} X_j \cos \omega j$$

$$B(\omega) = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} X_j \sin \omega j$$

is called *periodogram*.

It is

$$I_{X,N}(\omega) = \frac{1}{N} \left| \sum_{j=1}^{N} X_j e^{-i\omega j} \right|^2 .$$

An important property of the periodogram is given in the following theorem. Remark that the result is valid for <u>all</u> stationary processes:

<u>Theorem</u> (discrete Wiener-Khintchine)

Let $X_1, \ldots, X_N$ be observations of a stationary process $(X_t)_{t \in \mathbb{Z}}$. We assume

$\bar{X}_N = \frac{1}{N} \sum_{k=1}^{N} X_k = 0$. (Otherwise replace X_i by $X_i - \bar{X}_N$.)

Then

$$I_{X,N}(\omega) = \sum_{r=-(N-1)}^{(N-1)} \hat{R}_N(r) \cos r\omega$$

with

$$\hat{R}_N(r) = \frac{1}{N} \sum_{k=1}^{N-r} X_k X_{k+r} .$$

(the estimator for the covariance function, see 4.1)

This means that $I_{X,N}(\omega)$ is nothing else - apart from the factor 2π - than the finite Fourier transform of the autocovariance function.

Let us consider the more realistic model of a harmonic process which includes the errors of observation:

$$X_t = \sum_{i=1}^{K} A_i \cos(\omega_i t + \phi_i) + \varepsilon_t$$

where additively $(\varepsilon_t)_{t\in\mathbb{Z}}$ is white noise, $E(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$ with an unknown parameter σ_ε^2 and (ε_t) independent of the (ϕ_i). In this case we get for the expected value of $I_{X,N}$:

$$E(I_{X,N}(\omega)) = \sigma_\varepsilon^2 + \frac{1}{4N} \sum_{i=1}^{K} A_i^2 \left(\frac{\sin^2(\frac{1}{2}N(\omega+\omega_i))}{\sin^2(\frac{1}{2}(\omega+\omega_i))} + \frac{\sin^2(\frac{1}{2}N(\omega-\omega_i))}{\sin^2(\frac{1}{2}(\omega-\omega_i))} \right)$$

Is ω close to one of the ω_i's, $i = 1,\ldots,n$ then

$$E(I_{X,N}(\omega)) \sim \sigma_\varepsilon^2 + \frac{1}{4} N A_i^2 .$$

Under certain conditions for the process $(X_t)_{t\in\mathbb{Z}}$ we get for the variance of $I_{X,N}$:

$$\mathrm{Var}(I_{X,N}(\omega)) = \mathrm{const}\ \sigma_\varepsilon^4 + O(\frac{1}{N}).$$

In the case of a purely discrete spectrum we have found a satisfying estimator for the spectral distribution function.

b) More complicated is the case of processes with purely continuous spectrum. A general model for these processes is the general linear process $(X_k)_{k\in\mathbb{Z}}$ with

$$X_k = \sum_{l=-\infty}^{\infty} g_l \varepsilon_{k-l} \quad , \qquad (g_l)_{l\in\mathbb{Z}} \in l^2 .$$

(We have generalized the model by extending the summation from $-\infty$ to $+\infty$.)

We further assume that the autocovariance function $R(r)$ is absolutely summable i.e. $\sum_{r=-\infty}^{\infty} |R(r)| < \infty$, then the spectral density function on $h(\omega)$ is continuous for all ω.

Let us consider $E(I_{X,N}(\omega))$ for this case:

$$E(I_{X,N}(\omega)) = \sum_{r=-(N-1)}^{(N-1)} E(\hat{R}_N(r))\cos r\omega$$

$$= \sum_{r=-(N-1)}^{(N-1)} (1 - \frac{|r|}{N})R(r)\cos r\omega$$

and therefore

$$E(I_{X,N}(\omega)) \underset{N\to\infty}{\longrightarrow} 2\pi\, h(\omega)$$

i.e. $I_{X,N}(\omega)$ is an asymptotically unbiased estimate of $h(\omega)$, but for the variance we get

$$\mathrm{Var}(I_{X,N}(\omega)) \to \text{const}\ h^2(\omega)$$

i.e. $I_{X,N}$ is a non consistent estimator and therefore useless for practical purposes.

Although the periodogram is itself an inconsistent estimate, we shall see that the following procedures of constructing consistent estimates of the spectral density function are essentially based on the periodogram by using some sort of smoothing procedure.

Instead of the estimator

$$\hat{h}_{X,N}(\omega) = \frac{1}{2\pi} \sum_{r=-(N-1)}^{(N-1)} \hat{R}_{X,N}(r)\cos r\omega$$

for the spectral density function we now use

$$\tilde{h}_{X,N}(\omega) = \frac{1}{2\pi} \sum_{r=-M}^{M} \hat{R}_{X,N}(r)\cos r\omega ,$$

with $M < N-1$, $M \in \mathbb{N}$, $M \to \infty$ if $N \to \infty$ and $\frac{M}{N} \to 0$,

we may expect

$$\mathrm{Var}(\tilde{h}_{X,N}(\omega)) \sim \frac{M}{N} \mathrm{Var}(\hat{h}_{X,N}(\omega)) = O(\frac{M}{N})$$

The estimate $\tilde{h}_{X,N}$ can be regarded as a special case of the more general form of the estimate

$$\hat{h}_{X,\gamma_N,N}(\omega) = \frac{1}{2\pi} \sum_{r=-(N-1)}^{(N-1)} \gamma_N(r)\hat{R}_{X,N}(r)\cos r\omega$$

with a function $\gamma_N: \mathbb{Z} \to \mathbf{R}_+$, the socalled *window function*. The finite Fourier transform of γ_N

$$W_N(\theta) = \frac{1}{2\pi} \sum_{r=-(N-1)}^{N-1} \gamma_N(r)\cos r\theta$$

is called the *spectral window*.

Usually we assume that γ_N resp. W_N have the following properties:

(a) $\gamma_N(r) = \gamma_N(-r)$ for all $r \in \mathbb{Z}$, $N \in \mathbb{N}$

(b) $W_N(\theta) \geq 0$ for all $\theta \in [-\pi,\pi]$, $N \in \mathbb{N}$

(c) $\int_{-\pi}^{\pi} W_N(\theta)d\theta = 1$ for all $N \in \mathbb{N}$

(d) $\int_{-\pi}^{\pi} W_N^2(\theta)d\theta < \infty$ for all $N \in \mathbb{N}$

(e) For all $\varepsilon > 0$ is $\lim_{N\to\infty} \sup_{|\theta|>\varepsilon} W_N(\theta) = 0$

(f) $\left(\sum_{r=-(N-1)}^{N-1} \frac{|r|}{N} \gamma_N^2(r)\right)\left(\sum_{r=-(N-1)}^{N-1} \gamma_N^2(r)\right)^{-1} \longrightarrow_{N\to\infty} 0.$

Then under general assumptions for the process $(X_k)_{k\in\mathbb{N}}$ one can show that

$$E(\hat{h}_{X,\gamma_N,N}(\omega)) \underset{N\to\infty}{\sim} \int_{-\pi}^{\pi} h_X(\theta)W_N(\omega-\theta)d\theta \underset{N\to\infty}{\longrightarrow} h_X(\omega)$$

$$\mathrm{Var}(\hat{h}_{X,\gamma_N,N}(\omega)) \sim \frac{C_{\gamma_N}}{N} h_X^2(\omega) \qquad \text{for } \omega \neq 0,\ \omega \neq \pm\pi,$$

with

$$C_{\gamma_N} = \sum_{r=-(N-1)}^{(N-1)} \gamma_N^2(r) ,$$

this means that with a suitable window function γ_N $\hat{h}_{X,\gamma_N,N}$ is an asymptotically unbiased and consistent estimator for the spectral density function of processes with purely continuous spectrum.

The bias $b(\omega)$ is given by

$$b(\omega) = E(\hat{h}_{X,\gamma_N,N}(\omega)) - h_X(\omega)$$

Examples:

(In the following $M \in \mathbb{N}$ is chosen as above, i.e. $M < N-1$, $M \to \infty$ if $N \to \infty$ and $\frac{M}{N} \to 0$ if $N \to \infty$.)

RECTANGULAR WINDOW

$$\gamma_N(r) = \begin{cases} 1 & |r| \leq M \\ 0 & |r| > M \end{cases}$$

$$\mathrm{Var}(\hat{h}_{X,\gamma_N,N}(\omega)) \sim \frac{2M}{N} h_X^2(\omega)$$

BARTLETT WINDOW

$$\gamma_N(r) = \begin{cases} 1 - \frac{|r|}{M} & |r| \leq M \\ 0 & |r| > M \end{cases}$$

$$\mathrm{Var}(\hat{h}_{X,\gamma_N,N}(\omega)) \sim \frac{2M}{3N} \cdot h_X^2(\omega)$$

$$b(\omega) \sim -\frac{1}{M} \cdot \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} |r| R_X(r) \cos r\omega$$

DANIELL WINDOW

$$\gamma_N(r) = \frac{\sin(\pi r/M)}{\pi r/M}$$

$$\mathrm{Var}(\hat{h}_{X,\gamma_N,N}(\omega)) \sim \frac{M}{N} h_X^2(\omega)$$

$$b(\omega) \sim \frac{\pi^2}{6M^2} h_X''(\omega)$$

HANNING WINDOW

$$\gamma_N(r) = \begin{cases} \frac{1}{2}(1+\cos\frac{\pi r}{M}) & |r| \leq M \\ 0 & |r| > M \end{cases}$$

$$\mathrm{Var}(\hat{h}_{X,\gamma_N,N}(\omega)) \sim \frac{3M}{4N} h_X^2(\omega)$$

$$b(\omega) \sim \frac{\pi^2}{4M^2} h_X''(\omega)$$

Resolvability and bandwidth

The asymptotic behaviour of the variance and bias of $\hat{h}(\omega)$ depends on the chosen window. We get

$$\mathrm{Var}(\hat{h}(\omega)) = O(\frac{M}{N})$$

and

$$\mathrm{bias}(\hat{h}(\omega)) = b(\omega) = O(\frac{1}{M^a})$$

with the socalled *characteristic exponent* a, which is the largest integer (>0) such that

$$\lim_{x\to o} (\frac{1-k(x)}{|x|^a})$$

exists, is finite and non-zero, with $k(\frac{r}{M}) = \gamma_N(r)$.

Therefore the effect of increasing M is to increase the variance and decrease the bias, while decreasing M decreases the variance and increases the bias.

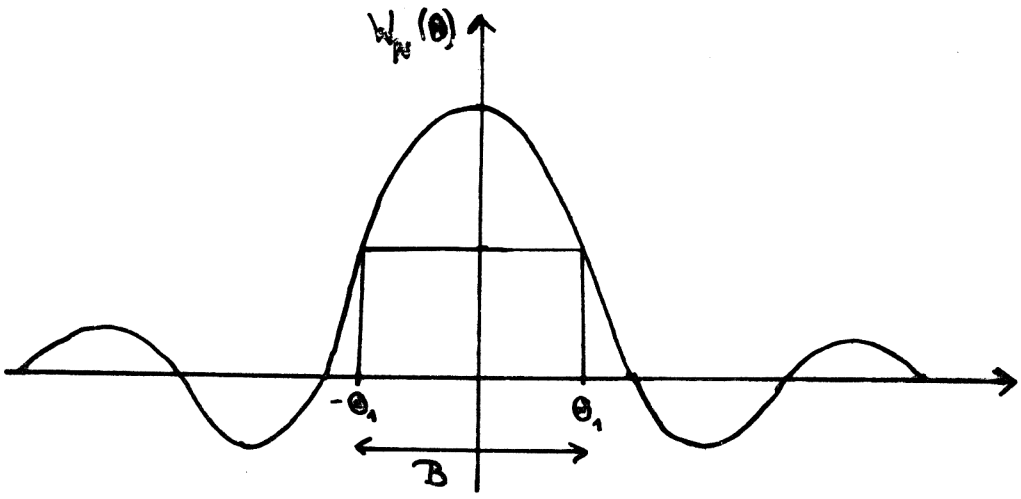
To value the "quality" of a window we have to consider different criteria. One characteristic of a window is the *resolvability*, this means, if the true spectral density function $h(\omega)$ has two distinct peaks at ω_1 and ω_2 then the estimate $\hat{h}(\omega)$ should also have two distinct peaks at ω_1 and ω_2. In order to separate or "resolve" the values of $h(\omega)$ at ω_1 and ω_2 we must choose M sufficiently large so that the "width" of $W_N(\theta)$ is not greater than the distance between ω_1 and ω_2. There is a lot of definitions

for the "width" of a window, the socalled spectral *bandwidth*. We will follow engineering context and define:

If $W_N(\theta)$ is an even function of θ with the maximum at $\theta = 0$, then the *bandwidth* B is given by

$$B = 2\theta_1$$

with θ_1 such that $W_N(\pm\theta_1) = \frac{1}{2} W_N(0)$.



There is now the important result that the product of the variance and the bandwidth is constant.

$$\mathrm{Var}(\hat{h}(\omega)) \cdot B = \text{constant}.$$

This means, if we make M large to get a small bandwidth and therefore a good resolvability we must accept a large variance and vice versa.

TIME SERIES ANALYSIS

A time series is a collection of observations made sequently in time. Examples occur in a variety of fields, ranging from economics to engineering. I will start with some examples.

Economic time series are f.e.

- share prices on successive days,
- export totals in successive months,
- average incomes in successive months,
- company profits in successive years.

Physical time series f.e.

in meteorology, marine science, geophysics,

- rainfall on successive days,
- air temperature in successive hours, days or months.

Marketing time series are f.e.

- sales figures in successive weeks or months.

Demographic time series.

There are several possible objectives in analysing a time series. These objectives may be classified as

- description

  (The first step in the analysis is usually to plot the data and to obtain simple descriptive measures of the main properties of the series, f.e. investigation of the trend, of seasonal fluctuations, of the autocorrelation.)

- explanation

  (f.e. when observations are taken on two or more variables, may be possible to use the variation in one time series to explain the variation in another series. This may lead to a deeper understanding of the mechanism which generated a given time series.)

- prediction (or forecasting)

  Given an observed time series, one may want to predict the future values of the series.

- control

  When a time series is generated which measures the quality of a process (f.e. a manufacturing process), the aim of analysis may be to control the process.

Sources of variation of a time series are f.e.

- seasonal effects

  (Many time series, such as sales figures and temperature readings, exhibit a variation which is annual in period.)

- other cyclic changes

  (Apart from seasonal effects, some time series exhibit variation at a fixed period due to some other physical cause. An example is the daily variation in temperature.)

- trend

  (This may be loosely defined as long term change in the mean.)

- other irregular fluctuations

  (After trend and cyclic variations have been removed from a set of data, we are left with a series of residuals, which may or may not be random.

In practice the following model of a time series is often used:

$$X_t = \mu_t + S_t + Y_t \qquad t = 1,\ldots,N$$

X_t observation at time t

μ_t trend (time dependent mean of X_t)

S_t seasonal (or other cyclic) fluctuation

Y_t residual, i.e. a zero mean stationary process

The analysis of a time series depends on whether one wants to measure the trend or remove it, to measure the seasonal components or remove them. We will consider some analysing techniques.

A. Estimation of the trend

Let us suppose that we can express μ_t as a linear combination of known functions $\phi_1(t),\ldots,\phi_q(t)$

$$\mu_t = \Theta_1\phi_1(t) + \ldots + \Theta_q\phi_q(t)$$

with unknown parameters $\Theta_1,\ldots,\Theta_q$

(This method is known as 'regression' analysis.)

The task is then to estimate the parameters $\Theta_1,\ldots,\Theta_n$. This can be done by the 'least squares method' i.a. to minimize the function

$$F(\Theta_1,\ldots,\Theta_q) = \sum_{t=1}^{N} (X_t-\mu_t)^2$$

or by the 'maximum likelihood' approach. (More about the last one in a later section.)

A case of considerable interest is the case that the trend μ_t is a polynomial of degree q-1, i.e.

$$\mu_t = \Theta_1 + \Theta_2 t + \ldots + \Theta_q t^{q-1}.$$

For q = 2 we speak of linear regression.

B. Removing the trend

We can sometimes remove trends by passing the observations $\{X_t\}$ through a suitable linear filter.

If μ_t is a (q-1)th degree polynomial in t

$$\mu_t = \theta_1 + \theta_2 t + \ldots + \theta_q t^{q-1},$$

then it is well known that the q-th differences of μ_t will be zero. Let us consider for a while that

$$X_t = \mu_t + Y_t \qquad \text{(i.e. there are non cyclic components in the series).}$$

We consider $\nabla^q X_t$ (where $\nabla = (1-B)$ is the standard difference operator

$$\text{i.e. } \nabla X_t = (1-B)X_t = X_t - X_{t-1}.$$

We get

$$X'_t := \nabla^q X_t = \nabla^q \mu_t + \nabla^q Y_t = \nabla^q Y_t.$$

The spectral density functions of $\nabla^q X_t$ and Y_t are related by

$$h_{X'}(\omega) = |1-\exp(-i\omega)|^{2q} h_y(\omega) = 2^{2q}(\sin^2(\tfrac{\omega}{2}))^q h_y(\omega).$$

Hence, we may estimate $h_y(\omega)$ by first estimate $\hat{h}_{X'}(\omega)$ and then compute

$$\hat{h}_y(\omega) = 2^{-2q}(\sin^2(\tfrac{\omega}{2}))^{-q} \cdot \hat{h}_{X'}(\omega) \qquad \omega \neq 0.$$

If there are cyclic components in the time series, we need methods to remove them.

C. Removing cyclic components

Let us assume we have still removed the trend and consider

$$X_t = S_t + Y_t.$$

The cyclic component S_t may be removed by operating on X_t with one of the following filters:

- Moving average filter:

  Let S_t be periodic with period P.

  If P is odd, say $P = 2r+1$, we may remove S_t by operating on X_t with the filter

$$X'_t := \frac{1}{2r+1}(X_{t-r} + \ldots + X_t + \ldots + X_{t+r}) = \phi(B)X_t$$

i.e. X'_t is formed by taking a moving average of X_t over $(2r+1) = P$ points. Since S_t is periodic with period P the filter $\phi(B)$ will clearly reduce S_t to constant value, we may write

$$X'_t = \text{const.} + \phi(B)Y_t.$$

The spectral density function of Y_t can now be estimated as described above.

If P is even, we cannot construct a 'symmetric' filter of the above form and the procedure is modified slightly. $P = 2r$, we define

$$X'_t = \frac{1}{2r}\left(\frac{1}{2}X_{t-r} + X_{t-r+1} + \ldots + X_{t+r-1} + \frac{1}{2}X_{t+r}\right).$$

Note that in applying these filters we lose r points at the beginning and at the end. If we start with N values of X_t we can compute only (N-2r) values of X'_t.

- p-step difference filter

Another way of removing S_t is to use the filter $(1-B^p)$,

$$X'_t = (1-B^p)X_t = X_t - X_{t-p}.$$

It will remove any component in X_t which is periodic with period p

$$X'_t = (1-B^p)Y_t = Y_t - Y_{t-s}.$$

A good and short introduction in time series analysis you will find in [11].

IDENTIFICATION OF LINEAR SYSTEMS - SOME METHODS

We will introduce some methods for the identification of linear timeinvariant systems. We consider the following single input/single output system

$$y_t + \sum_{i=1}^{p} a_i y_{t-i} = \sum_{i=1}^{p} b_i u_{t-i} + \xi_t \ .$$

We assume that we have N observations:

$$\left.\begin{array}{ll} u_t & \text{input} \\ y_t & \text{output} \\ \xi_t & \text{residuals} \end{array}\right\} \quad t = 1,\ldots,N$$

The problem is to identify the parameters $a_1,\ldots,a_p;b_1,\ldots,b_p$.

One method to solve this identification problem is the 'least squares' approach.

We will introduce first the basic ideas of this method. The method of least squares is, perhaps, one of the oldest estimation procedures and was first developed independently by Gauss and Legendre in the early 19th century. The main idea of the least squares approach can be motivated as follows: We assume that $(Y_t)_{t\in I}$, $I = \{1,\ldots,N\}$ is a stochastic process for which the mean value is a linear function of a parameter vector Θ, i.e.

$$E[Y_t] = x_t^T\Theta,$$

with a known vector x_t, $t = 1,\ldots,N$. We want to get a good estimate in the least squares sense of Θ from a realization of the stochastic process (Y_t). This means we seek a value of $\hat{\Theta}$ which minimizes

$$S = \sum_{t=1}^{N} (y_t - x_t^T\Theta)^2.$$

This equation can be written in the form

$$S = (Y - X\Theta)^T(Y-X\Theta)$$

with

$$Y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{pmatrix} \qquad X = \begin{pmatrix} x_1^T \\ \cdot \\ \cdot \\ \cdot \\ x_N^T \end{pmatrix} .$$

Differentiating with respect to Θ shows that the value $\hat{\Theta}$ minimizes S if

$$(X^TX)\hat{\Theta} = X^TY .$$

If $X^T \cdot X$ is invertible, then there is a unique solution which can be expressed as

$$\hat{\Theta} = (X^TX)^{-1}X^TY.$$

This equation is often called the 'least squares estimator'. For the case X^TX is singular the above equation does not have a unique solution and there is a family of least squares estimates which may be determined in any particular case by the usual methods for solving linear equations.

We will now use this approach to identify our system above:

$$y_t + \sum_{i=1}^{p} a_i y_{t-i} = \sum_{i=1}^{p} b_i u_{t-i} + \xi_t$$

The equation can be expressed in the form

$$y_t = x_t^T\Theta + \xi_t \qquad t = 1,\ldots,N$$

with

$$\Theta^T = (a_1,\ldots,a_p;b_1,\ldots,b_p)$$

$$x_t^T = (-y_{t-1},\ldots,-y_{t-p};u_{t-1},\ldots,u_{t-p})$$

or it can be expressed in matrix form

$$Y = X\Theta + \Sigma$$

with

$$Y^T = [y_1,\ldots,y_N]$$

$$X^T = [x_1,\ldots,x_N]$$

$$\Sigma = [\xi_1,\ldots,\xi_N]$$

By the least squares approach we get an estimation for Θ:

$$\hat{\Theta} = (X^T X)^{-1} X^T Y$$

(Remark that the estimate is <u>not</u> linear in the y_t's.)

If (ξ_t) are uncorrelated random variables it can be shown the $\hat{\Theta}$ is a consistent estimator, this means that $\hat{\Theta}$ is an asymptotically unbiased estimator and that the variance of the estimator converges to the variance of Θ in the mean squares sense.
(Under suitable conditions $\hat{\Theta}$ is even a strongly consistent estimator, [10] p. 103 f.)

If the (ξ_t) are correlated and we know the autocorrelation function, the so-called *generalized least squares* approach can be used. Under these conditions for (ξ_t) it is possible to model ξ_t as follows

$$\xi_t = \sum_{k=0}^{n} g_k \varepsilon_{t-k} = G(z)\varepsilon_t.$$

This result is a special case of a more general situation, the socalled *spectral factorization*. One can show that under quite general assumptions any stationary process (X_t) with a purely continuous spectral density can be represented as a linear combination of the terms of an uncorrelated process (ε_t)

$$X_t = \sum_{u=0}^{\infty} g_u \varepsilon_{t-u}.$$

The coefficients g_u can be directly determinated by the spectral density. See f.e. [9], p. 730 ff, [10], p. 72.

Our system can now be expressed in the following form:

$$A(z)y_t = B(z)u_t + G(z)\varepsilon_t$$

or

$$A(z)y_t^* = B(z)u_t^* + \varepsilon_t$$

with

$$y_t^* = G^{-1}(z)y_t$$
$$u_t^* = G^{-1}(z)u_t.$$

We are now in the situation to apply the above least squares procedure and get estimates of the parameters $a_1,\ldots,a_p,b_1,\ldots,b_p$.

If Θ does not depend on the data, this means if we consider the system

$$y_t = \sum_{i=1}^{p} b_i u_{t-i} + \xi_t,$$

ξ_t uncorrelated, then $\hat{\Theta}$ is a *linear* estimator and it can be shown that it is a socalled BLUE-estimator, that means the best linear unbiased estimator.

By far the most general and most powerful method of estimation is the *maximum likelihood* approach. This method can be used to any type of estimation problem provided only that we can write down the joint probability density function of the observations.

Let us assume we have a sample $X_1,\ldots,X_n$ of n independent random variables from a distribution with density function $p(x|\Theta)$ with an unknown parameter Θ. The joint probability density function of $(X_1,\ldots,X_n)$ may be written as

$$p(x|\Theta) = p(x_1|\Theta)p(x_2|\Theta)\ldots p(x_n|\Theta).$$

If

$$p(x^1|\Theta) > p(x^2|\Theta)$$

where x^1, x^2 are two realizations of $(X_1,\ldots,X_n)$, then we may say that x^1 is *more likely* than x^2.

If we consider $p(x|\Theta)$ as a function of Θ (with x fixed), then we call it the *likelihood function* of Θ.

If

$$p(x|\Theta_1) > p(x|\Theta_2)$$

then we may say that Θ_1 is a *more plausible* value of Θ than Θ_2. The idea in

the maximum likelihood approach is now to estimate Θ by its most plausible value, this means to minimize the likelihood function $p(x|\Theta)$ with respect to Θ. In many cases it is more convenient to work with the function

$$L(x|\Theta) = \log(p(x|\Theta),$$

the socalled *log-likelihood-function*. We may equally well determine $\hat{\Theta}$ by maximizing $L(x|\Theta)$ rather than by maximizing $p(x|\Theta)$.

The least squares approach considered above works in the case of models which are linear in the parameters. The maximum likelihood approach works in a far more general class of problems.

For simplicity we will show how maximum likelihood works at the above problem. We want to identify as before the parameters of the system given by

$$y_t + \sum_{i=1}^{p} a_i y_{t-i} = \sum_{i=1}^{p} b_i u_{t-i} + \xi_t, \quad t = 1,\ldots,N.$$

We assume that ξ_t are independent random variables with a Gauss-distribution with zero mean and variance σ^2. Let us first assume σ^2 is known. We have to estimate $\Theta = (a_1,\ldots,a_p,b_1,\ldots,b_p)$.

The likelihood-function is given by

$$p(y_1,\ldots,y_N|u_1,\ldots,u_N;\Theta) = \prod_{t=1}^{N} p(y_t|y_1,\ldots,y_{t-1};u_1,\ldots,u_t;\Theta) \quad \textit{(Bayes' rule)}.$$

Let $w_t(\Theta) = y_t + \sum_{i=1}^{p} a_i y_{t-i} - \sum_{i=1}^{p} b_i u_{t-i}$.

(This means $w_t(\Theta)$ is an observation of ξ_t for a particular value of Θ.)

Rule for transformation of density functions gives us

$$p(y_t|y_1,\ldots,y_{t-1};u_1,\ldots,u_t;\Theta) = p_{\xi_t}(w_t(\Theta)|\Theta)\cdot\left|\det\left(\frac{\partial \xi_t}{\partial y_t}\right)\right| = p_{\xi_t}(w_t(\Theta|\Theta),$$

since $\frac{\partial \xi_t}{\partial y_t} = 1$.

The likelihood function is then given by

$$p(y_1,\ldots,y_N|u_1,\ldots,u_N;\Theta) = \prod_{t=1}^{N} p_{\xi_t}(w_t(\Theta)|\Theta) \qquad (*).$$

Since we have assumed that (ξ_t) has a Gauss-distribution $(0,\sigma^2)$, we get

$$p_{\xi_t}(w_t(\Theta)|\Theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{w_t^2(\Theta)}{2\sigma^2}\right)$$

and therefore

$$p(y_1,\ldots,y_N|u_1,\ldots,u_N;\Theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \sum_{t=1}^{N} \exp\left(-\frac{w_t^2(\Theta)}{2\sigma^2}\right).$$

For the log-likelihood-function we get

$$\begin{aligned} L(\Theta) &= \log p(y_1,\ldots,y_N|u_1,\ldots,u_N;\Theta) \\ &= -N \log \sigma - \frac{N}{2} \log 2\pi - \sum_{t=1}^{N} \frac{w_t^2(\Theta)}{2\sigma^2} . \end{aligned}$$

If we do not know σ^2, we have to estimate σ^2 too and consider the above log-likelihood-function as a function of Θ and σ^2.

Differentiating $L(\Theta,\sigma)$ with respect to σ yields:

$$\frac{\partial L(\Theta,\sigma)}{\partial \sigma} = -N \cdot \frac{1}{\sigma} + \frac{1}{\sigma^3} \sum_{t=1}^{N} w_t^2(\Theta)$$

$$\frac{\partial L}{\partial \sigma} = 0 \quad \Rightarrow \quad \sigma^2 = \frac{1}{N} \sum_{t=1}^{N} w_t^2(\Theta) \qquad (**)$$

We see that for any fixed Θ equation (**) maximizes the log-likelihood-function with respect to σ, therefore we have first to maximize $L(\Theta,\sigma)$ with respect to Θ and then compute σ^2 from (**)

$$\begin{aligned} \frac{\partial L(\Theta,\sigma)}{\partial \Theta} &= \left(\frac{\partial L}{\partial a_1}, \ldots, \frac{\partial L}{\partial b_1}, \ldots, \frac{\partial L}{\partial b_p}\right) \\ &= \frac{1}{\sigma^2} \sum_{t=1}^{N} w_t(\Theta)\left(\frac{\partial w_t(\Theta)}{\partial a_1}, \ldots, \frac{\partial w_t(\Theta)}{\partial a_p}, \frac{\partial w_t(\Theta)}{\partial b_1}, \ldots, \frac{\partial w_t(\Theta)}{\partial b_p}\right) \\ &= \frac{1}{\sigma^2} \sum_{t=1}^{N} w_t(\Theta)(y_{t-1},\ldots,y_{t-p},-u_{t-1},\ldots,-u_{t-p}) \\ &= \frac{1}{\sigma^2} \sum_{t=1}^{N} (y_t - x_t^T(\Theta)(-x_t^T) \end{aligned}$$

with

$$w_t(\Theta) = y_t - x_t^T\Theta \quad \text{and} \quad -x_t^T := (y_{t-1},\ldots,y_{t-p},-u_{t-1},\ldots,-u_{t-p}).$$

Using some notations for X, Y as in the least squares case we get

$$\frac{\partial L}{\partial \Theta} = 0 \quad \Rightarrow \quad 0 = X^T Y - (X^T X)\hat{\Theta}$$

$$\Rightarrow \quad \hat{\Theta} = (X^T X)^{-1}(X^T Y) \quad \text{(unique if } X^T X \text{ nonsingular)}$$

$$\hat{\sigma^2} = \frac{1}{N} \sum_{t=1}^{N} w_t^2(\hat{\Theta}) = (\frac{1}{N}(Y - X\hat{\Theta})^T (Y - X\Theta)$$

Remark: In the case the $\{\xi_t\}$ are Gauss-variables we get the same result for $\hat{\Theta}$ as for the least squares procedure.

If the (ξ_t) have another distribution than a Gauss distribution we can compute the likelihood function with equation (*).

DATA REDUCTION AS A SPECIAL PRINCIPLE IN FATIGUE ANLYSIS

It is essential that critical safety components on a passenger car, for example suspension components, do not fail during the service life of the vehicle. On the other hand suspension components are designed for a finite life, i.e. they may be subjected to infrequent peak stresses which they will not withstand if repeated indefinitely.

Before the individual components can be fatigue tested in the laboratory for example with servo-hydraulic actuators, it is first necessary to define the loading for which they are to be designed. This loading is called the "reference loading" which must be available as a time function over a long distance (for instance 300.000 km), in order to feed the servo-hydraulic actuators with the corresponding signals. Natuaraly these "reference loading" cannot be measured as a path of a stochastic process over 300.000 km, so annother approach is necessary.

The method used in practise is to take short component loading measurements with strain gauges on public roads over short distances about 100 km supplemented with measurements of extreme maneuvers (such as sharp acceleration and braking) and driving on extremely rough roads on factory proving grounds. These stochastic load time functions are subjected to certain data reduction principles such as Markov counting and Rainflow counting methods. By extrapolation and superposition of the given individual measurements one obtains the reference data. The required load time function can then be determined by using an on-line reconstruction principle.

Here we discribe two of the mostly used data reduction principles which are the Markov counting and the Rainflow counting method. Afterwards we give some insides in the ideas used for the on-line reconstruction principles. For both data reduction principles it is necessary to divide the measurement

range into a finite number of classes (Fig. 1). For simplicity we assume that these classes are given by the numbers 1,2,...,n. Such a discrete process is reduced to a time function which has only peaks and troughs. This means that all values which are not local maxima or minima are crossed out (Fig. 1).

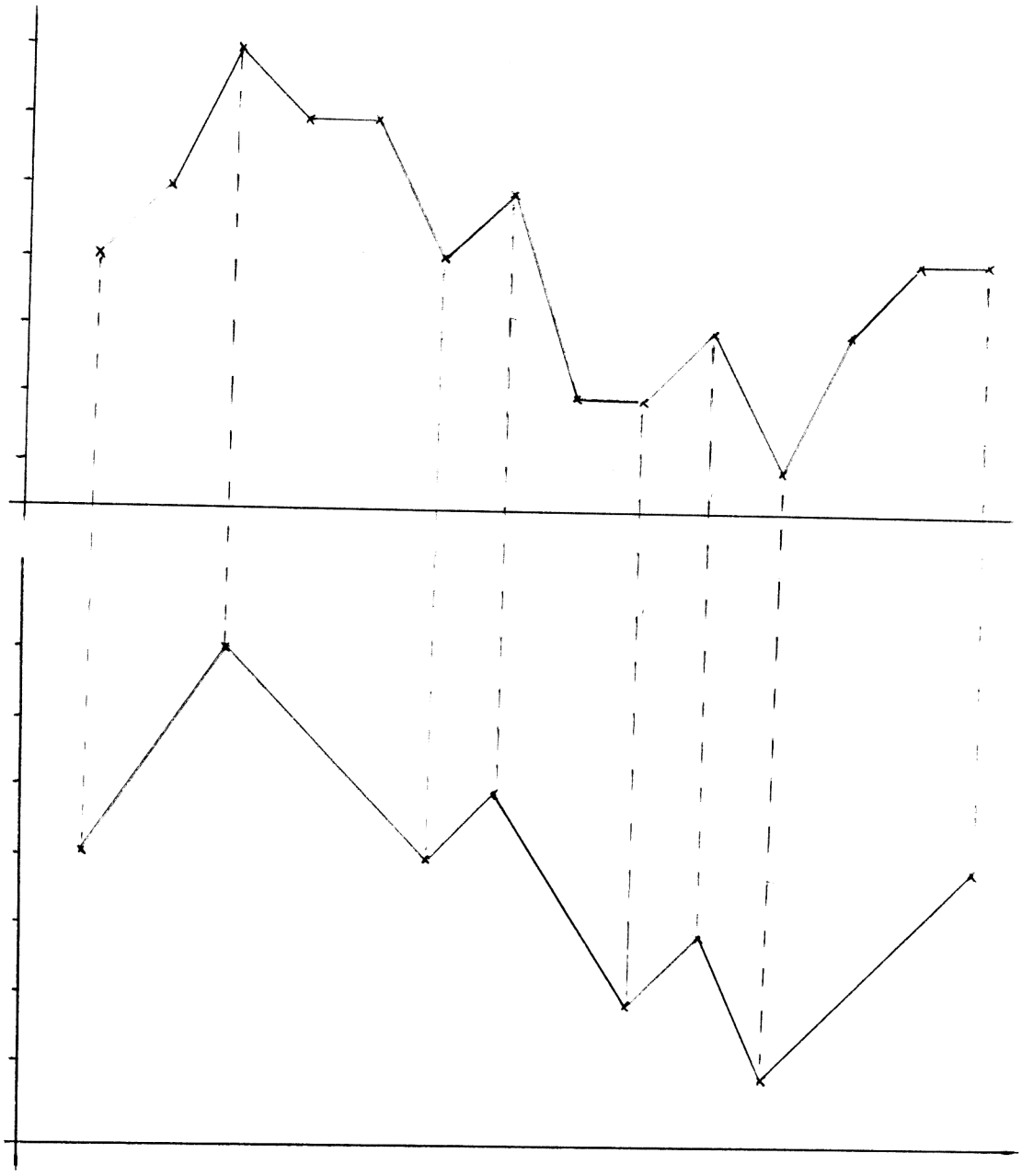
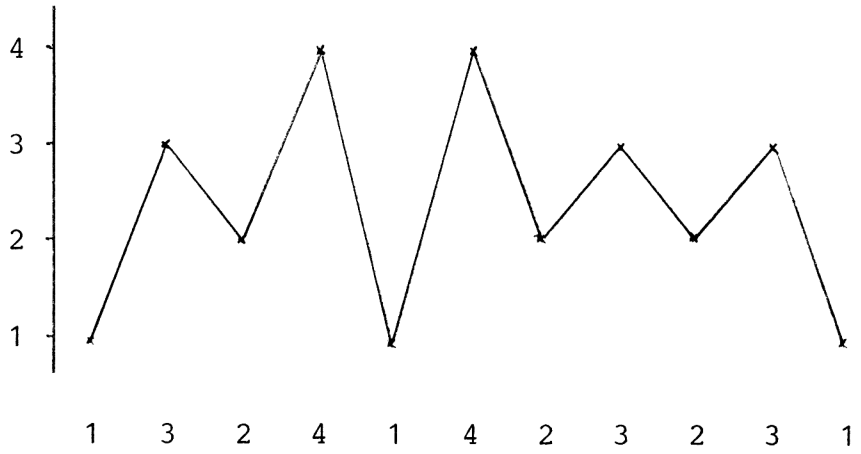


Fig. 1 : Reduction of the load time function to a peak/trought sequence.

Markov counting method

The Markov counting method counts all load variations from level "i" to level "j". This counting result is stored in the so-called Markov counting matrix (Fig. 2).



Markov counting matrix A :

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

a_{ij} : number of load variations from level "i" to level "j"

Fig. 2 : Markov counting

There is some structure on the Markov matrices which is easy to prove

For $i,j \in \{1,2,\ldots,n\}$ define

$$R_{1i} = \sum_{j=i}^{n} a_{ij}$$

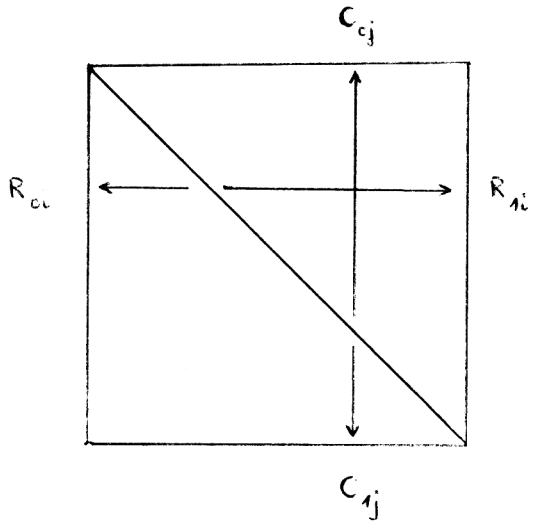
$$R_{oi} = \sum_{j=1}^{i} a_{ij}$$

(all transitions from level "i")

$$C_{1j} = \sum_{i=j}^{n} a_{ij}$$

$$C_{oj} = \sum_{i=1}^{j} a_{ij}$$

(all transitions into level "j")



If the level "i" is neither a starting point nor an end point then we have that

$$C_{1i} = R_{1i} \quad \text{and} \quad C_{oi} = R_{oi} \; .$$

So it is easy to see that not every n×n - Matrix A from $\mathbb{N}_o^{n\times n}$ is a Markov matrix. But there are some good principles which transform any given matrix $A \in \mathbb{N}_o^{n\times n}$ into an admissible Markov matrix. Therefore we may assume that we always have an admissible Markov matrix. To explain the reconstruction principle for Markov

matrices let us consider the following example (Fig. 3)

$$A = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

1 3 1 4 1 ⟶ STOP

1 3 1 4 2 4 1 ⟶ STOP

1 3 1 4 2 3 2 4 1 ⟶ STOP

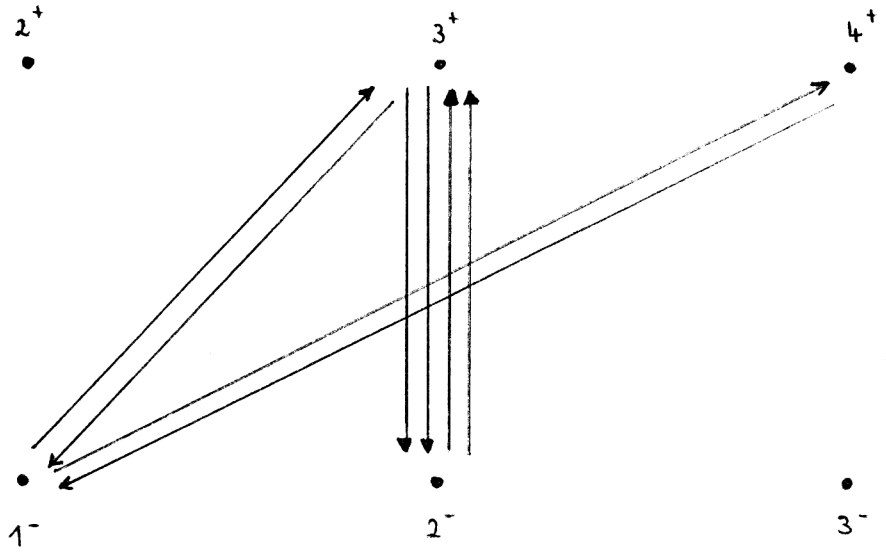
1	3	1	4	2	3	2	3	2	4	1

Fig. 3 : Reconstruction from the Markov matrix

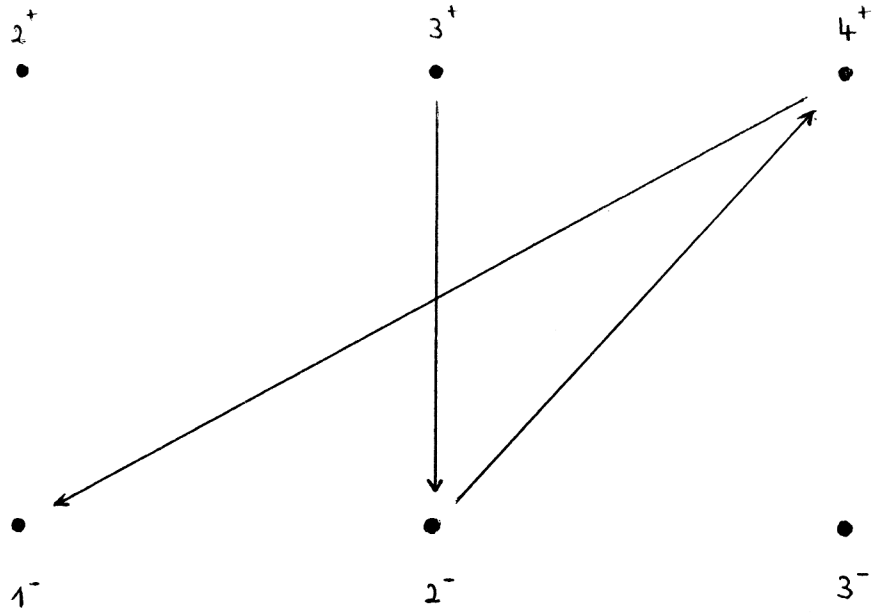
We see that it is not easy to find one reconstruction from the given Markov matrix. Our aim is to construct an on-line reconstruction algorithm which chooses among all load time functions which belong to the given Markov matrix one time function with equal probability. To solve this problem it is better to use the language of graph theory. Every Markov matrix is equivalent to a certain graph. This relation is very simple. The knots are given by the discrete stress levels distinguished whether one is reaching or leaving a given class with an upward or downward transition. The edges are given by the transitions of the Markov matrix. In this language our problem of reconstruction is the same as to find an Eulerian cycle for the given graph. This can be done by constructing a spanning tree. This

is the subgraph of the given graph with the property that from every knot there is a unique way into a fixed endpoint.

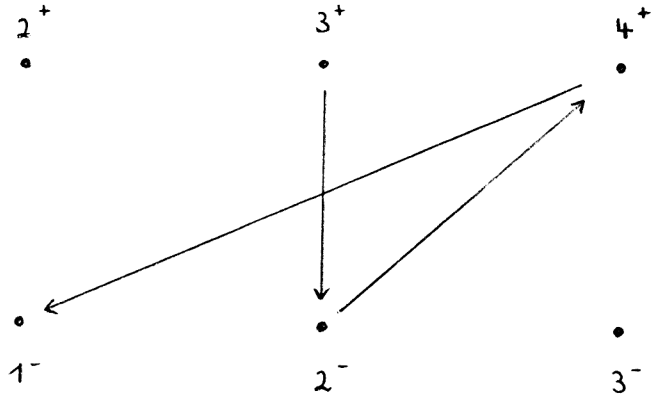
Graph which is equivalent to the above matrix A:



A spanning tree at this graph:



If we know a spanning tree it is very easy to find an Eulerian cycle. We subtract the spanning tree from the given graph and then we may choose any way from the starting point to one of the next possible knots. Whenever there is no further edge we choose the unique edge given by the spanning tree. This is demonstrated in the following example (Fig. 3).



$$V = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

$$A-V = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

reconstruction:

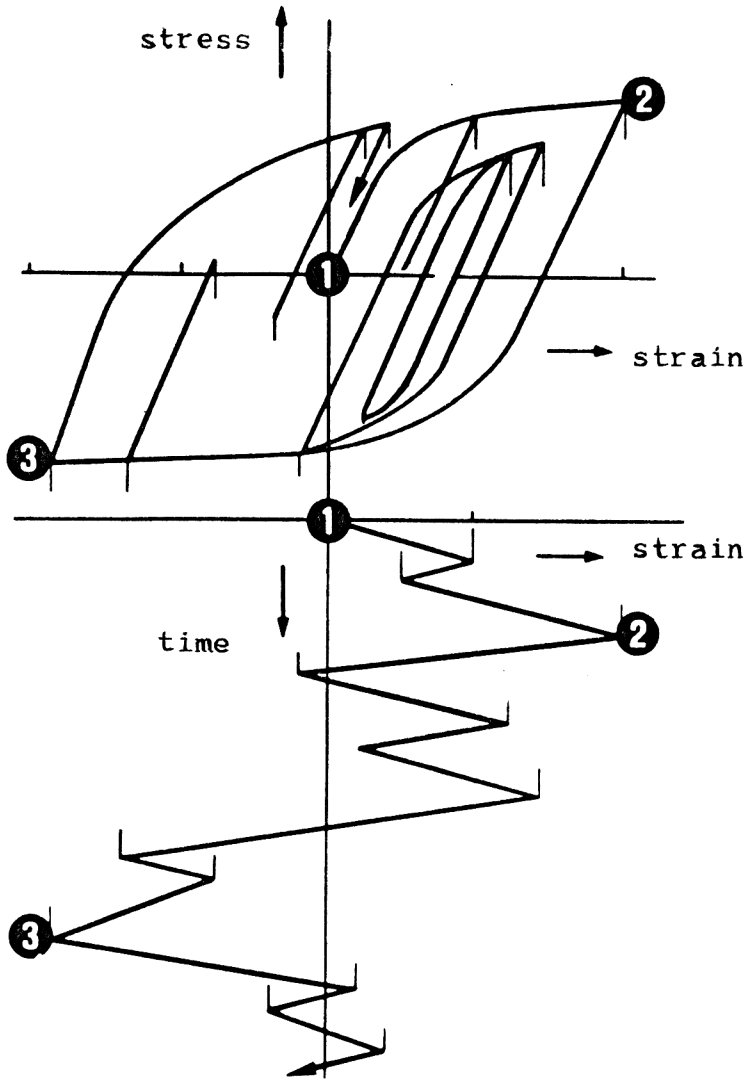
1 3 2 3 1 4 2 3 2 4 1

Fig. 3: How to use a spanning tree.

Rainflow counting method

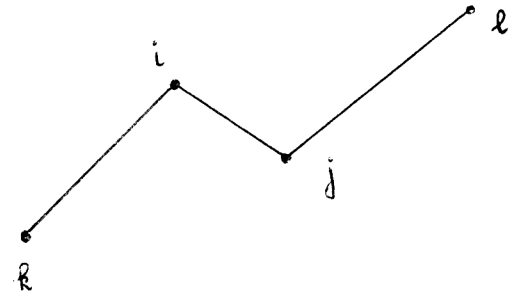
The Rainflow counting method counts all closed hysterisis loops in the stress/strain diagram (Fig. 4). These can be seen as intermediate pulse in the load time function. The residuum left when all the intermediate pulses are eliminated can be seen in Fig. 4 as the main points 1, 2 and 3.

Fig. 4: Hysterisis loop



The counting result is stored in the so-called Rainflow matrix A and the Residuum R. The elements a_{ij} of A are the number of closed hysterisis loops from level "i" to level "j". These loops are also called Rainflow cycles and may be defined as follows:

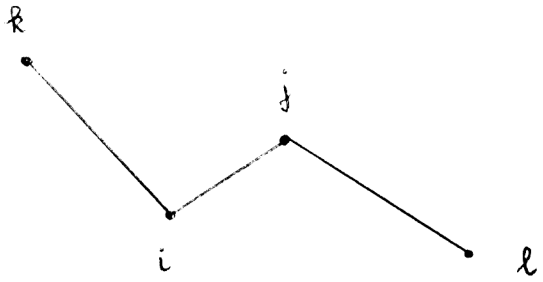
Rainflow cycle from "i" to "j":



$i > j$

$j \geq k$

$\ell \geq i$



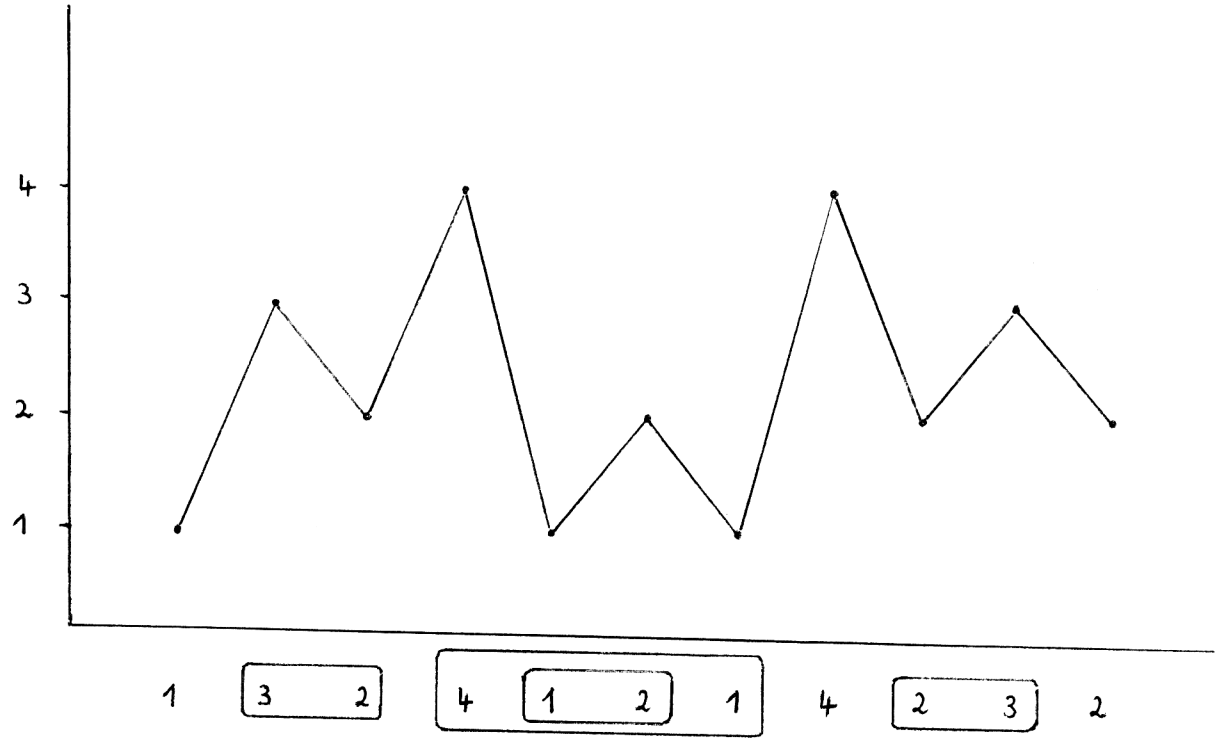
$i < j$

$j \leqq k$

$\ell \leqq i$

The result of the Rainflow counting method is shown in the following example (Fig. 5).

Fig. 5: Rainflow counting



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$R = (1, 4, 2)$$

The on-line reconstruction from the Rainflow matrix is much more complicated as the reconstruction from the Markov matrix.

It would be to exhaustive to explain it here. But all these methods are described in the references [12], [13].

References

[1] Zadeh, L.A.: From circuit theory to system theory. Proc. IRE 50, 856-865 (1962)

[2] Guidorzi, R.P.: Invariants and Canonical Forms for Systems Structural and Parametric Identification. Automatica, Vol. 17, No. 1, pp. 117-133 (1981)

[3] Mehra, R.K.: Identification in Control and Econometrics; Similarities and Differences, Annals of Social and Economic Measurement, Jan. 1974

[4] Kalman, R.E.; Falb, P.L.; Arbib, M.A.: Topics in Mathematical System Theory. McGraw-Hill, New York (1969)

[5] Willems, J.C.: From Time Series to Linear Systems, Part I and II, Mathematics Institute, University of Groningen, Groningen 1985

[6] Kalman, R.E.: Mathematical Description of Linear Dynamical Systems. SIAM J. Control 1, pp. 152-192 (1963)

[7] Hazewinkel, M.: On Identification and the Geometry of the Space of Linear Systems, Proc. Bonn, Jan. 1979, Workshop on Stochastic Control and Stochastic Differential Equations, Lect. Notes Control and Information Sciences, 16, Springer, 1979, pp. 401-415

[8] Akaike, H.: A New Look at the Statistical Model Identification, IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, Dec. 1974

[9] Priestley, M.B.: Spectral Analysis and Time Series, London, Acad. Press, 1981, Vol. 1+2

[10] Goodwin, G.C.; Payne, R.L.: Dynamic Systems Identification Experiment Design and Data Analysis, Mathematics in Science and Engineering, Vol. 136, Acad. Press 1977

[11] Chatfield, C.: The Analysis of Time Series: An Introduction, London, New York 1980

[12] Krüger, W.; Scheutzow, M.; Beste, A.; Petersen, J.: Markov- und Rainflow-Rekonstruktionen stochastischer Beanspruchungszeitfunktionen. VDI-Fortschrittsbericht, Reihe 18, Nr. 22, 1985

[13] de Jonge, J.B.: Counting Methods for the Analysis of Load-Time-Histories. National Aerospace Laboratory, NLR Memo. SB 80/106/U, Holland, 1980

BISHER SIND FOLGENDE BERICHTE ERSCHIENEN:

Nr. 1 FORSCHUNG

W.G. Eschmann und Ralph Götz
Optimierung von Gelenksechsecken

Nr. 2 WEITERBILDUNG

H. Neunzert, M. Schulz-Reese
Mathematische Weiterbildung

Nr. 3 FORSCHUNG

W. Krüger
The Trippstadt Problem

Nr. 4 WEITERBILDUNG

H. Neunzert, M. Schulz-Reese, K.E. Hoffmann
Mathematics in the University and Mathematics in Industry - Complement or Contrast?

Nr. 5 FORSCHUNG

A.K. Louis
The Limited Angle Problem in Computerized Tomography

Nr. 6 FORSCHUNG

W. Krüger
Regression für Ellipsen in achsenparalleler Lage
(Interner Bericht)

Nr. 7 FORSCHUNG

Th. Mietzner
Umströmung von Ecken und Kanten

Nr. 8 FORSCHUNG

W. Krüger, J. Petersen
Simulation und Extrapolation von Rainflow-Matrizen

Nr. 9 FORSCHUNG

W. Krüger, M. Scheutzow u. A. Beste, J. Petersen
Markov- und Rainflow-Rekonstruktionen stochastischer Beanspruchungszeitfunktionen

Nr. 10 FORSCHUNG

Th. Mietzner
Umströmung von Ecken und Kanten, Teil 2

Nr. 11 FORSCHUNG

H. Ploss
Simulationsmethoden zur Lösung der Boltzmanngleichung

Nr. 12 FORSCHUNG

M. Keul
Mathematische Modelle für das Zeitverhalten stochastischer Beanspruchungszeitfunktionen

Die Berichte der Arbeitsgruppe Technomathematik
können angefordert werden bei:

UNIVERSITÄT KAISERSLAUTERN
Fachbereich Mathematik
Arbeitsgruppe Technomathematik
Erwin-Schrödinger-Straße

6750 Kaiserslautern

ARBEITSGRUPPE TECHNOMATHEMATIK AM FACHBEREICH MATHEMATIK DER UNIVERSITÄT KAISERSLAUTERN

Leiter: Prof. Dr. H. Neunzert, Universität Kaiserslautern

Die Arbeitsgruppe Technomathematik hat es sich zur Aufgabe gemacht, neue Formen und Möglichkeiten einer Kooperation zwischen Universität und Industrie im Bereich der Mathematik zu erarbeiten und durchzuführen. Dabei beschäftigt sich die Arbeitsgruppe mit den folgenden Schwerpunkten:

EINBEZIEHUNG KONKRETER FRAGESTELLUNGEN AUS DER INDUSTRIE IN DIE MATHEMATISCHE FORSCHUNG.

Im Rahmen des von der VW-Stiftung geförderten Forschungsprojekts "Technomathematik" werden mathematische Probleme aus der industriellen Praxis in Form von Problemseminaren, Diplomarbeiten und Forschungsaufträgen bearbeitet. Als Beispiele für schon bearbeitete oder in Bearbeitung befindliche Probleme seien genannt

- die Optimierung von Kurbelgetrieben, Nocken und Felgen;
- die analytische und numerische Untersuchung spezieller strömungsdynamischer und akustischer Probleme;
- die Simulation stochastischer Prozesse in der Zuverlässigkeitsanalyse.

PRAXISORIENTIERTE GESTALTUNG DER MATHEMATISCHEN AUSBILDUNG IM HINBLICK AUF EINE BESSERE VORBEREITUNG DER ABSOLVENTEN AUF DIE BERUFSWIRKLICHKEIT.

Dies geschieht z.B. durch den Studiengang "Technomathematik"; die wesentlichen Lernziele sind dabei:

- Bildung mathematischer Modelle für technische Probleme,
- Kenntnis von mathematischen Methoden zur analytischen und numerischen Auswertung der Modelle,
- Beherrschung des Computers als Werkzeug,
- Kommunikationsfähigkeit mit Ingenieuren.

Auch in die Mathematikausbildung der Ingenieure sollen Modellbildung und moderne, insbesondere numerische und stochastische Methoden verstärkt integriert werden.

MATHEMATISCHE WEITERBILDUNG FÜR DEN PRAKTIKER.

Das aus dem "Modellversuch zur mathematischen Weiterbildung" hervorgegangene Konzept für eine mathematische Weiterbildung für Ingenieure, Naturwissenschaftler und Mathematiker wird weiterentwickelt und fortgesetzt. Die angebotenen Kurse dienen der

- Unterstützung bei der Bewältigung praktischer Probleme,
- Anpassung an den neuesten wissenschaftlichen Erkenntnisstand,
- Einordnung des praktisch-beruflichen Wissens in einen theoretisch-wissenschaftlichen Rahmen,
- Auffrischung von Hochschulwissen.

Die Arbeitsgruppe Technomathematik setzt sich aus Professoren und Mitarbeitern der Universität Kaiserslautern und einer Gruppe von Mathematikern an der Technischen Hochschule Darmstadt unter der Leitung von Prof. Dr. Törnig zusammen.