Introduction to Stochastic Processes

AE4304 lecture # 1, edition 2017-2018

prof. dr ir Max Mulder

m.mulder@lr.tudelft.nl

Introduction to Stochastic Processes

For this introduction the following material was used:

- **Chapter 1 of Priestley, M.B. (1981). Spectral Analysis and Time Series, Academic Press Ltd., London**
- Chapter 6 of Stevens, B.L., & Lewis, F.L. (1992). *Aircraft Control and Simulation*, Wiley & Sons
- Chapters 1, 2 and 3 of Lecture notes *Aircraft Responses to Atmospheric Turbulence*

Chapter 1 of (Priestley, 1981) is put on the Blackboard. It provides an excellent overview of the field of analyzing stochastic processes in the time and frequency domain.

Importance of stochastic systems and signals

The nature of spectral analysis

For those who are really interested (Priestley chapter)

Periodic and non-periodic functions: Fourier analysis
The Fourier series
The Fourier integral

Energy distributions; Signal spectra

Random processes; Stationary random processes

Time series, and the use of spectral analysis in practice

Importance of stochastic signals and systems

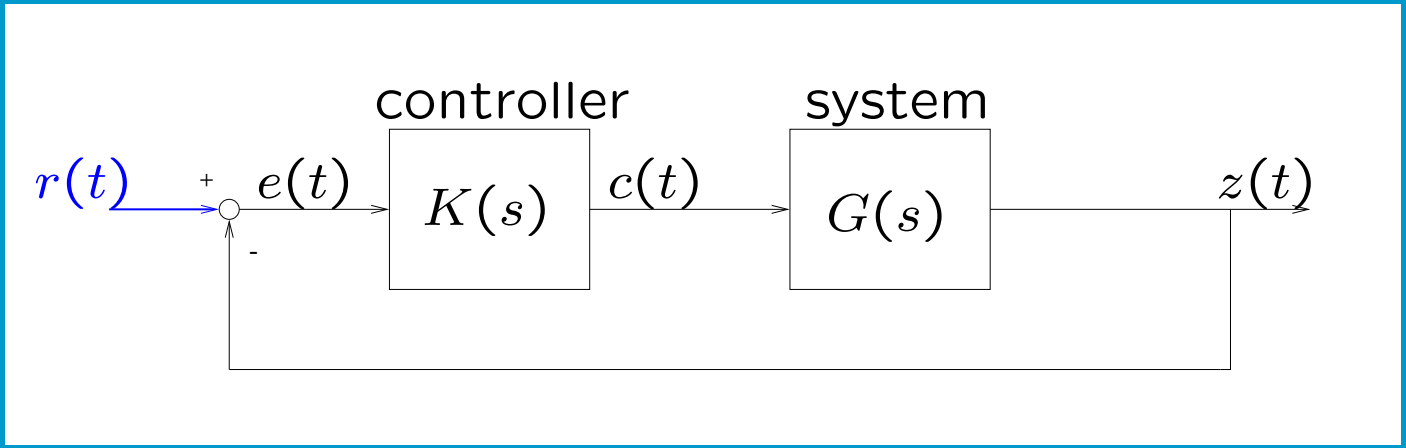
Almost all quantitative phenomena occurring in science (except perhaps the "pure" mathematics) are subject to some form of "randomness", i.e., elements that cannot be described using a "model" or "function", and consequently, all these real-life phenomena should be treated as random processes as opposed to deterministic functions.

The theory of stochastic (or random) signals and systems is an essential tool in the analysis and synthesis of real-life processes.

WARNING: The AE4304 lecture series can only provide a first introduction into this theory. Its scope is fairly limited and all students (in particular those that do the MSc profile "Control and Simulation") are recommended to follow at least one additional "signals and systems"-like course in their curriculum.

Example: the design of a simple servo control system

First start with the deterministic case, i.e., all systems are linear time-invariant (LTI), and the signal to be followed can be described using a mathematical function (such as a sinusoid, or a sum of sinusoids).



The principle design requirement is to design the controller $K(s)$ in such a way that the system output $z(t)$ follows the reference signal $r(t)$ as close as possible $\rightarrow$ minimize the error $e(t)$.

Thus: $\frac{Z(s)}{R(s)} = \frac{K(s)G(s)}{1+K(s)G(s)} \approx 1$ (and $\frac{E(s)}{R(s)} = \frac{1}{1+K(s)G(s)} \approx 0$).

Is this necessary for all frequencies? No, it is only necessary for *those frequencies that are below the bandwidth ω_r of reference signal $r(t)$*.

Solution: achieve a high loop gain $K(s)G(s)$ for $\omega \leq \omega_r$.

We see that it is important to know *at what frequencies the signal to be followed has "power"*.

BUT In reality, we are dealing with all sorts of **uncertainties**. Those concern the system involved (is it really linear? is it really time-invariant?) and the signals in the closed loop.

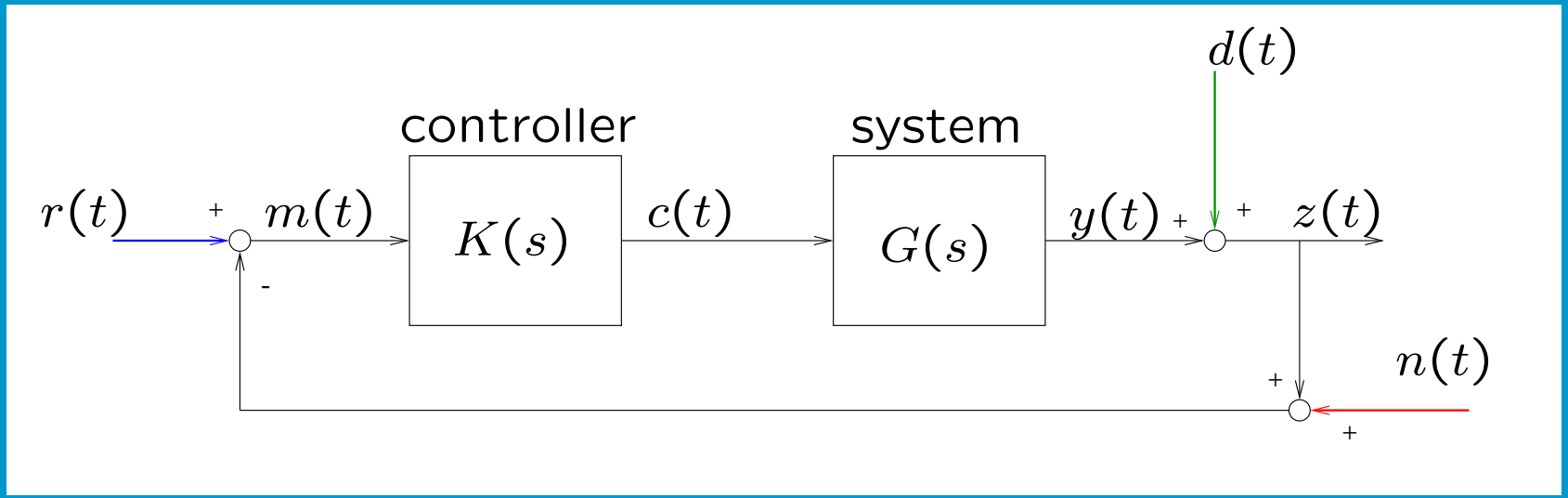
Some examples are:

- In reality, the system is almost certainly non-linear, and the analysis above only holds within the operating range of the linearization (i.e., only for small perturbations). E.g., consider the human controller.

- In reality, we may have to deal with *disturbances* on the system that we do not know and that are hard to predict. E.g., consider the turbulence acting on the aircraft.

- In reality, we cannot measure the signals in the loop accurately. All sensors have their own peculiarities in terms of bias, drift, all sorts of random inaccuracies, etc.

As a second and more realistic case, consider the same problem, but now in the presence of a random, unknown *system disturbance* $d(t)$ as well as a random, unknown *measurement noise* $n(t)$.



Note that a more common representation of the disturbance $d(t)$ is to put it in front of the system $G(s)$.

System disturbances $d(t)$ are typically low-frequency, i.e., the signal has "power" up to a certain frequency ω_d. Measurement noise $n(t)$ is typically high-frequency, i.e., the signal has "power" from a frequency ω_n.

Note that the reference signal $r(t)$ is also typically "low-frequency", i.e., it contains "power" up to frequency ω_r.

Now:

$$E(s) = R(s) - Z(s) \qquad \text{tracking error}$$
$$M(s) = R(s) - (Z(s) + N(s)) \qquad \text{'real' error}$$
$$Z(s) = G(s)K(s)M(s) + D(s)$$

Solving for $E(s)$ and $Z(s)$ yields:

$$E(s) = \frac{1}{1 + K(s)G(s)}\,(R(s) - D(s)) + \frac{K(s)G(s)}{1 + K(s)G(s)}N(s). \qquad \text{"error"}$$

$$Z(s) = \frac{K(s)G(s)}{1 + K(s)G(s)}\,(R(s) - N(s)) + \frac{1}{1 + K(s)G(s)}D(s). \qquad \text{"output"}$$

Define the system **sensitivity** $S(s) = (1 + K(s)G(s))^{-1}$ and the **complementary sensitivity** (or, co-sensitivity) $T(s) = K(s)G(s)S(s)$, so that $S(s) + T(s) = 1$ (in the scalar case).

Then:

$$\boxed{\begin{array}{rcl} E(s) & = & S(s)(R(s) - D(s)) + T(s)N(s) \\ Z(s) & = & T(s)(R(s) - N(s)) + S(s)D(s) \end{array}}$$

To achieve small tracking errors, $S(s)$ must be small at frequencies where $R(s) - D(s)$ is large. This is called **disturbance rejection**. Also, $T(s)$ must be small at frequencies where $N(s)$ is large. This is called **sensor noise rejection**.

Unfortunately, $S(s)$ and $T(s)$ cannot be small at the same time (their sum equals 1, remember?), so in general we have to compromise. A rule of thumb that holds in most practical cases is to get:

- $S(s)$ small ($T(s)$ large) at low frequencies where $R(s)$ dominates, i.e., a large "loop gain" $K(s)G(s)$ at low frequencies,
- $S(s)$ large ($T(s)$ small) at high frequencies where $N(s)$ dominates, i.e., a small "loop gain" $K(s)G(s)$ at high frequencies.

Again, we see that in order to meet these requirements **it is of the utmost importance to know at what frequencies the signals** $d(t)$ **and** $n(t)$ **have "power"** $\rightarrow$ Spectral Analysis !!

The nature of spectral analysis

"When we look at an object we immediately notice two things. First, we notice whether or not the object is well illuminated, i.e., we observe the strength *of the light (...) from the object, and secondly we notice the* colour *of the object. In making these simple observations our eyes have, in fact, carried out a very crude form of a process called* spectral analysis" (Priestley, 1981)

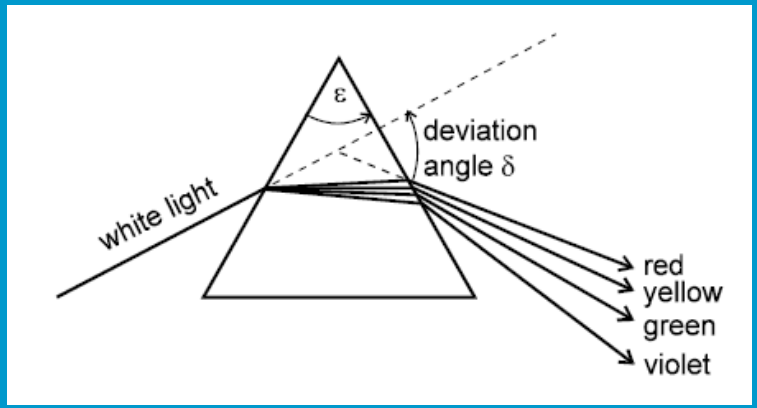
$$\text{strength} \approx \text{amplitude}$$

$$\text{colour} \approx \text{frequency}$$

Humans observe the "overall effect". That is, although we are able to say that light appears "blueish" or "reddish", we cannot say immediately how much blue or red it contains relative to other colours.

Newton's "opticks" (1704)

The components carried in the light have frequencies f_i, wavelengths $\lambda_i = 1/f_i$, the deviation angle when passing a prism is δ_i:

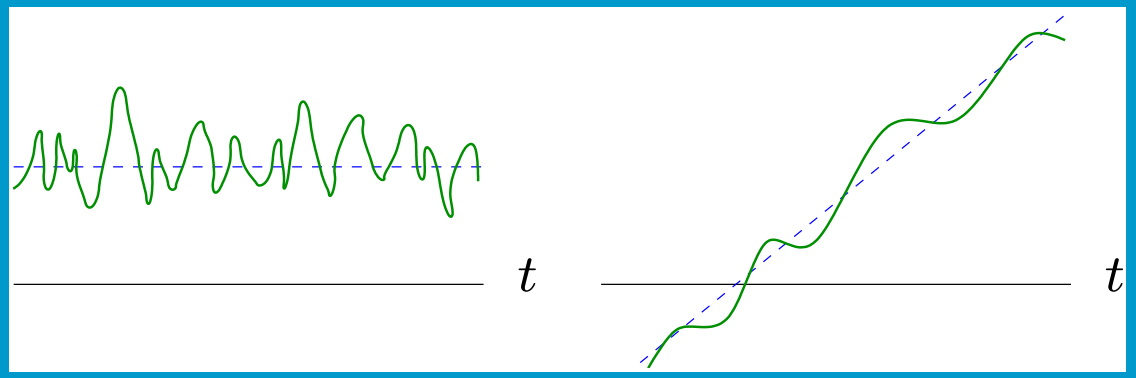


Red light: $\lambda = 700\ \mu$ m; green light $\lambda = 500\ \mu$ m.

Since δ increases when f increases, the (use of an optical device such as a) prism allows us to "split" the light into its various constituent colours, and determine the **strength** of the various components which are present. The prism allows us to "see" the *spectrum* of the light.

Note: white light $\leftrightarrow$ 'white noise'.

Light is a "wave"-like phenomenon, for which spectal analysis seems particularly appropriate. We will see, however, that we may usefully apply spectral analysis techniques to **any type of process which fluctuates in some form, but which tends to maintain a "steady" value**.



The slides from here follow the first chapter of Priestley. Very interesting, but the mathematical notations used are different from those in the lecture notes, which may be confusing.

Periodic functions

A function $f(x)$ is **periodic** if:

$$f(x) = f(x + kp), \text{ for all } x, \tag{1}$$

where k has an integer value, 0, ± 1, ± 2, ... The **period** p is the smallest number such that Eq. (1) holds. When there is no value of p (other than zero) where Eq. (1) holds, the function is **non-periodic**.

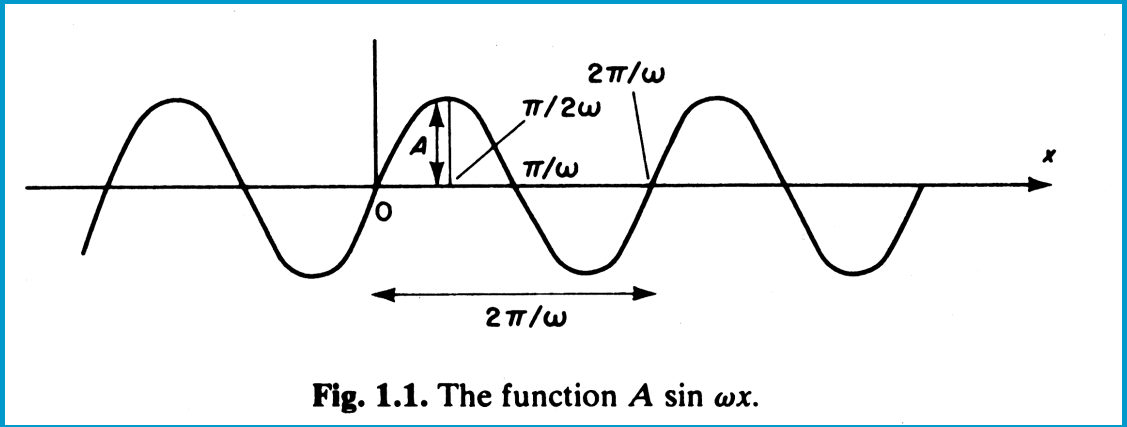


Fig. 1.1. The function A sin ωx.

Note: $p = 1/f$, and $f = \omega/(2\pi)$, so $\omega = \frac{2\pi}{p}$

Periodic functions: the Fourier Series

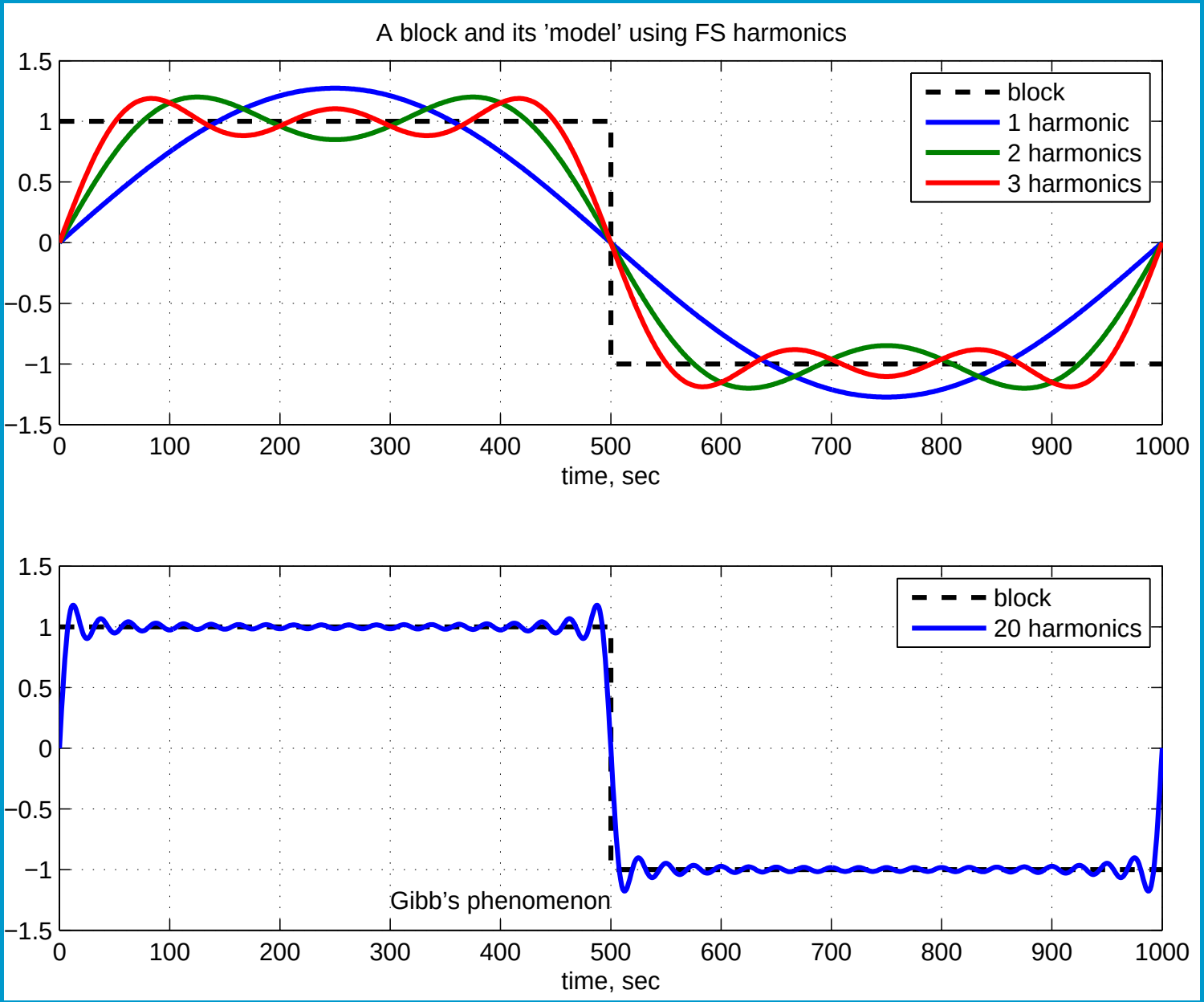
Fourier's (1768-1830) theorem states that any "well-behaved" periodic function $f(x)$ can be expressed as a (possibly infinite) sum of sine and cosine functions, the **Fourier series**:

$$f(x) = \sum_{r=0}^{\infty} \left[a_r \cos(r \left(\frac{2\pi}{p}\right) x) + b_r \sin(r \left(\frac{2\pi}{p}\right) x) \right] \tag{2}$$

where a_0, a_1, ..., b_0, b_1, ..., are *constants* that may be determined from the form of $f(x)$.

Note that the first term (r=0) is a constant, the second term (r=1) represents sine and cosine waves with period p, the third term (r=2) represents sine and cosine waves with period $p/2$, etc.

Hence, sine and cosine functions are the "building blocks" of **all** periodic functions: these can all be "constructed" from a suitable combination of these elements.


A block and its 'model' using FS harmonics
block
1 harmonic
2 harmonics
3 harmonics
time, sec
block
20 harmonics
Gibb's phenomenon
time, sec

Non-periodic functions: the Fourier Integral

A non-periodic function can be considered a periodic function with an *infinite period* ($p \to \infty$). When p is increased, the values of the coefficients in the Fourier series (a_0, a_1, ..., b_0, b_1, ...) are reduced. Also, the distance between the frequencies $2\pi r/p$ and $2\pi(r+1)/p$ of neighboring terms tends to zero, so that in the limit the summation in (2) becomes an integral:

$$f(x) = \int_0^\infty [g(\omega)\cos(\omega x) + k(\omega)\sin(\omega x)]\,\mathrm{d}\omega \tag{3}$$

where $g(\omega)$ and $k(\omega)$ are *functions* whose form may be determined from the form of $f(x)$. This integral is known as the **Fourier Integral**.

Note: an important requirement in deriving this expression is that the function $f(x)$ is "absolutely integrable", i.e., $\int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x < \infty$.

Rewrite (2) and (3) using complex numbers

Recall:

$$\sin\omega t = \tfrac{1}{2j}(e^{j\omega t} - e^{-j\omega t}) \qquad \cos\omega t = \tfrac{1}{2}(e^{j\omega t} + e^{-j\omega t})$$

If A_r is a (complex valued) sequence such that:

$$A_r = \begin{cases} \frac{1}{2}(a_r - jb_r) & r > 0 \\ a_0 & r = 0 \\ \frac{1}{2}(a_{|r|} + jb_{|r|}) & r < 0 \end{cases}$$

Then the Fourier series, (2), becomes two-sided:

$$\boxed{f(x) = \sum_{r=-\infty}^{r=\infty} A_r e^{j\omega_r x}} \tag{4}$$

where $\omega_r = r\left(\frac{2\pi}{p}\right)$, $r = 0, \pm 1, \pm 2$, etc.

Again, $\frac{2\pi}{p}$ is the **fundamental frequency**: one rotation of the complex exponential in p seconds.

Then $|A_r| = \sqrt{a_r^2 + b_r^2}$ is the amplitude, $\angle A_r = atan(\frac{-b_r}{a_r})$ is the phase, and ω_r is the angular frequency of the complex exponential function.

Similarly, for aperiodical functions, introduce the (complex valued) function $P(\omega)$:

$$P(\omega) = \begin{cases} \frac{1}{2}\left(g(\omega) - jk(\omega)\right) & \omega > 0 \\ g(0) & \omega = 0 \\ \frac{1}{2}\left(g(|\omega|) + jk(|\omega|)\right) & \omega < 0 \end{cases}$$

Then the Fourier integral, (3), becomes:

$$\boxed{f(x) = \int_{-\infty}^{\infty} P(\omega) e^{j\omega x} \mathrm{d}\omega} \tag{5}$$

The function $P(\omega)$ is called the **Fourier transform** of $f(x)$.

Note: Priestley defines

$$P(\omega) = \mathcal{F}\{f(x)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-j\omega x} \mathrm{d}x$$

as the Fourier transform of a function $f(x)$, and:

$$f(x) = \mathcal{F}^{-1}\{P(\omega)\} = \int_{-\infty}^{\infty} P(\omega) e^{j\omega x} \mathrm{d}\omega$$

as the inverse Fourier transform of a function $f(x)$.

As will be discussed in the lecture notes (§3.2.4) there is no universal agreement on where to put the $1/2\pi$-term, that is, some authors (like Priestley) put this term in the Fourier transform, whereas others (like in the lecture notes) put this term in the inverse Fourier transform.

Further note that throughout this course AE4304 we will use ω as it is closer related to our 'control' and dynamics courses. Remember from AE2235-II, that when we use f, there is no need for any scaling at all in the Fourier transforms!

In the following, we adopt the definition of the lecture notes.

Again consider the Fourier series, (4), and the Fourier integral, (5):

$$f(x) = \sum_{r=-\infty}^{r=\infty} A_r e^{j\omega_r x}$$

$$f(x) = \int_{-\infty}^{\infty} P(\omega) e^{j\omega x} \mathrm{d}\omega$$

Comparing (4), for periodic functions, and (5), for non-periodic functions, we now see the essential difference between periodic and non-periodic functions – namely, that whereas a periodic function can be expressed as a sum of cosine and sine terms over a *discrete set of frequencies* ω_r (ω_0, ω_1, ω_2, ...), a non-periodic function can be expressed only in terms of cosines and sines which cover the *whole continuous range of frequencies*, i.e., from 0 to infinity.

From this point, Priestley continues with a theoretic account of the Fourier-Stieltjes integrals which unify the Fourier series and the Fourier integral. As we take a more practical approach in this lecture, you need only to read this part, until §1.5.

Energy distributions

The *energy* carried by a sine or cosine term is defined as being proportional to the *square* of its amplitude.

Consequently, in the case of a periodic process, the contribution of a term of the form $(a_r \cos \omega_r x + b_r \sin \omega_r x)$ equals $|A_r|^2$. If we plot the squared amplitudes $|A_r|^2$ against the frequencies ω_r the graph obtained shows the **relative contribution of the various sine and cosine terms to the total energy**. This graph is called an "*energy spectrum*" or "*power spectrum*".

For any signal $x(t)$:

$$\text{Energy } E = \int_{-\infty}^{\infty} x^2(t)dt$$

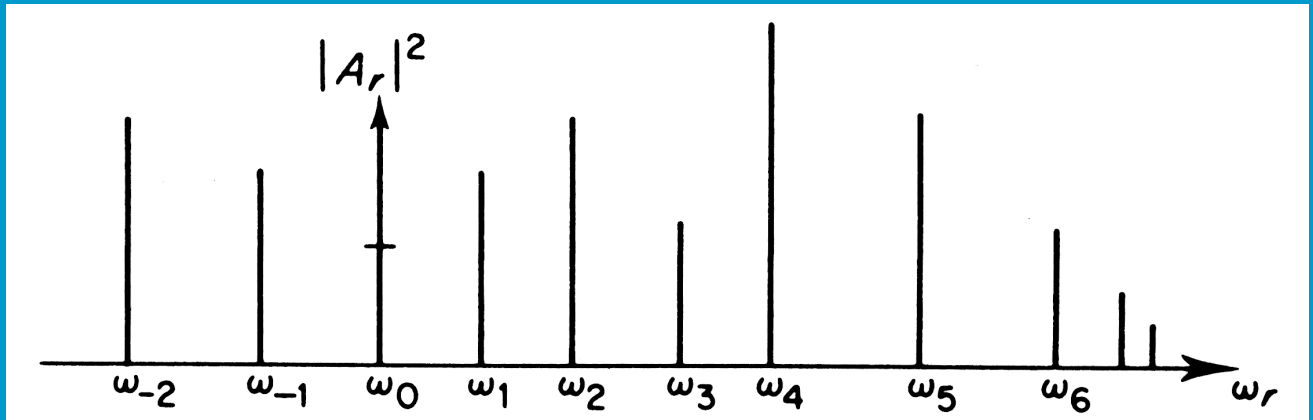


Fig. 1.5. Discrete spectrum.

For a periodic process with period p the only frequencies of interest are those which are integer multiples of the fundamental frequency (in this graph $\omega_1 = (2\pi/p)$). The total energy of the process is divided up entirely among this *discrete set* of frequencies. Hence, a periodic process has a **discrete spectrum**.

Note that the spectrum is an even function ($x(-t) = x(t)$).

Further note the energy at $\omega_0 = 0$, the "zero frequency": what does it mean?

In the case of a non-periodic process, *all* frequencies contribute, and the total energy is spread over the whole *continuous* range of frequencies. Thus, as we plot $|P(\omega)|^2$ as a function of ω, the type of energy distribution so obtained is called a **continuous spectrum**.

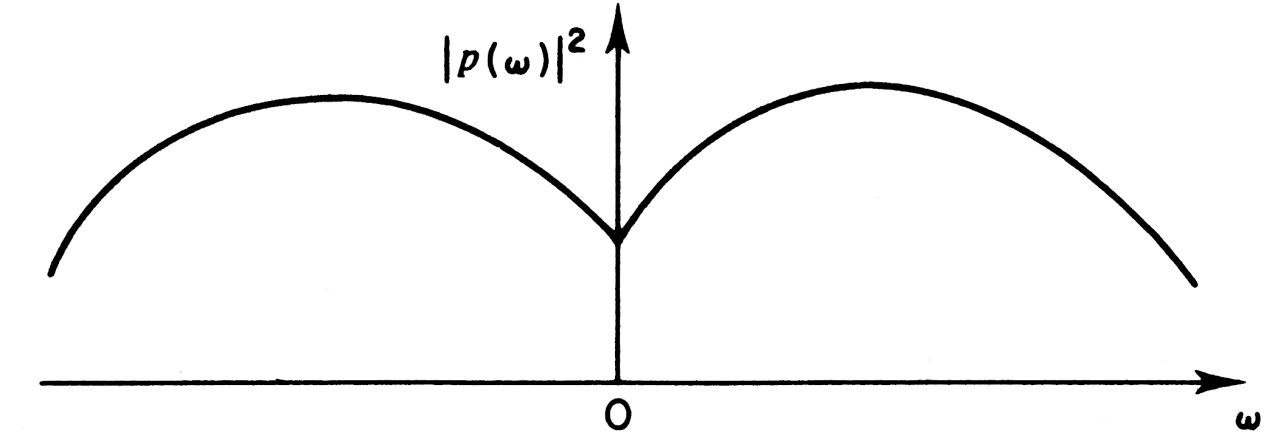


Fig. 1.6. Continuous spectrum.

Again, the spectrum is an even function.

Random processes

Deterministic functions $f(x)$ are functions which are such that, for each value of x, we have a rule, often in terms of a mathematical formula, that enables us to calculate the *exact* value of $f(x)$. E.g., $f(x) = \sin(\omega x)$.

They form the domain of classical mathematical analysis, but in reality **almost all the processes which we encounter are not of this type**. Consider for example, a record of the variations in thickness of a piece of yarn (NL: "garen"), plotted as a function of length.

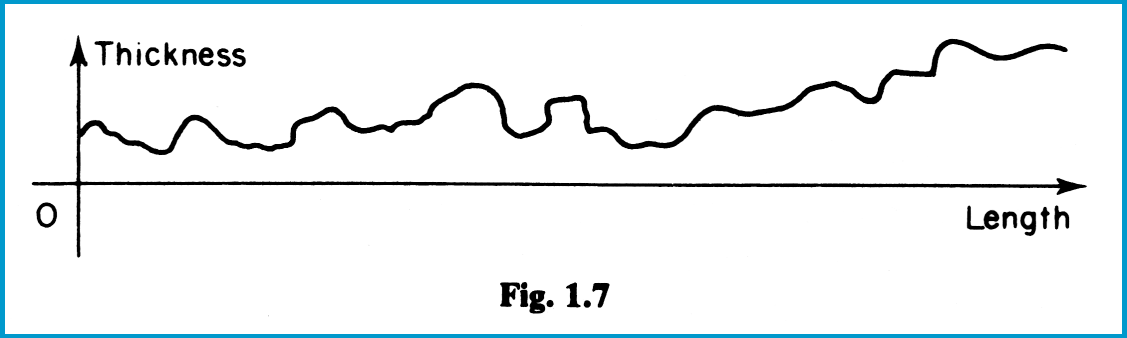


Fig. 1.7

No deterministic function is available that describes this process. A function *might* be found that could give a good approximation to the graph over a finite interval, say, 0 to L, but we would find that outside this interval the function is no longer a valid approximation, no matter how large we would make L. Why does this happen?

It turns out that the manufacturing process involves a (large) number of factors that cannot be controlled precisely, and whose effects on the thickness of the yarn may simply *never* be fully understood.

The variations in thickness along the length of the yarn have a **random character** in the sense that *we cannot determine theoretically what will be the precise value of the thickness at each point of the yarn:* **rather, at each point there will be a whole range of possible values**.

The only way to describe the thickness of the yarn as a function of length is to specify, at each point along the length, a *probability distribution* which describes the relative "likeliness" of each of the possible values. I.e., the thickness is a **random variable** and the complete function (thickness against length) is called a **random (or stochastic) process**.

The term "stochastic" originates from the ancient Greek: $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa o\varsigma$ means "the art of guessing".

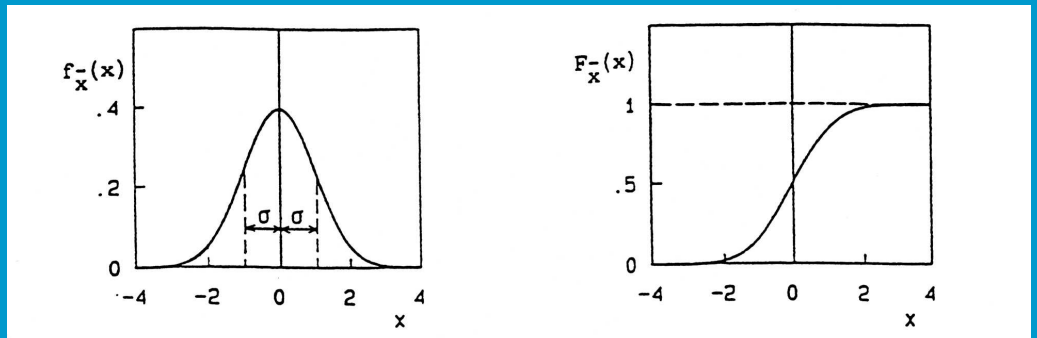


Figure 2.7: The Gaussian probability density- and distribution functions (plotted here with $\mu_{\bar{x}} = 0,\ \sigma_{\bar{x}} = 1$).

Recall: $f_{\bar{x}}(x) = \frac{dF_{\bar{x}}(x)}{dx}$.

Random variables

Consider now a random variable, $x(t)$ (i.e., time-varying) which arises from an experiment that may be repeated under identical conditions.

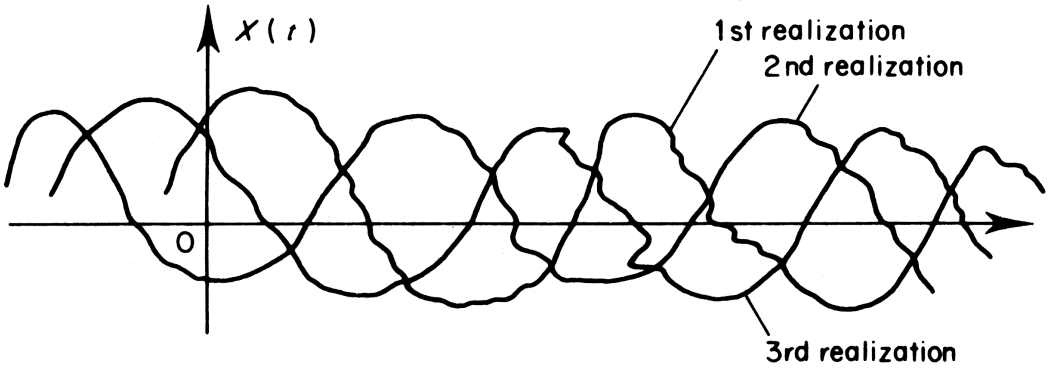


Fig. 1.8. Realizations of a random process.

E.g., here $x(t)$ is the variation in yarn thickness w.r.t. the average, as a function of time.

Hence, *an observed record of a random process is merely one record out of a whole collection of possible records which we might have observed*. The collection of *all* possible records is called the **ensemble**, and each particular record is called a **realization** of the process.

Note that the ensemble is a theoretical abstraction.

That is, the set of all possible observations is infinitely large!

Equivalently, one could consider the ensemble as one single realization but with an **infinite observation time**.

In this lecture we will see that the length of the observation time will play a very large role, as the *theory* is largely developed for infinite observation time (Fourier transforms) but the *practical* implementation is for finite observation time (DFT, Fourier Series).

Origins of "randomness"

In the real world, the random character of most of the processes which we study arise as a result of:

1. "inherent" random elements in (or acting on) the physical system,
   (e.g., the effects of turbulence on the aircraft motion)
2. the fact that it is impossible to describe the behavior of the process in other than probabilistic terms,
   (e.g., the behavior of humans in basically any imaginable task)
3. measurement errors made while observing the process variables.
   (e.g., any sensor contains "noise")

Almost all quantitative phenomena occuring in science are subject to one or more of these factors, and consequently should be treated as random processes as opposed to deterministic functions.

Stationary random processes

In many physical problems we encounter random processes which may be described loosely as being in a state of "statistical equilibrium". That is, if we take any realization of such a process and divided it up into a number of time intervals, the various sections of the realization look "pretty much" the same.

We express this type of behaviour more precisely by saying that, in such cases *the statistical properties of the process do not change in time*.

Random processes that possess this property are called **stationary**, and all processes which do not possess this property are called **non-stationary**.

In this lecture we will confine ourselves to stationary random processes. Read §1.8 and skip §1.9 of Priestley.

Time series analysis: use of spectral analysis in practice

Generally, the object of study is a random process $x(t)$ of which we have a finite number of observations (samples) in time, say $t = 1$, 2, ..., N. That is, we have a (portion of) a single realization, or, a so-called **time series**.

Our aim is **to infer as much as possible about the properties of the whole stochastic process** of which we have measured this single realization. The four most important applications are:

I **The direct physical application**

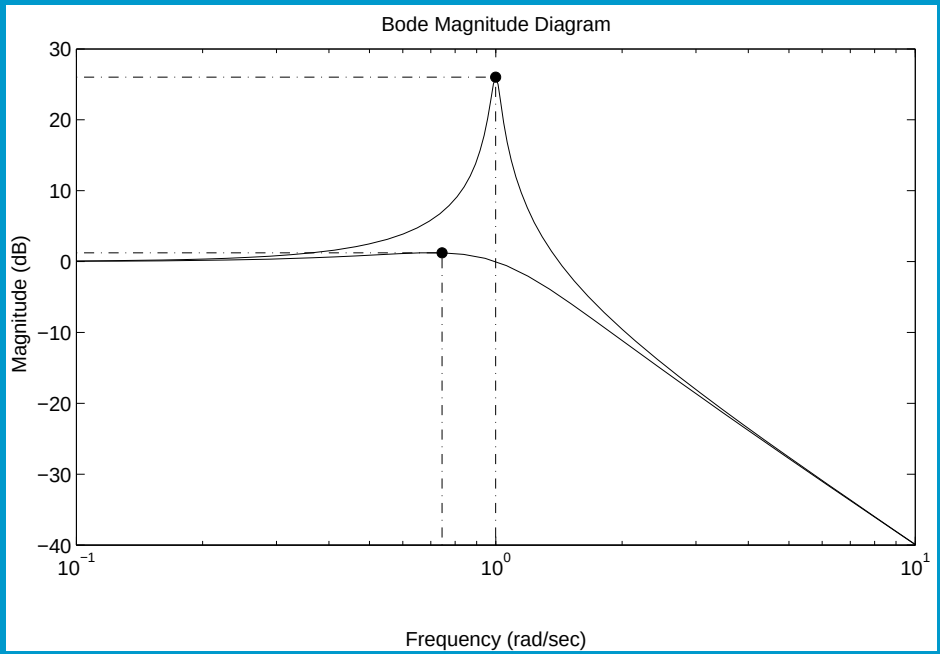
II Use in statistical model fitting

III **The estimation of transfer functions**

IV Prediction and filtering

I The direct physical application

If a structure has a number of "resonant" frequencies then it is crucial to design it so that these resonant frequencies are not "excited" by the random driving force.



Magnitude-plot of a mass-spring-damper system for two values of damping ζ (0.5 and 0.025), resonant frequency ω_0 at 1 rad/s.

A design requirement is to have the resonant frequencies of the structure fall in a region where the spectrum of the random vibrations has a very low energy content. Hence, it is in this case crucial to know the form of the "vibration" spectrum.

II Use in statistical model fitting

When studying a process it is often customary to gain some insight into the *probabilistic structure* of the process by attempting to describe its behavior in terms of a statistical model: "model fitting".

Common models are the **auro-regressive (AR)** model:

$$x_k + a_1 x_{k-1} + \cdots + a_N x_{k-N} = \epsilon_k, \tag{6}$$

where a_1, ... a_N are the constants and ϵ_k is a "purely random" process, so-called **white noise**, and the **moving-average (MA)** model:

$$x_k = \epsilon_k + b_1 \epsilon_{k-1} + \cdots + b_M \epsilon_{k-M}, \tag{7}$$

where b_1, ... b_M are the constants and ϵ_k is again the white noise. Mixing (6) and (7) yields the ARMA model.

Note: "white noise" is called after "white light", which contains *all* colours, i.e., all frequency components in the spectrum have an equal magnitude.

As we will see later (III), each type of model gives rise to a spectrum which has a characteristic shape. If, therefore, we can estimate the shape of the spectrum *directly from the observations* without assuming that the process conforms to a particular type of model, then the spectral "shape" provides a very useful indication of the type of model which should be fitted to the data.

However, when we have identified the general form of the model we are still faced with the problem of *estimating the numerical values of the "parameters"* (i.e., the constants a_1, a_2, ..., a_N in (6) and b_1, b_2, ..., b_M in (7)) which arise in the model.

III Estimation of transfer functions

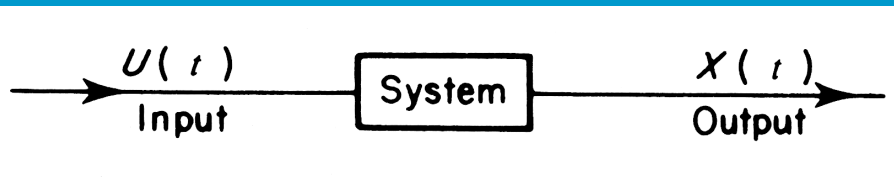


Fig. 1.10

If we assume that a process behaves like a Linear Time-Invariant (LTI) system, then at any time instant the current value of the output process $x(t)$ is a linear combination of present and past values of the input process $u(t)$:

$$x(t) = \int_0^\infty g(\tau)u(t-\tau)\mathrm{d}\tau, \tag{8}$$

the well-known **convolution integral** (or: $x(t) = g(t) \star u(t)$), with $g(t)$ the **impulse-response function** of the system.

If $u(t)$ is a sine wave with frequency ω, $u(t) = Ae^{j\omega t}$, then:

$$\begin{aligned} x(t) &= \int_0^\infty g(\tau) A e^{j\omega(t-\tau)} \mathrm{d}\tau \\ &= A \left\{ \int_0^\infty g(\tau) e^{-j\omega\tau} \mathrm{d}\tau \right\} e^{j\omega t} \\ &= A \cdot \Gamma(\omega) \cdot e^{j\omega t} \\ &= A \cdot |\Gamma(\omega)| e^{j\angle\Gamma(\omega)} \cdot e^{j\omega t} \\ &= \underbrace{\{A \cdot |\Gamma(\omega)|\}}_{change\ in\ amplitude} \cdot \underbrace{\left\{ e^{j(\omega t + \angle\Gamma(\omega))} \right\}}_{change\ in\ phase} \end{aligned} \tag{9}$$

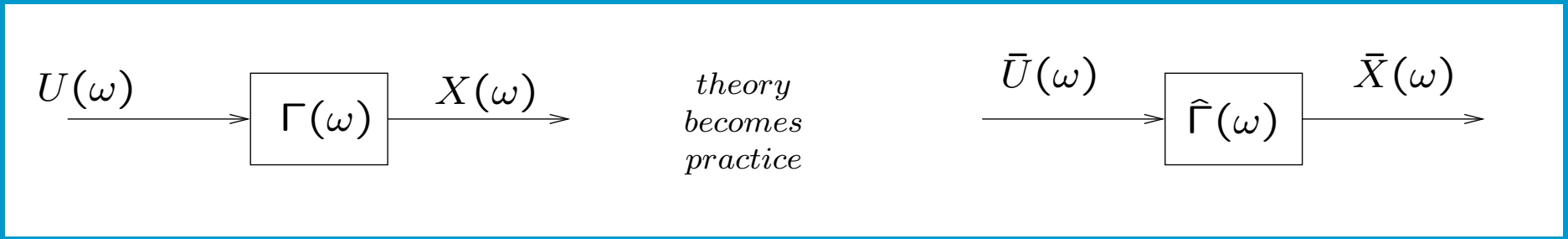
We see that, *if the input of an LTI system is a sine wave, the output is also a sine wave with exactly the same frequency but with modified amplitude and phase.*

$\Gamma(\omega)$ is the **frequency response function** of the system, the Fourier transform of the system **impulse response function** $g(t)$:

$$\boxed{\Gamma(\omega) = \int_0^\infty g(\tau) e^{-j\omega\tau} \mathrm{d}\tau = |\Gamma(\omega)| e^{j\angle\Gamma(\omega)}} \tag{10}$$

In many applications, the frequency response is used to describe the system behavior. For simple systems, $\Gamma(\omega)$ can be calculated theoretically, but for more complex systems this is impossible. In these cases, $\Gamma(\omega)$ is determined using a time series, i.e., the "form" of $\Gamma(\omega)$ is **estimated** from observed records (realizations) of $x(t)$ and $u(t)$.

That is, from measured input and output signals we can obtain an **estimate** of the system's frequency response: $\hat{\Gamma}(j\omega)$:



This estimate $\hat{\Gamma}(j\omega)$, obtained in the frequency domain, is called the **estimated** frequency response function (FRF).

Recall that, for each frequency, the power spectrum is proportional to the squared modulus of the amplitude of that frequency component, see the section on energy distributions.

Assume $x(t)$ and $u(t)$ are stationary random processes. Then, the squared modulus of each frequency component in $x(t)$ equals the squared modulus of the corresponding frequency component in $u(t)$ multiplied by $|\Gamma(\omega)|^2$ (see Eq. (9)):

$$\boxed{\{\text{Powerspectrum } x(t)\} = \{\text{Powerspectrum } u(t)\} \cdot |\Gamma(\omega)|^2} \tag{10}$$

Thus, when we can obtain *estimates* of the power spectral densities of $u(t)$ and $x(t)$, $|\Gamma(\omega)|^2$ can be *estimated* by the quotient of the spectral density estimates of the process output signal and the process input signal.

Note that we have obtained an estimate for $|\Gamma(\omega)|^2$, and NOT for $\Gamma(\omega)$! We get the amplitude characteristic $|\Gamma(\omega)|$ of the system, but "miss" the *phase* characteristic $\angle\Gamma(\omega)$. For this purpose so-called *cross-spectral densities* are used, which are discussed in Chapters 3 and 4 of the lecture notes.

The method above is known as the "non-parametric" approach, where the only assumption that needs to be made about the system is that it is an LTI system. In practice, however, the method is also used for systems which are known to be non-linear, such as in the case where the characteristics of a human controller are determined from experimental data. This has proven to be a feasible approximation.

When we have some more knowledge about the system, e.g.,, we know $\Gamma(\omega)$ is a rational function of ω, i.e.,

$$\Gamma(\omega) = \frac{b_0 + b_1 j\omega + ... + b_M (j\omega)^M}{1 + a_1 j\omega + ... + a_N (j\omega)^N}, \tag{11}$$

then we only need to estimate the $M + N + 1$ *parameters* a_1, a_2, ..., b_0, b_1, ..., from the data. In this case we can use the spectral analysis techniques as before to get the spectral density estimate $\hat{\Gamma}(j\omega)$ of the transfer function $\Gamma(j\omega)$, and parameterize the function (11) by calculating the parameters such that a criterium is minimized. Hence, a two-step approach.

Note: look at the similarity of the problem in (III) where we "parameterize" the system properties in the frequency-domain, with the problem in (II), where essentially a time-domain technique is used. In practice, both methods are used simultaneously.

IV Prediction and filtering

Prediction

Given the observed values of a random process at all past time points $\{x(s); s \leq t\}$ and wish to predict the value it will assume at some specific future time point τ, i.e., $x(t+\tau)$ for $\tau > 0$.

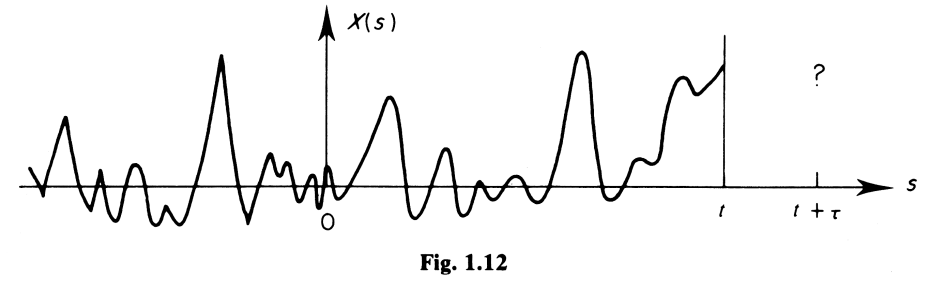


Fig. 1.12

The problem is to find a "function" of the given observations which makes the value of a criterion $\mathcal{M}$ as small as possible:

$$\mathcal{M} = \text{avg}\{x(t+\tau) - \hat{x}(t+\tau)\}^2,$$

i.e., minimize the *mean-square prediction error*.

Take for instance a predictor function of a discrete-time random process: $\widehat{x}_{k+K} = \sum_{j=0}^{K} a_j x_{k-j}$,
then we have to find the sequence a_j so as to minimize $\mathcal{M}$. It can be shown that the **optimal** form of this sequence is determined *uniquely* by the power spectral density function of x_k. Hence, in order to obtain an optimal predictor for a random process, we need to know the power spectral density of that process.

Filtering

The problem of *linear filtering* is in fact a more general version of the prediction problem and arises when we are unable to make accurate observations on the process x_k directly, but rather observe the process y_k where $y_k = x_k + n_k$, and n_k a noise disturbance.

The problem then becomes: given a record of past values of y_k, say at k, $k-1$, $k-2$, ..., construct a *linear filter* function of the form:

$$\widehat{x}_{k+K} = \sum_{j=0}^{\infty} b_j y_{k-j},$$

which yields the best approximation of x_{k+K} in the sense that the mean-square error is minimized. Note that K may be positive, negative, or zero, since it might be of interest to "estimate" the *unobserved value of* x_k corresponding to a "future" (i.e., the prediction problem), "present" (the filtering problem) or "past" (i.e., smoothing the time series) time point.
Hence, the filtering problem may be regarded on as that of "filtering out" the noise disturbance n_k, so as to reveal the underlying process x_k. It is obvious that this technique is crucial in any environment where the sensors have to obtain measurements of the process variables at hand.

Again, the **optimal** choice of the sequence b_j depends *entirely on the spectral and cross-spectral properties of the processes* x_k *and* n_k.

Contents of Part One of the AE4304 lectures

Part One

Chapter 2 *Scalar Stochastic Processes*

Chapter 3 *Spectral Analysis of Continuous-Time Stochastic Processes*

Chapter 4 *Spectral Analysis of Discrete-Time Stochastic Processes*

Chapter 5 *Multivariate Stochastic Processes*