STATISTICAL THEORY OF
NOISE AND MODULATION
6.573

C.T. WHITEHEAD FALL '60

2739 N

Name

6.573 : Statistical Theory of Noise & Modulation

Prof. Siebert

Text: Random Signals & Noise
Davenport & Root

Review of Probability (Chapters 2,3,4):

Definition of probability:

Sidestepping the "relative frequency" – "degree of belief" arguments, probability can be defined by four axioms:

I. Let $A = \{a_i : a_i$ is a possible event or outcome of a specific experiment$\}$
Then $\exists$ a real non-negative number $P(a_i)$ for each $a_i \in A$.
This number is defined as the probability of the event a_i.

If $a_i \in A$ and $a_j \in A$, then $(a_i \overset{\text{and}}{\wedge} a_j) \in A$, $(a_i \overset{\text{or}}{\vee} a_j) \in A$,
and $(\overset{\text{not}}{\sim} a_i) \in A$.

II. The probability of a certain event is unity.

III. If $a_i \in A$ and $a_j \in A$ are mutually exclusive events, then $P(a_i \text{ or } a_j) = P(a_i) + P(a_j)$

IV. If $P(b_i)$ is defined for all $b_i \in B \subset A$ then $P(b_1 \text{ or } b_2 \text{ or} \dots)$ is defined. If the b_i further are mutually exclusive, then

$$P(b_1 \text{ or } b_2 \text{ or} \dots) = \sum_{b_i \in B} P(b_i)$$

These axioms require that $0 \le P(a_i) \le 1$, that

$P(a_i) + P(\sim a_i) = 1$; and for very large numbers of trials approach the relative frequency.

$P(a_j) = 0 \Leftarrow a_j$ never occurs [the implication cannot be reversed]

Similarly $P(a_i) = 1 \Leftarrow a_i$ always occurs. This seems a flaw in the definition.

<u>Joint probabilities</u>:

Let $A = \{a_i\}$ and $B = \{b_j\}$ $\ni$ $\sum_A P(a_i) = 1 = \sum_B P(b_i)$
~~and the $b_j \in B$ are mutually exclusive~~.

Then $\exists$ $P(a_i, b_j)$ $\ni$ $0 \leq P(a_i b_j) \leq 1$ where

$P(a_i b_i)$ is <u>the probability of the joint event</u> $(a_i b_i) \in A \times B$.

If the $b_j \in B$ are mutually exclusive, then

$$P(a_i) = \sum_B P(a_i b_j) \;. \geq P(a_i, b_k)$$

<u>Conditional probabilities</u>:

$$P(a|b) \equiv \frac{P(a,b)}{P(b)} \quad \text{so that} \quad P(a,b) = P(a|b)P(b)$$

$$P(a|b) \geq P(a,b) \quad \text{but } P(a|b) \gtreqless P(a)$$

$$\sum_{b \in B} P(a|b) = P(a)$$

$$\sum_{a \in A} P(a|b) = 1$$

<u>Statistical independence</u>:

If $P(a|b) = P(a)$, then $P(a|b)P(b) = P(a\,b) = P(a)P(b) = P(b|a)\,P(a)$
$\Rightarrow P(b|a) = P(b)$

If $P(a|b) = P(a)$ or $P(b|a) = P(b)$, then a and b are said to be <u>statistically independent</u>.

For more than two events, pairwise independence does not assure complete independence. For this we must require

$$P(a_i^{(1)} a_j^{(2)} \dots a_k^{(n)}) = P(a_i^{(1)})P(a_j^{(2)})P(a_k^{(n)}) \;,\; i \neq j \neq k \neq i \;,\; i,j,k = 1,2,\dots,n \,.$$

Random variables:

Sample space is the set of all possible outcomes of an experiment. The sample space depends on the experiment. Let S be the sample space.

$$\text{If } A \subset S,\ A = \{a_i\}, \text{ and } S = \{s\}$$

$$P(A) = P(s \in A)$$

A random variable $x(s)$ is a function whose domain is the sample space S.

Probability distribution functions:

The probability distribution function of a random variable is

$$P(x \le X) = \Pr\{x \le X\} \ ; \text{ or joint, } P(x \le X, y \le Y)$$

and is well defined for continuous or discrete random variables.

~~Finite~~ Discrete random variables:

If x can ~~[illegible]~~ assume only a finite number of values in a region, x is a discrete random variable. There is thus a finite probability $P(x_i)$ that $x = x_i$.

Continuous random variables:

A random variable may be able to take on any value over a continuous region. In such a case, the probability distribution *may* be continuous over the region (if there are no single points which have finite probability). Such a variable is a continuous random variable.

Probability density functions:

For a continuous random variable, we can define a probability density function $P(x)$, where

$$P(x) \equiv \frac{d}{dx_1} P(x_1 \leq x) \, . \quad \geq 0$$

If there is some $x = x_0$ which has a finite probability of occuring, this will appear as an impulse of area $P(x_0)$ in the probability density function.

~~If there are an infinite number of~~

If S is a subset of $A = \{x\}$, then

$$P(S) = \int_{x \in S} P(x)\,dx$$

Note also that $P(A) = 1 = \int_{x \in A} P(x)\,dx$

Since both S and A are assumed continuous, we have for the probability of a particular point $x_0 \in S \subset A$

$$P(x_0) = \lim_{S \to x_0} \int_{x \in S} P(x)\,dx = \int_{x_0}^{x_0} P(x)\,dx = 0 \, .$$

Thus the probability $P(x)$ of any particular point is zero, although obviously the event x is not impossible.

If $P(x)$ has impulses, we can still perform the integration to get $P(S)$, since the integral of a unit impulse is by definition a unit step.

Joint probability densities:

$$P(x,y) = \frac{\partial^2}{\partial x \partial y} P(x_1 \le x, y_1 \le y)$$

$$P(x) = \int_{-\infty}^{\infty} P(x,y)\,dy$$

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, P(x,y) = 1$$

$$P(S) = \iint_{x,y \in S} P(x,y)\,dx\,dy \quad , \quad S \subset \{x,y\}$$

Conditional density functions:

$$P(y_1 \le y \mid x_1 = x) = \lim_{\Delta x \to 0} P(y_1 \le y \mid x - \Delta x \le x_1 \le x)$$

$$= \lim_{\Delta x \to 0} \frac{\int_{-\infty}^{y} \int_{x-\Delta x}^{x} P(x_1, y_1)\, dy_1\, dx_1}{\int_{x-\Delta x}^{x} P(x)\, dx_1} = \lim_{\Delta x \to 0} \frac{\int_{-\infty}^{y} P(x, y_1)\, dy_1\, \Delta x}{P(x)\;\Delta x}$$

$$= \frac{\int_{-\infty}^{y} P(x, y_1)\, dy_1}{P(x)} = P(y_1 \le y \mid x)$$

So $\quad P(y|x) = \frac{d}{dy} P(y_1 \le y \mid x) = \underline{\underline{\frac{P(x,y)}{P(x)}}}$

$$P(y \in S \mid x) = \int_{y \in S} P(y|x)\, dy$$

If $S = \{y\}$, $\int_{-\infty}^{\infty} P(y|x)\, dy = 1$.

Statistically independent random variables:

Just as statistical independence was defined for the events which are the set of possible outcomes of ~~an~~ two experiments so can we define statistical independence for two random variables by associating each variable with an experiment. If the variable takes on a certain value for each possible outcome of the experiment ($x = x_\alpha$ when event a_α occurs), then the variables x, y are statistically independent if

$$P(x,y) = P(x)P(y) \quad \text{or,} \quad P(x) = P(x|y)$$

For continuous random variables, this becomes

$$P(x_1 \le x, y_1 \le y) = P(x_1 \le x)\, P(y_1 \le y)$$

By partial differentiation with respect to x and y,

$$p(x,y) = p(x)\,p(y).$$

Functions of random variables:

Let x be a random variable and let $y \equiv g(x)$. Then If $S_x = \{x\}$, and $A \subset S$, there is a space $B \subset S_y \equiv \{y\}$ such that

$$P(x \in A) = P(g(x) \in B) = P(y \in B)$$

Assuming x and y to be continuous, we require

$$p_x(x)\,dx = p_y(y)\,dy = p_y(y)\frac{d}{dx}g(x)\,dx$$

or $$p_x(\cdot) = p_y(\cdot)\frac{d\,g(x)}{dx} \quad \text{or} \quad p_y(\cdot) = p_x(\cdot)\frac{1}{\frac{dg(x)}{dx}}$$

So $$\boxed{p_y(y) = p_x(g^{-1}(y)) \left|\frac{d\,g^{-1}(x)}{dy}\right|}$$

magnitude sign means always ignore a $(-)$ sign before $\frac{dg^{-1}(x)}{dy}$.

If we have several variables,

$$x_n = f_n(y_1, \ldots, y_N), \quad n = 1, \ldots, N$$

then

$$\boxed{P_y(y_1, \ldots, y_N) = P_x(f_1, \ldots, f_N) \left| \frac{\partial(f_1, \ldots, f_N)}{\partial(y_1, \ldots, y_N)} \right|}$$

Random processes:

A random process is defined to be a series of "experiments" over ~~the~~ time. ~~The outcome of the~~ The output of the process is the sequence of outcomes of consecutive experiments. By tagging each experiment, we define a parameter, ~~which ... describes the~~ ~~the~~ the value of which ~~it~~ tells us the probability distribution over the possible outputs of the process at any time. This parameter may be discrete or continuous.

Suppose we have N coins numbered $1, 2, \ldots, N$, and our random process consists of tossing the coins in order. Then the parameter describing this process is n, the number of the coin being tossed and the probability distribution is $(P_n, 1-P_n)$. This is a discrete parameter random process.

If the process was instead an ~~m~~astable flip-flop for which ~~the~~ probability of changing state was a function of time, $P_r\{\text{change state}\} = P_c(t)$, the parameter would be t. This is a continuous parameter random process.

The output of a random process may be continuous or discrete, independent of the nature of the parameter.

A stationary random process is one in which the probability distribution over possible outputs is the same for all time.

Statistical averages (Expectation):

An arithmetic average of the outcomes of an experiment is

$$\bar{x} = \sum_{k=1}^{M} x_k \frac{n(x_k)}{N} \qquad \text{for discrete r.v.}$$

As $N \to \infty$, this is presumably well behaved &:

$$\bar{x} = \sum_{k=1}^{M} x_k P(x_k) \equiv E(x)$$

Or, if x is continuous,

$$\bar{x} = \int_x x P(x)\,dx$$

Expectation of a function of a random variable:

Let $y = g(x)$, $P(x)$ is known.

$$E(y) = E[g(x)] = \sum_y y P_y(y) = \sum_x g(x) P_x(x)$$

or $E(y) = \int_x g(x) P(x)\,dx \;.= \int_y y P_y(y)\,dy.$

Functions of severable variables:

$$E[g(x,y)] = \int_x\int_y g(x,y) P(x,y)\,dx\,dy \qquad \text{continuous}$$

or $$E[G(x,y)] = \sum_x \sum_y G(x,y)\, P(x,y)$$

Notice that the expectation is a linear ~~function~~ operator:

$$E\Big[\sum_k a_k x_k\Big] = \sum_k a_k\, E(x_k).$$

Average of a random process:

For either a continuous-parameter or discrete-parameter random process, the expectation can be defined for each value of the parameter by ~~using~~ averaging over the probability distribution denoted by that value of the parameter (i.e. as a function of the parameter).

Let the output of a random process be $x(t) = x_t$ for a continuous parameter process. Then

$$E[g(x_t)] = \int_{x_t} g(x_t)\, P(x_t)\, dx_t = \int_x g(x)\, P_t(x)\, dx$$

This is the ensemble average of $g(x)$ and may be a function of the parameter t.

Moments of a probability distribution:

$$E(x^n) = \int_{-\infty}^{\infty} x^n P(x)\, dx \equiv n^{th} \text{ moment of } P(\cdot) .$$

Note that the moments depend on the probability distribution function and not on the dummy variable to which it is applied.

The central moments of a distribution are taken about the mean of the distribution rather than the origin:

$$E[(x-m_x)^n] = \int_{-\infty}^{\infty} (x-m_x)^n P(x)\, dx \equiv \mu_n \equiv n^{th} \text{ central moment of } P(\cdot).$$

For joint probability distributions, the joint moments can be similarly defined:

$$E(x^n y_k) = \iint x^n y^k\, P(x,y)\, dx\, dy$$

and $\mu_{nk} \equiv E\left[(x-m_x)^n (y-m_y)^k\right]$

$\mu_{11} \equiv$ covariance

Characteristic functions of probability distributions:

We can Fourier transform the probability distribution $P(\cdot)$ if $P(\cdot)$ is defined on a real variable, so that

$$M_x(j\nu) \equiv E[e^{j\nu x}] = \int_{-\infty}^{\infty} e^{j\nu x} P(x)\,dx \qquad , \quad \operatorname{Im}\{\nu\} = 0.$$

$M_x(j\nu) \equiv$ characteristic function of $P(\cdot)$.

We can reverse the transformation under most conditions so that the distribution can be found from its characteristic function:

$$P(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_x(j\nu)e^{-j\nu x}\,d\nu$$

If the distribution is discrete, we can analogously define a characteristic function (or if we use impulse integration properly, reduce the above definition to):

$$M_x(j\nu) = \sum_x P(x)e^{j\nu x}$$

The characteristic function always exists for a true probability distribution:

$$|M_x(j\nu)| = \left|\int_{-\infty}^{\infty} e^{j\nu x}P(x)\,dx\right| \le \int_{-\infty}^{\infty} |e^{j\nu x}|P(x)\,dx = \int_{-\infty}^{\infty} P(x)\,dx = 1 .$$

Moment generating functions:

If we transform $P(\cdot)$ by a similar transform such that $u = j\nu$ is constrained to be real,

$$M_x(u) \equiv \int_{-\infty}^{\infty} e^{ux}P(x)\,dx$$

& reverse: $$P(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_x(u)e^{-ux}\,dx.$$

This transform need not always exist:

$$|M_x(u)| \le \int_{-\infty}^{\infty} |e^{ux}|P(x)\,dx \quad \text{which is not always bounded.}$$

Moment generation:

If we differentiate $M_x(jv)$ re: v, we ~~can~~ differentiate under the integral and bring down a factor (jx) under the integral. Thus

$$\frac{d}{dv} M_x(jv)\Big|_{v=0} = j\int_{-\infty}^{\infty} x\, e^{jvx} p(x)\,dx\Big|_{v=0} = j m_x = j\,E(x)$$

Obviously we can extend this to any number of differentiations so:

$$\boxed{E(x^n) = (-j)^n \frac{d^n}{dv^n} M_x(jv)\Big|_{v=0}} = \frac{d^n}{d(jv)^n} M_x(jv)\Big|_{v=0}$$

All moments (& hence, all derivatives at zero) may not exist.

In some cases a Taylor series expansion of $M_x(jv)$ about $v=0$ may exist & be simple to find. Thus

$$M_x(jv) = \sum_{n=0}^{\infty} E(x^n)\frac{(jv)^n}{n!} = \sum_{n=0}^{\infty} \frac{d^n M_x(jv)}{dv^n}\Big|_{v=0} \frac{v^n}{n!}$$

Similarly, the moments can be found from the moment generating function if it exists. The moment generating function will exist only if all moments exist.

Joint characteristic functions:

$$M(jv_1, jv_2) = E\left[e^{jv_1x + jv_2y}\right]$$

$$M(0,0) = 1$$

$$E(x^n y^k) = (-j)^{n+k} \frac{\partial^{n+k}}{\partial v_1^n \partial v_2^k} M(jv_1, jv_2)\Big|_{v_1, v_2 = 0}$$

If $x_1, \ldots, x_N$ are statistically independent random variables,

$$M(jv_1, \ldots, jv_N) = \prod_{i=1}^{N} M_{x_n}(jv_n)$$

Chebyshev inequality:

$$E[|g(x)|] = \int_{\text{all}} |g(x)|\, p(x)\, dx \;\geq\; K \int_S p(x)\, dx \quad \text{where} \quad S = \{x : |g(x)| \geq K\}$$

$$= K\, P[|g(x)| \geq K]$$

or

$$\boxed{P[|g(x)| \geq K] \;\leq\; \frac{E[|g(x)|]}{K}} \xrightarrow{\times} \cancel{g(x) \geq 0}$$

In particular, let $g(x) = |x - m_x|$

Then $\quad P[|x-m_x|^2 \geq K^2\sigma^2] \leq \dfrac{E[|x-m|^2]}{K^2\sigma^2} = \dfrac{\sigma^2}{K^2\sigma^2} = \dfrac{1}{K^2}$

or $\quad \boxed{P[|x - m_x| \geq K\sigma] \;\leq\; \frac{1}{K^2}}$

Sums of statistically independent random variables:

Consider x, y $[p_x(x) + p_y(y)]$ as statistically independent r.v.'s

Let $z = x + y$.

$$p_z(z) = \int_{-\infty}^{\infty} p_x(x)\, p_y(z-x)\, dx = p_x(x) \otimes p_y(x)$$

Thus, to get the distribution of a sum of random variables, we convolve ~~their~~ the distributions. Or, we multiply their characteristic functions.

Central Limit Theorem:

Conditions for the existence of a central limit theorem are:

(1) $\overline{x}$ and $\overline{x^2}$ exist and all samples are statistically independently made from the same distribution

or (2) $\overline{x}$, $\overline{x^2}$, and $\overline{x^3}$ exist and all samples are statistically independent.

Let $x_1, \dots, x_N$ be statistically independent samples from a given distribution of mean m & variance σ^2.

Define $\xi_i = \dfrac{x_i - m}{\sigma}$

and $\xi = \sum_{i=1}^{N} \xi_i$

Let $\varphi_i(v)$ be the characteristic function for ξ_i

$$\varphi_i(v) = E[e^{+iv\xi_i}] \quad , i = 1, \dots, N$$

$$\varphi(v) = [\varphi_i(v)]^n = [1 + \tfrac{1}{2}v^2 + v^2 f(v)] \quad , \lim_{v\to 0} f(v) = 0$$

Let $u = v\sqrt{N}$, $\eta = \dfrac{\xi}{\sqrt{N}}$

$$\varphi_\eta(u) = \left[1 + \frac{u^2}{2N} + \frac{u^2}{N} f\left(\frac{u}{\sqrt{N}}\right)\right]^N$$

$$\cancel{\lim_{N\to\infty} \varphi_\eta(u) = e^{\frac{u^2}{2}[1+f(\frac{u}{\sqrt{N}})]}}$$

$$\lim_{N\to\infty} \varphi_\eta(u) = e^{\frac{u^2}{2}\left[1 + \lim_{N\to\infty} f\left(\frac{u}{\sqrt{N}}\right)\right]} = e^{\frac{u^2}{2}}$$

$$\boxed{e^{\frac{u^2}{2}} \leftrightarrow P\left(\frac{\xi}{\sqrt{N}}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2N}} \quad \text{or the Normal distribution}}$$

Note that several distributions reproduce themselves on addition of variables, but all resemble the Normal in some loose sense & still approach it in the limit.

<u>Note on limits in variable & transform domains:</u>

$$\left.\begin{array}{l} \varphi_N(v) \leftrightarrow P_N(x) \\ \lim_{N\to\infty} \varphi_N(v) = \varphi(v) \\ \lim_{N\to\infty} P_N(x) = P(y) \qquad , y = \sum^N x \end{array}\right\} \text{BUT: } \varphi(v) \not\leftrightarrow P(y)$$

~~However~~ This is because if Ψ is discrete for any $N < \infty$ and the limit of $P(y)$ is still discontinuous.

If we consider distribution functions, we do however get complete correspondence:

~~$P_N(x) \leftrightarrow \Phi_N(v)$~~

$$\begin{array}{ccc} P_N(x) & \xrightarrow{\lim N\to\infty} & P(y) \\ \updownarrow & & \updownarrow \\ \Phi_N(v) & \xrightarrow{\lim N\to\infty} & \Phi(v) \end{array}$$

Gaussian random variables:

Consider a set of statistically independent random variables $\{x_1, \dots, x_n\}$, all having a Normal distribution with mean 0 & variance σ^2.

$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}$$

Define $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ so $X_t = [x_1 \cdots x_n]$

Then

$$\boxed{P(X) = \frac{1}{(2\pi)^{n/2}} e^{\frac{-X_t X}{2}}}$$

Let $AX = Y$ where $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ & $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$

Now $X = A^{-1} Y$

so

$$P(Y) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{Y_t A_t^{-1} A^{-1} Y}{2}} |J|$$

$$J = \det \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & & \\ \frac{\partial x_n}{\partial y_1} & \cdots & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix} = \det A^{-1} = \frac{1}{\det A}$$

since $x_1 = a_{11}^{-1} y_1 + a_{12}^{-1} y_2 + \cdots$, $(a_{ij}^{-1}) = A^{-1}$

& then

$$\underline{P(Y) = \frac{1}{(2\pi)^{n/2} \det A} e^{-\frac{Y_t A_t^{-1} A^{-1} Y}{2}}}$$

Now $E(y_i y_j) = E\left[(a_{i1}x_1 + \cdots)(a_{j1}x_1 + \cdots)\right]$

$= a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots a_{in}a_{jn}$, since x_i + x_j are statistically independent.

$$= \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} a_{j1} \\ \vdots \\ a_{jn} \end{bmatrix}$$

Hence, the matrix $(E[y_i y_j])$ is

$$\begin{bmatrix} E(y_1^2) & E(y_1 y_2) & \cdots & \\ E(y_2 y_1) & \cdots & & \\ \cdot & & & \\ E(y_n y_1) & \cdots & & E(y_n^2) \end{bmatrix} = AA_t \equiv \Lambda = \underline{\text{covariance matrix}}$$

$$(AA_t)^{-1} = A_t^{-1}A^{-1} = \Lambda^{-1} \quad + \quad \det A = \sqrt{\det \Lambda}$$

so $$P(Y) = \frac{1}{(2\pi)^{n/2}(\det\Lambda)^{1/2}} e^{-\frac{Y_t \Lambda^{-1} Y}{2}}$$ for zero means for all y_i

More generally, $Y \to Y - M$
$\Lambda_{ij} = E[(y_i - m_i)(y_j - m_j)]$

and $$P(Y) = \frac{1}{(2\pi)^{n/2}(\det\Lambda)^{1/2}} e^{-\frac{(Y_t - M_t)\Lambda^{-1}(Y - M)}{2}}$$

Two gaussian random variables:

$$\Lambda = \begin{bmatrix} \sigma_1^2 & \mu_{12} \\ \mu_{21} & \sigma_2^2 \end{bmatrix}, \quad \mu_{12} = \mu_{21}$$

Define $\rho = \frac{\mu_{12}}{\sigma_1\sigma_2} \equiv$ correlation coefficient

$$\Lambda = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\det \Lambda = \sigma_1^2\sigma_2^2 - \rho^2(\sigma_1^2\sigma_2^2) = \sigma_1^2\sigma_2^2(1-\rho^2)$$

$$\Lambda^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ \frac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{bmatrix}$$

$$\begin{bmatrix} y_1-m_1 & y_2-m_2 \end{bmatrix} \Lambda^{-1} \begin{bmatrix} y_1-m_1 \\ y_2-m_2 \end{bmatrix} = \frac{\sigma_2^2(y_1-m_1)^2 + \sigma_1^2(y_2-m_2)^2 - 2\rho\sigma_1\sigma_2(y_1-m_1)(y_2-m_2)}{2\sigma_1^2\sigma_2^2(1-\rho^2)}$$

Properties of multi-dimensional gaussian r.v.'s:

(1) $p(X) = \frac{1}{(2\pi)^{n/2}|\Lambda|^{1/2}} \exp\left\{-\frac{1}{2}(X_t - M_t)\Lambda^{-1}(X-M)\right\}$

(2) $E[x_1x_2x_3x_4] = E(x_1x_2)E(x_3x_4) + E(x_1x_3)E(x_2x_4) + E(x_1x_4)E(x_2x_3)$

(3) If Λ is a diagonal matrix, then the variables are statistically independent.

(4) A linear transformation on X gives a new set Y of random variables which are also gaussian.

$$p(Y) = \frac{1}{(2\pi)^{n/2}|\Lambda|^{1/2}|A|} \exp\left\{-\frac{1}{2}(Y_t - N_t)\Lambda_y^{-1}(Y-N)\right\}$$

where $Y = AX$, $N = AM$, $\Lambda_y = A\Lambda_x A_t$

(5) Λ is of positive semi-definite quadratic form:

If $\Lambda = (\mu_{ij})$, then $\sum_{ij} \mu_{ij} a_i a_j = a_t \Lambda a \geq 0$, for any a_i, a_j.

Proof: $0 \leq E\left[\left(\sum_i x_i a_i\right)^2\right] = E\left[\sum_i \sum_j x_i x_j a_i a_j\right] = \sum_{ij} \mu_{ij} a_i a_j$.

(6) $M_x(i V) = E\left[e^{\sum i v_i x_i}\right] = E\left[e^{i V_t X}\right] = E\left[e^{i X_t V}\right]$

$$= \frac{1}{(2\pi)^{n/2}|\Lambda|^{1/2}} \iint \cdots \int \exp\left\{-\tfrac{1}{2}(X_t - M_t)\Lambda^{-1}(X-M) - i X_t V - i V_t X\right\} dX$$

Completing the square gives as the exponential term:

$$-\frac{(X_t - M_t)\Lambda^{-1}(X-M) - i V_t (X-M) - i (X_t - M_t) V - V_t \Lambda V}{2} + i V_t M - \tfrac{1}{2} V_t \Lambda V$$

$$= \underbrace{-\tfrac{1}{2}(X_t - M_t - i V_t \Lambda)\Lambda^{-1}(X - M - i \Lambda V)} + i V_t M - \tfrac{1}{2} V_t \Lambda V$$

In the integration, the first part will integrate to unity with the constants ahead of the integrals. The remainder is not a function of X, so

$$M_x(i V) = e^{i V_t M - \frac{1}{2} V_t \Lambda V}$$

(7) Given any positive semi-definite matrix Λ_x, there exists a matrix A such that $\Lambda_y = A \Lambda_x A_t$ is a diagonal matrix.

This corresponds to rotating coordinates so that they line up with the major & minor axes of the distribution. Scale is not changed.

E.g.: $\Lambda = \begin{bmatrix} \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 \end{bmatrix}$, $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \leftrightarrow \begin{array}{l} y_1 = x_1 \cos\theta + x_2 \sin\theta \\ y_2 = -x_1 \sin\theta + x_2 \cos\theta \end{array}$

$$A \Lambda A_t = \sigma^2 \begin{bmatrix} 1+\rho\sin 2\theta & \rho\cos 2\theta \\ \rho\cos 2\theta & 1+\rho\sin 2\theta \end{bmatrix} \xrightarrow[\theta=45^\circ]{} \begin{bmatrix} 1+\rho & 0 \\ 0 & 1+\rho \end{bmatrix} \Rightarrow A \to \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Sampling:

Suppose we take n samples from a distribution, and the values we measured were $x_1, \dots, x_n$

$$\boxed{\underline{\text{Sample mean}} \equiv m = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

Suppose the true mean is m_x; then how close to m_x is m?

$$E(m) = \frac{1}{N}\sum_{i=1}^{N} E(x_i) = \frac{1}{N}\sum_{i=1}^{N} m_x = m_x$$

This is an unbiased statistic; it is not affected by the measurements we make.

Let $s^2 = \sum_{i=1}^{N} \frac{(x_i - m)^2}{N}$ as an estimate of the variance

$$s^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - 2m \underbrace{\sum_{i=1}^{N} \frac{x_i}{N}}_{m} + m^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - m^2$$

$$E(s^2) = \frac{1}{N}\sum_{i=1}^{N} E(x_i^2) - E(m^2) = E(x^2) - \frac{1}{N^2} E\left[\left(\sum_{i=1}^{n} x_i\right)^2\right]$$

Assuming statistical independence:

$$E(s^2) = E(x^2) - \frac{1}{N^2}\left[N E(x^2) + (N^2 - N) m_x^2\right]$$

$$= \frac{N-1}{N}\left[E(x^2) - E^2(x)\right] = \underbrace{\frac{N-1}{N}\sigma_x^2}\ \rightarrow \text{a } \underline{\text{biased statistic}}$$

If we formed the quantity $\left(\frac{N}{N-1} s^2\right) \equiv q^2$, this would be an unbiased statistic.

Convergence:

$$\lim_{N\to\infty} E[(m-m_x)^2] = 0$$

or, $\text{l.i.m.}\,(m) = m_x$ = limit in the mean (m) with unity probability.
This does not imply certainty, only $P=1$.
This is convergence in the mean.

If we require that $P_r\{|m-m_x| < \epsilon\} \to 1$ as $N\to\infty$, this is called convergence in probability, or $P\lim m = m_x$.

Weak law of large numbers:

$$E[(m-m_x)^2] = \frac{1}{N^2}\sum\sum E(x_i x_j) - m_x^2 = \sigma^2$$

$$\sigma^2 = \frac{1}{N}E(x^2) - \frac{1}{N}m_x^2 = \frac{\sigma_x^2}{N} \quad \text{for stat ind r.v.'s.}$$

This does not, however, prove that if we consider a sequence of measurements of m, & average them, that this average converges to m_x.

Strong law of large numbers:

$$P_N\{|m-m_x| \geq \epsilon\} \leq \delta \quad \text{for all } N > N_\delta$$

I.e., the probability will go to zero & stay there beyond some N_δ.

Plausibility: Let x be a r.v. such that $x = \begin{cases} 1 & \text{if } A \\ 0 & \text{if not } A \end{cases}$

$$\left.\begin{aligned} m &= \frac{n(A)}{N} \\ \& \; E(m) &= P(A) \end{aligned}\right\}$$ The requirements that relative frequency limits converge to the probability require here that the strong law hold.

Example:

Sample is known to come from a normal distribution of unknown variance and mean of 0 or 1.

Samples taken are x_1, x_2, x_3

$$m = \frac{1}{3}(x_1 + x_2 + x_3)$$

$$s^2 = \frac{1}{3}\left[(x_1 - m)^2 + (x_2 - m)^2 + (x_3 - m)^2\right]$$

Let $y = \frac{m}{s}$; assume $m_x = 0$; $\sigma_x = \sigma_x$

$$\left.\begin{aligned} \text{Let } \xi_1 &= \tfrac{1}{\sqrt{3}}(x_1 + x_2 + x_3) \\ \xi_2 &= a x_1 + b x_2 + c x_3 \\ \xi_3 &= d x_1 + e x_2 + f x_3 \end{aligned}\right\} \text{orthogonal transformation so that } \xi_1, \xi_2, \xi_3 \text{ are independent.}$$

$$\frac{m}{s} = \frac{\frac{1}{\sqrt{3}}\xi_1}{\sqrt{\frac{1}{3}(x_1^2 + x_2^2 + x_3^2) - m^2}} = \frac{\frac{1}{\sqrt{3}}\xi_1}{\sqrt{\frac{1}{3}(\xi_1^2 + \xi_2^2 + \xi_3^2) - \frac{1}{3}\xi_1}} = \frac{\xi_1}{\sqrt{\xi_1^2 + \xi_2^2}}$$

Thus the sample mean & sample variance are statistically ~~independent~~ independent.

$$p(y) = \frac{1}{2(1+y^2)^{3/2}} \quad : \text{Student's distribution}$$

Regression:

$P(y|x) \to$ regression curve of y on x $\}$ locus of modes of $\begin{cases} P(y|x) \text{ over } x \\ P(x|y) \text{ over } y \end{cases}$
$P(x|y) \to$ " " " x on y $\}$

These two curves will in general be different.

Estimation:

Let $x_p(y)$ be our estimate of x given y.
Now we try to minimize our mean-square error:

$$\min_{x_p} E\left[(x_p(y) - x)^2\right]$$

$$= \min \iint [x_p(y) - x]^2 P(x|y) P(y)\, dx\, dy$$

$$= \min \int dy\, P(y) \int dx\, [x_p(y) - x]^2 P(x|y)$$

Since $P(y) \geq 0$, it will be sufficient to ~~minimize~~ find

$$\min_{x_p} \int [x_p(y) - x]^2 P(x|y)\, dx \qquad \text{for each } y.$$

$$\frac{d}{dx_p(y)} \int [x_p(y) - x]^2 P(x|y)\, dx = 2x_p(y) - 2\int x P(x|y)\, dx$$

$$\Rightarrow \boxed{x_p(y) = E(x|y)} \quad \text{or} \quad \int [x_p(y) - x] P(x|y)\, dx = 0$$ for an optimal (non-linear) solution to $x_p(\cdot)$

Via calculus of variations, we have $x_p(y)$ as our solution:

Let $\Psi = \iint [x_p(y) + \epsilon\eta(y) - x]^2 P(x|y) P(y)\, dx\, dy$

$$\left.\frac{\partial \Psi}{\partial \epsilon}\right|_{\epsilon=0} = 2\iint \eta(y) [x_p(y) - x] P(x|y) P(y)\, dx\, dy = 0 = 2\int dy\, P(y)\, \eta(y) \int dx\, [x_p(y) - x] P(x|y)$$

$$\Rightarrow \int [x_p(y) - x] P(x|y)\, dx = 0 \quad \text{as before}$$

This is the solution for an optimal filter in the mean square error sense. The problem now is to determine $P(x|yz\cdots)$ so that we can evaluate $E(x|yz\cdots) = x_p(y,z,\cdots)$.

Linear filter:

Suppose we restrict $x_p(y)$ to be a linear function of y; i.e. $x_p(y) = a + by$. Then

$$\min_{x_p} E\left[(x_p(y)-x)^2\right] = \min_{a,b} E\left[(a+by-x)^2\right]$$

$$= \min_{a,b}\left[a^2 + 2ab\,E(y) + b^2 E(y^2) - 2aE(x) - 2bE(xy) + E(x^2)\right]$$

Partial differentiation re: a and b gives

$$\left.\begin{aligned} 2a + 2bE(y) - 2E(x) &= 0 \\ 2aE(y) + 2bE(y^2) - 2E(xy) &= 0 \end{aligned}\right\} \text{ linear in } a \,\&\, b.$$

Solving for a & b:

$$b = \frac{-E(x)E(y) + E(xy)}{\sigma_y^2} = +\frac{\sigma_x}{\sigma_y} E(\xi\eta) \text{ where } \begin{cases} \xi = \dfrac{x - E(x)}{\sigma_x} \\ \eta = \dfrac{y - E(y)}{\sigma_y} \end{cases}$$

$$= +\rho\frac{\sigma_x}{\sigma_y}, \quad \rho \equiv E(\xi\eta)$$

$$a = E(x) + \rho\frac{\sigma_x}{\sigma_y}E(y) = \frac{E(x)\sigma_y^2 - E(y)\left[E(xy) - E(x)E(y)\right]}{\sigma_y^2}$$

So, $x_p(y) = E(x) + \rho\frac{\sigma_x}{\sigma_y}E(y) + \rho\frac{\sigma_x}{\sigma_y}y \Rightarrow \sigma_x\xi_p = +\rho\frac{\sigma_x}{\sigma_y}\eta \Rightarrow \boxed{\xi_p = +\rho\eta = \frac{\mu_{11}}{\sigma_x\sigma_y}\eta}$

Note that for linear filtering, prediction depends only on first & second moments. Also, since we get linear constraints on the filter parameters $(a+b)$, we can always solve for the filter.

For many y_i, let $x_p = a + \sum b_i y_i \rightarrow$ ~~$\int h(\tau)y(t-\tau)d\tau$~~ $\rightarrow \int h(z)x.(y-z)dz$.

Linear filtering continued:

$\xi_p = \rho\eta$ & also, $\eta_p = \rho\xi$ } best linear estimates.

If $\rho = 1$, $\xi_p = \eta$ and $E[(\xi_p - \xi)^2] = E[(\eta-\xi)^2] = 2(1-\rho) = 0.$

i.e., $\xi_p = \xi$ with probability 1. (?)

Suppose $y = x^2$ & we are given $p(x)$ symmetric about $x = 0$.

Now $E(xy) = E(x^3) = 0$

Thus x & y are completely dependent but uncorrelated.

Let $p(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\}$

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2 - 2\rho xy + \overbrace{y^2 - (1-\rho^2)y^2}^{\rho^2 y^2}}{2(1-\rho^2)}\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{(x-\rho y)^2}{2(1-\rho^2)}}$$

$E(x|y) = \rho y = x_p(y)$

$\Rightarrow$ For gaussian distributed variables, the optimal filter is a linear filter.

Random Processes:

Let $x = x(s,t)$ where s is a random event under $P(s)$. Each value of t defines a random variable $x_t(s)$. We have classified a number of random variables according to the parameter $\underline{t}$. ~~Thus~~

This defines a random process as a process whose output at time t is a sample from $x_t(s)$.

Present probability theory is based on relative frequencies of occurrence over repeated experiments. We cannot repeat a time-varying process in time & must therefore consider relative frequency over an ensemble of similar experiments being simultaneously performed. Then $P(x) = P_t(x[s])$ where $P_t(\cdot)$ is found as the relative frequencies for all x occurring at time t, ~~as~~ taken over an infinite ensemble of experiments.

To describe the process, we must be able to find $P(x_{t_1}, x_{t_2}, \dots, x_{t_n})$ for any instants t_i + t_n.

Stationarity:

$$P(x_{t_i}, x_{t_j}) = P(x_{t_k}, x_{t_m}) \quad \text{for } \underline{\underline{all}}\ t_i, t_j, t_k, t_m.$$

Gaussian process:

A gaussian process is one for which $P(x_t, x_T)$ is gaussian for all t, T. To describe a gaussian process, we need only the first & second moments.

Let $E[x_t x_T] = R(t,T)$.

(1) $R(t,T) = R(T,t)$

(2) $R(t_1,t_2) = R(t_1 - t_2) = R(T) = R(-T)$ for a stationary process

(3) $R(0) = E(x_t^2) \geq R(T)$, all T for a stationary process

(4) $R(T) = E[x_t x_{t+T}] \Rightarrow E(x_t)E(x_{t+T}) = E(x)E(x) = m^2$ for stat. process

(5) $|R(t_1,t_2)|^2 \leq R(t_1,t_1)R(t_2,t_2)$ for a non stationary process.

Time integrals of a random process:

Let $\quad y(s) = \int_a^b x(s,t)\,dt$

Now $\quad E(y) = \int_a^b E[x(s,t)]\,dt$

So, if $\int_a^b E[|x(s,t)|] < \infty$, then $y(s)$ exists for all s with the possible exception of a set of probability zero.

Theorem: If $x(s,t)$ is stationary, then

$$\lim_{T\to\infty} y(s,T) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(s,t)\,dt = y(s)$$ and exists with probability one.

Further, if the process is ~~the process will be said to be~~ ergodic, ~~if~~ then

$$y(s) = y \equiv E(x)\,;$$

i.e., ~~if~~ the time average equals the ensemble average.

Mean of a random process (stationary):

$$\mathcal{M}(s) \equiv \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(s,t)\,dt$$

$$m \equiv E[x(s,t)]$$

We want to show that $\mathcal{M}(s) = m$ with probability one. To do so, we must show that $E[\mathcal{M}(s)] = m$ and that $E[(\mathcal{M}-m)^2] = 0$.

$$E(\mathcal{M}) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} E[x(s,t)]\,dt = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} m\,dt = m.$$

$$E[(\mathfrak{M}-m)^2] = E(\mathfrak{M}^2) - m^2$$

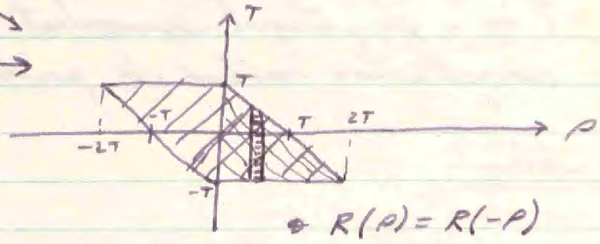
$$= \lim_{T\to\infty} \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} [E(x_t)E(x_\tau)]\,dt\,d\tau \;-\; m^2$$

$$= \lim_{T\to\infty} \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} [E(x_t)E(x_\tau) - m^2]\,dt\,d\tau$$

$$= \lim_{T\to\infty} \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} [R(t-\tau) - m^2]\,dt\,d\tau$$

$$= \lim_{T\to\infty} \frac{1}{4T^2}\int_{-T}^{T} d\tau \int_{-T-\tau}^{T-\tau} d\rho\,[R(\rho)-m^2] \qquad \text{where } \rho = t-\tau$$

$$= \lim_{T\to\infty} \frac{2}{4T^2}\int_{0}^{2T} d\rho \int_{-T}^{T-\rho} d\tau\,[R(\rho)-m^2] \quad \leftrightarrow$$



$\because R(\rho) = R(-\rho)$

$$= \lim_{T\to\infty} \frac{1}{2T^2}\int_{0}^{2T} d\rho\,[2T-\rho][R(\rho)-m^2]$$

$$= \lim_{T\to\infty} \frac{1}{T}\int_{0}^{2T} d\rho\,\left[1-\frac{\rho}{2T}\right][R(\rho)-m^2]$$

$$\le \lim_{T\to\infty} \frac{1}{T}\int_{0}^{2T} |R(\rho)-m^2|\,d\rho$$ since $1-\frac{\rho}{2T} \le 1$ over the range of integration

Thus, if $\int_{-\infty}^{\infty} |R(\tau)-m^2|\,d\tau < \infty$ & if $x(s,t)$ is stationary, the limit will be zero & $E[(\mathfrak{M}-m)^2] = 0$

so $\quad$ l.i.m. $\mathfrak{M} = m$

Proof that $R(0) \ge |R(\tau)| \ge 0$ for a stationary process:

$E[(x_t \pm x_\tau)^2] = 2R(0) \pm 2R(t-\tau) \ge 0 \quad$ since $R(\tau) = R(-\tau)$

~~Hence,~~ $\cancel{R(0) \ge R(t-\tau) = R(\rho)}$

Hence $\cancel{0 \le |R(0) \pm R(t-\tau)|}$ $|R(0)| \pm |R(\tau)| \ge 0$

or $\quad R(0) \ge \mp|R(\tau)| \ge |R(\tau)| \ge 0$

Property of $R(\tau)$:

$$E\left[\left(\int x(t)\,g(t)\,dt\right)^2\right] = E\left\{\iint x(t)\,x(\tau)\,g(t)\,g^*(\tau)\,dt\,d\tau\right\} \geq 0$$

for any complex function $g(\cdot)$.

$$= \iint E\{x(t)\,x(\tau)\}\,g(t)\,g^*(\tau)\,dt\,d\tau = \underline{\underline{\iint R(t-\tau)\,g(t)\,g^*(\tau)\,dt\,d\tau \geq 0}}$$

This is the integral equivalent of the requirement that the cross-correlation matrix be positive definite.

Problem:

$$x(s,t) = \begin{cases} +1 & \text{if \# of zero crossings between } t=0 \text{ \& } t=t \text{ is even} \\ -1 & \text{'' \quad odd.} \end{cases}$$

$$P_r\{K \text{ zero crossings in a time interval of length } t\} = P(K,t) = \frac{e^{-\lambda t}(\lambda t)^K}{K!}$$

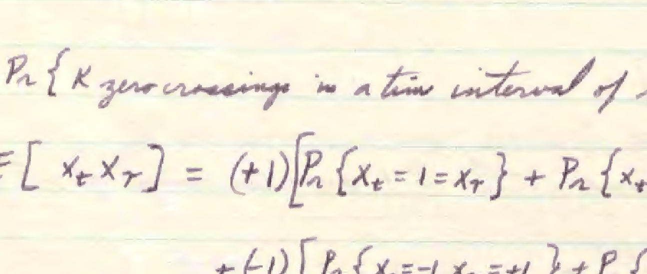
$$E[x_t x_\tau] = (+1)\left[P_r\{x_t = 1 = x_\tau\} + P_r\{x_t = -1 = x_\tau\}\right]$$

$$+(-1)\left[P_r\{x_t=-1, x_\tau=+1\} + P_r\{x_t=+1, x_\tau=-1\}\right]$$

$$= \sum_{K=2n} P(K,|t-\tau|) - \sum_{K=2n+1} P(K,|t-\tau|)$$

$$= e^{-\lambda|t-\tau|}\left[1 + \frac{\lambda^2|t-\tau|^2}{2!} + \frac{\lambda^4|t-\tau|^4}{4!} + \cdots\right] - e^{-\lambda|t-\tau|}\left[\frac{\lambda|t-\tau|}{1!} + \frac{\lambda^3|t-\tau|^3}{3!} + \cdots\right]$$

$$= e^{-\lambda|t-\tau|}\left[e^{-\lambda|t-\tau|}\right] = e^{-2|t-\tau|} = R(t-\tau)$$



<u>More properties of Correlation fns:</u>

$$R(\tau) \equiv E\{x_t x_{t+\tau}\}$$

$$\mathcal{R}(\tau) \equiv \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)\,x(t+\tau)\,dt \; .$$

For a stationary process, $R(\tau) = \mathcal{R}(\tau)$

Consider a periodic random process where

$$x(t) = \sum_{-\infty}^{\infty} a_n e^{j\frac{2\pi n t}{T}}$$

Then,

$$\mathcal{R}(\tau) = \frac{1}{T}\int_0^T x(t)\,x(t+\tau)\,dt$$

$$= \frac{1}{T}\int_0^T \sum_n a_n e^{j\frac{2\pi n t}{T}} \sum_m a_m e^{j\frac{2\pi m(t+\tau)}{T}}\,dt$$

$$= \frac{1}{T}\sum_n \int_0^T dt\, a_n e^{j\frac{2\pi n t}{T}} \sum_m a_m e^{j\frac{2\pi m t}{T}}\left[e^{j\frac{2\pi m \tau}{T}}\right]$$

$$= \sum_m |a_m|^2 e^{j\frac{2\pi m \tau}{T}}$$

Linear Systems:

System: Something with an input and an output, and some organized relationship between the two.

We will restrict ourselves to linear, time-invariant systems. These restrictions are defined in two axioms:

(1) If $\left.\begin{array}{l} x_1 \to y_1 \\ \text{and } x_2 \to y_2 \end{array}\right\}$ for $-\infty < t < \infty$,

then $a x_1 + b x_2 \to a y_1 + b y_2$ for $-\infty < t < \infty$

(2) $\left.\begin{array}{l} \text{If } x(t) \to y(t) \\ \text{then } x(t+T) \to y(t+T) \end{array}\right\}$ for $\left\{\begin{array}{l} -\infty < t < \infty \\ -\infty < T < \infty \end{array}\right\}$

We could use a differential equation representation of our systems, but these can be implicit, and we desire an explicit relation for the output in terms of the input. [e.g., $e = Ri + L\frac{di}{dt}$ is explicit if e is the output, implicit if i — there are many i's that will give the same e, but not vice versa.]

The basic foundation for linear system theory comes from (1) above. Namely,

if $x_k \to y_k$ then $x(t) = \sum_k a_k x_k \to \sum_k a_k y_k$

We want to choose a set of x_k's which can be used to represent every possible ~~and~~ input as a linear sum of the x_k's. To facilitate this, we would like the relation between x_k & y_k to be simple & the x_k's to be orthogonal over $-\infty < t < \infty$:

$$\int_{-\infty}^{\infty} x_i(t) x_j(t)\,dt = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

→ Such a set of functions is $X_s = e^{st}$, $-\infty < t < \infty$, where the range of our index $(k=s)$ has become uncountably infinite

→ The output for $x(t) = e^{st}$ can be found from our two ~~funds~~ axioms:

$\cancel{e^{st} \to y(t) \Rightarrow e^{s(t+\tau)} \to y(t+\tau) \Leftarrow e^{st}e^{s\tau} = y(t)e^{s\tau}}$

$$\left.\begin{array}{l} e^{st} \to y(t) \Rightarrow e^{s(t+\tau)} = e^{st}e^{s\tau} \to y(t+\tau) \\ \text{But } e^{st}e^{s\tau} \to y(t)e^{s\tau} \end{array}\right\} \Rightarrow y(t)e^{s\tau} = y(t+\tau);\ \text{all } t, \tau$$

In particular, for $t=0$, $y(\tau) = y(0)\,e^{s\tau}$

But $y(0)$ is just a constant (possibly dependent on s).
Hence,
$$\boxed{e^{st} \to H(s)e^{st}}$$
and we can represent the system by the function $H(s)$.

→ Now, how do we find $H(s)$?

Suppose we use the impulse response of the system $[h(t)]$, then

$$x(t) = e^{st} \to y(t) = \int_{-\infty}^{\infty} e^{s\tau} h(t-\tau)\,d\tau = \int_{-\infty}^{\infty} e^{st} e^{-s(t-\tau)} h(t-\tau)\,d(t-\tau)$$

$$= e^{st}\int_{-\infty}^{\infty} e^{-s\tau} h(\tau)\,d\tau \equiv H(s)e^{st}$$

Then we can define $\boxed{H(s) \equiv \int_{-\infty}^{\infty} e^{-s\tau} h(\tau)\,d\tau}$.

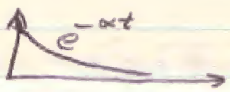
<u>Examples of system transforms</u>:

Suppose $h(t) = e^{-\alpha t}$ for $t \geq 0$ & $= 0$ for $t < 0$.

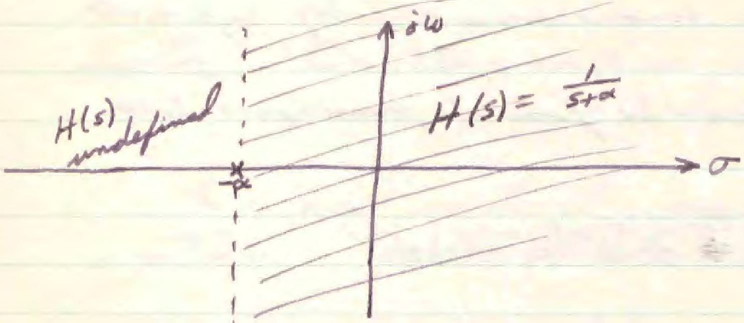
$$H(s) = \int_{0}^{\infty} e^{-(s+\alpha)t}\,dt = \frac{-1}{s+\alpha} e^{-(s+\alpha)t}\Big|_{t=0}^{t=\infty}$$

The evaluation exists at $t=\infty$ if & only if $Re\{s+\alpha\} > 0$;

i.e., ~~the~~ $H(s)$ is defined <u>only</u> for s such that $Re\{s\} > -\alpha$.

$h(t) =$ 

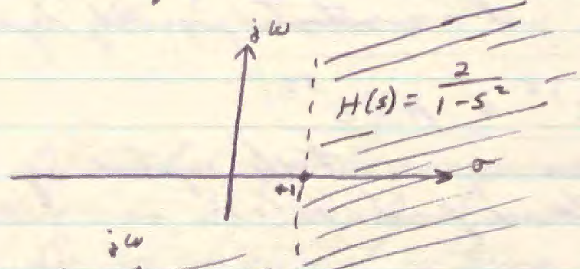
(cont.)



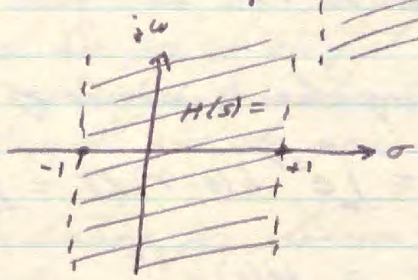
The region of convergence of $H(s)$ is that region of the s-plane where the input decay is slower than the natural decay.

In general, $H(s)$ will converge in a vertical strip in the s-plane:

$h(t) = e^{-t} - e^{+t} \longleftrightarrow$



$h(t) = e^{-|t|} \Longleftrightarrow$



etc.

Realizability, stability, etc.:

A linear system is realizable if $h(t) \equiv 0,\ t < 0$.

A system is stable if all bounded inputs give bounded outputs; i.e.,

$$\int_{-\infty}^{\infty} |h(t)|\,dt < \infty$$

In the s-plane, these requirements become:

Realizability: $H(s)$ converges in at least RHP; i.e., $H(s)$ blows up as an exponential at most.

Stability: $H(s)$ converges on the $j\omega$ axis. This must be since we require

$$|H(j\omega)| \le \int |h(t)||e^{j\omega t}|\,dt = \int |h(t)|\,dt.$$

Transform pair existence:

If $x(t)$ is such that (1) $\int_a^b |x(t)|\,dt < \infty$; $a, b \neq \pm\infty$

& (2) $\int_{-\infty}^{\infty} x(t)\,e^{-\sigma t}\,dt < \infty$ for some σ

then there exists a function $X(s)$ almost everywhere such that

$$x(t) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} X(s)\,e^{st}\,ds$$

and $$X(s) = \int_{-\infty}^{\infty} x(t)\,e^{-st}\,dt$$

If $x(t)$ is the input to a system with impulse response $h(t)$, then the regions of convergence of $X(s)$ and $H(s)$ must overlap for the ~~output~~ transform $Y(s)$ to exist. In this region, $x(t) = \sum_{\text{res of } X} X(s)\,e^{st}$, & $y(t) = \sum_{\text{res of } Y} Y(s)\,e^{st}$

Analytic Continuation:

To evaluate $h(t) = \frac{1}{2\pi j}\int_{\sigma-j\infty}^{\sigma+j\infty} H(s)\,e^{st}\,ds$, we maintain the form of $H(s)$ everywhere needed to evaluate the integral. This does <u>not</u> imply that $H(s)$ is an input-output relation over the entire s-plane.

<u>A representation for periodic random processes</u>:

Let $x(t)$ be a r.p. represented as

$$x(t) = \sum_{-\infty}^{\infty} x_n e^{j\frac{2\pi n t}{T}}$$; where the x_n are random variables, $x_{-n} = x_n^*$ so $x(t)$ is real.

Assume $E(x_n) = 0$
$E(x_n x_m^*) = \sigma_n^2 \, \delta_{nm}$

First, consider $E(x_n x_{-n}^*) = E(x_n^2) = E\left[(a+jb)^2\right] = E(a^2) - E(b^2) + 2j\,E(ab)$;

so, $E(a^2) = E(b^2)$ and $E(ab) = 0$ for $x_n = a + jb$.

$$E\left[x(t)\,x(t+\tau)\right] = \sum_{-\infty}^{\infty} E\{|x_n|^2\} e^{j\frac{2\pi n t}{T}} e^{j\frac{2\pi(-n)(t+\tau)}{T}}$$

$$= \sum \sigma_n^2 e^{j\frac{2\pi n \tau}{T}} = R(\tau)$$ which is periodic in τ.

<u>Note that a periodic r.p. represented thusly is wide-sense stationary.</u>

⟹ <u>Now, suppose we put ~~through~~ $x(t)$ through a linear filter</u>:

The output is $y(t)$ $\qquad \underset{(\text{r.p.})}{x(t)} \longrightarrow \boxed{h(t)} \longrightarrow y(t)$

$$y(t) = \sum x_n H\left(j\tfrac{2\pi n}{T}\right) e^{j\frac{2\pi n t}{T}}$$ by virtue of <u>superposition</u>.

Now, multiplying ~~each~~ x_n by $H\left(j\frac{2\pi n}{T}\right)$ (a constant) does not change any relevant properties of x_n; so

$$E\left\{x_n H\left(j\tfrac{2\pi n}{T}\right) x_m^* H\left(-j\tfrac{2\pi m}{T}\right)\right\} = \sigma_n^2 \left|H\left(j\tfrac{2\pi n}{T}\right)\right|^2 \delta_{mn}$$

Hence, $$R_y(\tau) = \sum_{-\infty}^{\infty} \sigma_n^2 \left|H\left(j\tfrac{2\pi n}{T}\right)\right|^2 e^{-j\frac{2\pi n \tau}{T}}$$; also periodic + wide-sense stationary

<u>Definition of a periodic random process</u>:

A periodic r.p. is one which is wide sense stationary and whose correlation function $R(\tau)$ is periodic in τ.

If a process satisfies these two requirements, then almost every sample function is a periodic function of time:

<u>Proof</u>: If a r.p. has a periodic correlation function & zero mean, it has a periodic formulation in time.

Let $R_x(\tau) = \sum_{-\infty}^{\infty} \sigma_n^2 e^{-j\frac{2\pi n\tau}{T}}$ & let $x(t)$ be real.

Let $x_n \equiv \frac{1}{T}\int_0^T x(t) e^{j\frac{2\pi n t}{T}} dt$ define x_n for each n.

Now, $E(x_n) = \frac{1}{T}\int_0^T E[x(t)] e^{j\frac{2\pi n t}{T}} dt = 0$

$$E(x_n x_m^*) = \frac{1}{T^2}\int_0^T\int_0^T E[x(t)x(\tau)] e^{j\frac{2\pi n t}{T}} e^{-j\frac{2\pi m \tau}{T}} dt\, d\tau$$

$$= \frac{1}{T^2}\int_0^T\int_0^T \sum_{k=-\infty}^{\infty} \sigma_k^2 e^{-j\frac{2\pi k(t-\tau)}{T}} e^{j\frac{2\pi n t}{T}} e^{-j\frac{2\pi m\tau}{T}} dt\, d\tau$$

$$\underline{\underline{= \sigma_n^2 \delta_{nm}}}$$

We want to show that $x(t)$ can be represented as $\sum x_n e^{j\frac{2\pi n t}{T}}$; i.e., $x(\cdot)$ is necessarily periodic:

$$E\left\{\left[x(t) - \sum_{-\infty}^{\infty} x_n e^{-j\frac{2\pi n t}{T}}\right]^2\right\} = E[x^2(t)] - 2\sum E[x_n x(t)] e^{j\frac{2\pi n t}{T}}$$

$$+ \sum_n \sum_m E(x_n x_m) e^{-j\frac{2\pi n t}{T}} e^{-j\frac{2\pi m t}{T}}$$

$$= R_x(0) + \sum_{-\infty}^{\infty} \sigma_n^2 - 2\sum E[x_n x(t)] e^{-j\frac{2\pi n t}{T}}$$

$$= 2R_x(0) - 2\sum E\left[x(t)\frac{1}{T}\int_0^T x(\tau) e^{j\frac{2\pi n\tau}{T}} d\tau\right] e^{-j\frac{2\pi n t}{T}}$$

$$= 2R_x(0) - 2\sum e^{-j\frac{2\pi n t}{T}} \frac{1}{T}\int_0^T E\{x(\tau) x(t)\} e^{j\frac{2\pi n\tau}{T}} d\tau$$

$$= 2R_x(0) - 2\sum_n e^{-j\frac{2\pi n t}{T}} \frac{1}{T}\int_0^T e^{j\frac{2\pi n\tau}{T}} \sum_k \sigma_k^2 e^{-j\frac{2\pi k(t-\tau)}{T}} d\tau$$

$$= 2R_x(0) - 2\sum_n \sigma_n^2 = 2R_x(0) - 2R_x(0) = 0.$$

(over)

Thus we have shown that ~~[illegible]~~ with probability one, $x(t)$ is represented as a periodic time process.

<u>Band-limited white random process</u>:

Let $x(t) = \sum_{-\infty}^{\infty} x_n \dfrac{\sin 2\pi w(t-\frac{n}{2w})}{2\pi w(t-\frac{n}{2w})}$

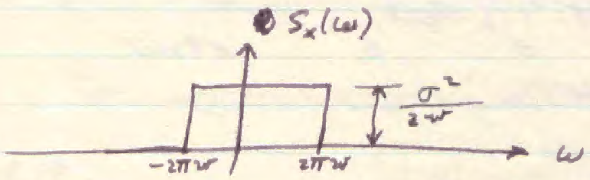
$E(x_n) = 0 \quad ; \quad E(x_n x_m) = \sigma^2 \delta_{mn}$

Then $R_x(t,\tau) = E[x(t)x(\tau)] = \sum \sigma^2 \dfrac{\sin 2\pi w(t-\frac{n}{2w}) \sin 2\pi w(t-\frac{n}{2w})}{2\pi w(t-\frac{n}{2w})\, 2\pi w(t-\frac{n}{2w})}$

$= \sigma^2 \dfrac{\sin 2\pi w(t-\tau)}{2\pi w(t-\tau)}$ (on faith)

& the process is wide sense stationary

$S_x(\omega) \equiv \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega\tau} d\tau$ is band-limited white:



If we let $x_n \equiv x(\frac{n}{2w})$, then

$E(x_n) = 0$ and $E(x_n x_m) = \sigma^2 \delta_{mn}$

& $x(t) = \text{l.i.m.} \sum_{-\infty}^{\infty} x_n \dfrac{\sin 2\pi w(t-\frac{n}{2w})}{2\pi w(t-\frac{n}{2w})}$

Karhunen - Loeve Theorem:

Suppose we have a r.p. $x(t)$ with zero-mean, and correlation function $R(t,\tau)$ which is not necessarily stationary. We want to represent $x(t)$ in the interval $0 < t < T$ as $x(t) = \sum_0^\infty x_n \varphi_n(t)$ where x_n are r.v. & $\varphi_n(\cdot)$ are definite funtions; also,

$$E(x_n) = 0 \quad \& \quad E(x_n x_m) = \sigma_n^2 \delta_{mn}$$

φ_n & φ_m are orthonormal: $\int_0^T \varphi_n(t)\varphi_m(t)\,dt = \delta_{nm}$

We now prove that such a representation exists:

$$\text{Let} \quad x_n \equiv \int_0^T x(t)\varphi_n(t)\,dt$$

$$\text{Now,} \quad E(x_n) = \int_0^T E[x(t)]\varphi_n(t)\,dt = 0$$

$$E(x_n x_m) = \int_0^T\!\!\int_0^T E[x(t)x(\tau)]\varphi_n(t)\varphi_m(\tau)\,dt\,d\tau$$

$$= \int_0^T\!\!\int_0^T R_x(t,\tau)\varphi_n(t)\varphi_m(\tau)\,dt\,d\tau$$

We must now select the φ's such that $E(x_n x_m) = \sigma_n^2 \delta_{mn}$;

i.e., $$\boxed{\int_0^T R_x(t,\tau)\varphi_n(t)\,dt = \sigma_n^2 \varphi_n(\tau)}$$

The solution to this equation is discussed in an appendix to Davenport & Root.

It is similar to finding a matrix $x]$ (here infinitely long) such that $A\,x] = \kappa\,x]$ where A is a square matrix.

The important question is whether the φ's exist or not, not what they are for some particular process.

→ If $R_x(t,\tau)$ is strictly positive definite, then the φ's do exist and form a complete set, so that

$$R_x(t,\tau) = \sum_0^\infty \sigma_n^2 \varphi_n(t)\varphi_n(\tau)$$

Statistical independence & linear filters:

$$x(t) \longrightarrow \boxed{h(t)} \longrightarrow y(t) \quad ; \quad y(t) = \int_{-\infty}^{\infty} h(\tau - t)\,x(t)\,dt$$

$$E\{y(t)\,y(\mu)\} = \iint E\{x(t)\,x(\mu)\}\,h(\tau - t)\,h(\nu - \mu)\,dt\,d\mu$$

$$= R_y(t, \mu) = \iint R_x(t - \mu)\,h(\tau - t)\,h(\nu - \mu)\,dt\,d\mu$$

$$= \iint R_x(\tau - \rho + \xi - \nu)\,h(\rho)\,h(\xi)\,d\rho\,d\xi, \quad \text{where} \quad \rho \equiv \tau - t \quad d\rho = -dt$$

$$\xi = \nu - \mu \quad d\xi = -d\mu$$

$$= \iint R_x(\tau - \nu + \xi - \rho)\,h(\rho)\,h(\xi)\,d\rho\,d\xi = \underline{R_y(\tau - \nu)}.$$

which is dependent only on $\tau - \nu$. Thus if $x(\cdot)$ is wide-sense stationary, so is $y(\cdot)$. If samples of $x(\cdot)$ taken at different times are statistically independent, then so are samples of the signal after it is passed through a linear filter.

Spectral power & correlation function transforms:

Transforming both sides of the above equation

$$R_y(\tau) = \iint R_x(\tau + \xi - \rho)\,h(\rho)\,h(\xi)\,d\rho\,d\xi$$

gives

$$S_y(\omega) = |H(j\omega)|^2\,S_x(\omega)$$

where $\quad S(\omega) \equiv \int_{-\infty}^{\infty} R(\tau)\,e^{-j\omega\tau}\,d\tau \quad \& \quad R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega)\,e^{j\omega\tau}\,d\omega$

The average power of the output is

$$R_y(0) = \frac{1}{2\pi}\int S_y(\omega)\,d\omega = \frac{1}{2\pi}\int S_x(\omega)\,|H(j\omega)|^2\,d\omega = \text{avg power (watts)}$$

$$S_y(\omega) = \text{avg power per cycle (watts/cycle)}$$

Now, suppose we consider a *finite* time interval, $-T < t < T$.

Let $X_T(\omega) \leftrightarrow x(t)$ in the interval.

Now, the actual measured average power in the input over this interval is $\frac{|X_T(\omega)|^2}{2T}$;

does $\lim_{T\to\infty} \frac{|X_T(\omega)|^2}{2T} \to S_x(\omega)$?: We might suspect so, but the limit does not converge:

Proof: First, are the means identical in the limit $T\to\infty$?

$$\cancel{E\left\{\frac{|X_T(\omega)|^2}{2T}\right\} = \frac{1}{2T}\int_{-T}^{T}\int E\{x(t)x(\tau)\}e}$$

$$\boxed{X_T(\omega) = \int_{-T}^{T} x(t)e^{-j\omega t}\,dt}$$

$$E\left\{\frac{|X_T(\omega)|^2}{2T}\right\} = \frac{1}{2T}E\{X_T(\omega)X_T^*(\omega)\} = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} E\{x(t)x(\tau)\}e^{-j\omega t}e^{j\omega\tau}\,dt\,d\tau$$

$$= S_T(\omega) = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} R_x(t-\tau)e^{-j\omega(t-\tau)}\,dt\,d\tau \neq \cancel{\frac{1}{2T}\int_{-T}^{T} S_T(\omega)d\tau}$$

Now, in the limit $T\to\infty$, the t-integration gives $S(\omega)$ and the τ-integration drops out the $2T$ denominator & the means of the two power measures are the same in the limit.

Does the variance go to zero?: The variance ~~of~~ of $S_T(\omega)$ is ~~the difference in $|X_T(\omega)|^2$ & $S_x(\omega)$ is~~

$$\text{var} = E\left\{\left[\frac{1}{2T}|X_T(\omega)|^2 - S_x(\omega)\right]^2\right\}$$

Let $S_T(\omega) \equiv \frac{1}{2T}|X_T(\omega)|^2$. We now want to show that

$$E\{[S_T(\omega)-S_x(\omega)]^2\} \geq 0 \quad \text{or} \quad E[S_T^2(\omega)] \geq S_x^2(\omega):$$

Now, $E[S_T(\omega)] = \frac{1}{2T}\int_{-T}^{T}\int_{-T}^{T} E\{x_t x_\tau\}e^{-j\omega t}e^{j\omega\tau}\,dt\,d\tau$

so $E[S_T^2(\omega)] = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T}\int_{-T}^{T} E\{x_t x_\tau x_u x_v\}e^{-j\omega t}e^{j\omega\tau}e^{-j\omega u}e^{j\omega v}\,dt\,d\tau\,du\,dv.$

(over)

For a gaussian process, we have shown that

$$E\{x(t)\,x(\tau)\,x(\mu)\,x(\nu)\} = R(t-\tau)R(\mu-\nu) + R(t-\nu)R(\mu-\tau) + R(t-\mu)R(\tau-\nu)$$

<u>so</u> $$E\{S_T^2(\omega)\} = \frac{1}{4T^2}\iiiint_{-T}^{T}\left[R(t-\tau)R(\mu-\nu)+R(t-\nu)R(\mu-\tau)+R(t-\mu)R(\tau-\nu)\right]e^{-j\omega(t+\mu-\tau-\nu)}\,dt\,d\tau\,d\mu\,d\nu$$

$$\cancel{\tfrac{1}{2}\,2E^2\{S_T(\omega)\} + \frac{1}{2T}\iint_{-T}^{T} R(t-\mu)\,e^{j}}$$

$$= 2E^2\{S_T(\omega)\} + \frac{1}{4T^2}\iiiint R(t-\mu)R(\tau-\nu)\,e^{-j\omega(t+\mu)}e^{j\omega(\tau+\nu)}\,dt\,d\tau\,d\mu\,d\nu$$

$$= 2E^2\{S_T(\omega)\} + \left|\frac{1}{2T}\iint R(t-\mu)\,e^{-j\omega(t+\mu)}\,dt\,d\mu\right|^2$$

Now in the limit $T\to\infty$, $E^2[S_T(\omega)] \to S_x^2(\omega)$, so:

$$\lim_{T\to\infty} E[S_T^2(\omega)] = 2S_x^2(\omega) + \lim_{T\to\infty}\left|\frac{1}{2T}\iint_{-T}^{T} R(t-\mu)\,e^{-j\omega(t+\mu)}\,dt\,d\mu\right|^2$$

& therefore, $\lim_{T\to\infty} E[S_T^2(\omega)] \geq 2S_x^2(\omega)$ QED

$\boxed{\text{\& thus, } S_T(\omega) \text{ does not converge to } S_x(\omega) \text{ with zero variance as } T\to\infty}$

$$\begin{tikzcd}
R_T(\tau) \arrow[r, leftrightarrow] \arrow[d, "T\to\infty"] & S_T(\omega) \arrow[d, "\times", "\text{does not converge}"'] \\
R(\tau) \arrow[r] & S(\omega)
\end{tikzcd}$$

Here is another example where convergence in the time domain does not yield similar convergence in the frequency domain.

Now consider

$$G_T(\xi) \equiv \int_{-\infty}^{\xi} S_T(\omega)\,d\omega$$

$$\lim_{T\to\infty} E[G_T(\xi)] = \int_{-\infty}^{\xi} S_x(\omega)\,d\omega$$

$$E[G_T^2(\xi)] = \int_{-\infty}^{\xi}\int_{-\infty}^{\xi} E[S_T(\omega) S_T(\rho)]\,d\omega\,d\rho$$

$$= \int_{-\infty}^{\xi}\int_{-\infty}^{\xi} d\omega\,d\rho\ \frac{1}{4T^2}\iiiint_{-T}^{T} E\{x_t x_\tau x_\mu x_\nu\}\, e^{-j\omega t} e^{j\omega\tau} e^{-j\rho\mu} e^{j\rho\nu}\,dt\,d\tau\,d\mu\,d\nu$$

$$= \iint_{-\infty}^{\xi} d\omega\,d\rho\ \frac{1}{4T^2}\iiiint_{-T}^{T}\left[R(t-\tau)R(\mu-\nu) + R(t-\nu)R(\mu-\tau) + R(t-\mu)R(\tau-\nu)\right] e^{-j\omega(t-\tau)} e^{-j\rho(\mu-\nu)}\,dt\,d\tau\,d\mu\,d\nu$$

$$= \iint_{-\infty}^{\xi} d\omega\,d\rho\left[\overset{E\{S_T(\omega)\}E\{S_T(\rho)\}}{\cancel{E\{S_T(\omega)S_T(\rho)\}}} + \left|\frac{1}{2T}\iint_{-T}^{T} R(t-\nu)e^{-j\omega t} e^{j\rho\nu}\,dt\,d\nu\right|^2 + \left|\frac{1}{2T}\iint_{-T}^{T} R(t-\mu) e^{j\omega t} e^{-j\rho\mu}\,dt\,d\mu\right|^2\right]$$

These last two integral terms will go to zero as $T\to\infty$ due to the $\frac{1}{2T}$ term (the integrals will stay finite).

Hence

$$\cancel{E[G_T^2(\xi)] = \iint_{-\infty}^{\xi} E[S_T(\omega) S_T(\rho)]\,d\omega\,d\rho = E}$$

$$E[G_T^2(\xi)] = \iint_{-\infty}^{\xi} E[S_T(\omega)]\,E[S_T(\rho)]\,d\omega\,d\rho = E^2[G_T(\xi)]$$

& $\operatorname{var}\{G_T(\xi)\} \to 0$ so $\lim_{T\to\infty} G_T(\xi) = \int_{-\infty}^{\xi} S_x(\omega)\,d\omega = G(\xi$.

$$\begin{tikzcd} R_T(\tau) \arrow[r] \arrow[d] & G_T(\xi) \arrow[d] \\ R(\tau) \arrow[r] & G(\xi) \end{tikzcd}$$

Thus, again the integral of the transform does converge in the limit.

Gaussian narrow band process :

First we define a complex r.p. $z(t) = x(t) + j\,y(t)$ where $x(t)$ and $y(t)$ are jointly gaussian, $E(x_t) = E(y_t) = 0$, $E(x_t x_\tau) = E(y_t y_\tau) = R(t-\tau)$, and $E(x_t y_\tau) = 0$. For simplicity, assume the spectrum is symmetric. ~~about a "carrier" frequency~~

$$E[z_t z_\tau^*] = E[x(t)x(\tau)] + E[y(t)y(\tau)] + jE[\overset{0}{x(\tau)y(t)}] - jE[\overset{0}{x(t)y(\tau)}]$$

$$= 2R(t-\tau).$$

Now, define a new random process by

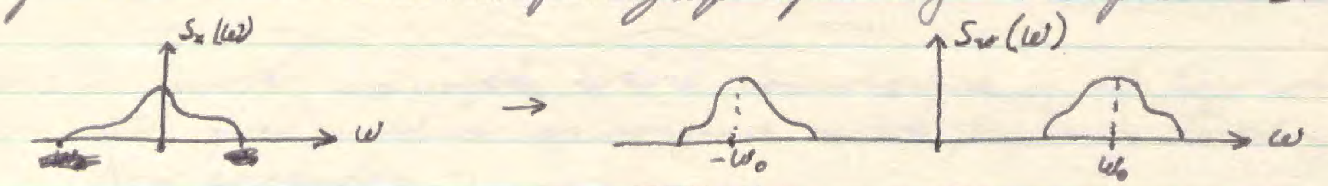
$$w(t) = \mathrm{Re}\{z(t)\,e^{j\omega_0 t}\} = x_t \cos\omega_0 t - y_t \sin\omega_0 t.$$

This process is still gaussian as the cos & sin time fns are just 'constant' multipliers.

$$E[w_t w_\tau] = R(t-\tau)[\cos\omega_0 t \cos\omega_0\tau + \sin\omega_0 t \sin\omega_0\tau]$$

$$= R(t-\tau)\cos\omega_0(t-\tau)$$

Thus, the process is wide sense stationary & since it is a gaussian process, completely stationary [recall that the correlation function (2nd moments) completely specifies a gaussian process].



Suppose $w(t) = E(t)\cos[\omega_0 t + \varphi(t)]$

$$= E(t)\cos\varphi(t)\cos\omega_0 t - E(t)\sin\varphi(t)\sin\omega_0 t$$

$\Leftrightarrow$ This process can be realized in the above formulation if

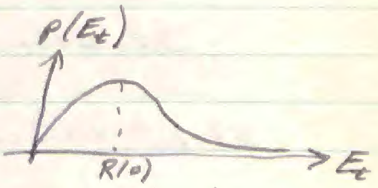
$$\left.\begin{aligned} x(t) &= E(t)\cos\varphi(t) \\ y(t) &= E(t)\sin\varphi(t) \end{aligned}\right\} \quad \begin{aligned} E(t) &= \sqrt{x^2(t) + y^2(t)} \\ \varphi(t) &= \tan^{-1}[y(t)/x(t)] \end{aligned}$$

$E(t) \leftrightarrow$ envelope of the process in time (AM)
$\varphi(t) \leftrightarrow$ phase of process (FM or PM)

E.g., suppose $\quad p(E_t, \varphi_t) = \dfrac{E(t)}{2\pi R(0)} e^{-\frac{x_t^2 + y_t^2}{2R(0)}}$

$$p(E_t, \varphi_t) = \frac{E(t)}{2\pi R(0)} e^{-\frac{E^2(t)}{2R(0)}}$$

$$\int_0^{2\pi} p(E_t, \varphi_t)\, d\varphi_t = \frac{E(t)}{R(0)} e^{-\frac{E^2(t)}{2R(0)}} = p(E_t)$$



Rayleigh dist

$$p(\varphi_t) = \int_0^{\infty} p(E_t, \varphi_t)\, dE_t = \frac{1}{2\pi}, \quad 0 \le \varphi_t < 2\pi$$

Thus we see that at any one instant of time, the phase and envelope amplitude are statistically independent.

Now, consider two time instants:

$$p[E_{t_1}, E_{t_2}, \varphi_{t_1}, \varphi_{t_2}] = p[E_1, E_2, \varphi_1, \varphi_2] \quad ; \quad \tau = t_2 - t_1 > 0$$

$$p(x_1, x_2, y_1, y_2) = \frac{1}{(2\pi)^2 |\Lambda|^{1/2}} e^{-\frac{[\underline{x}, \Lambda^{-1}\underline{x}]}{2}} = p(\underline{x}) \quad \text{where } \underline{x} = (x_1, x_2, y_1, y_2)$$

$$\Lambda = \begin{bmatrix} R(0) & R(\tau) & 0 & 0 \\ R(\tau) & R(0) & 0 & 0 \\ 0 & 0 & R(0) & R(\tau) \\ 0 & 0 & R(\tau) & R(0) \end{bmatrix} \quad ; \quad |\Lambda|^{1/2} = R^2(0) - R^2(\tau)$$

so,

$$p(E_1, E_2, \varphi_1, \varphi_2) = \frac{E_1 E_2}{(2\pi)^2 [R^2(0) - R^2(\tau)]} e^{-\frac{R(0)[E_1^2 + E_2^2] - 2R(\tau) E_1 E_2 \cos(\varphi_1 - \varphi_2)}{2[R^2(0) - R^2(\tau)]}}$$

After much hair, we find that the envelope process is not independent of the phase process.

Noise: Suppose to the preceding process we added a noise factor $P\sin(\omega_0 t+\psi)$; i.e., ~~random~~ sine wave of carrier frequency but random ~~noise~~ phase:

Let $u(t) = w(t) + P\sin(\omega_0 t + \psi)$; P is constant, ψ is r.v.

$$u(t) = x'(t)\cos\omega_0 t - y'(t)\sin\omega_0 t$$

$$= [x(t) + P\cos\psi]\cos\omega_0 t - [y(t) + P\sin\psi]\sin\omega_0 t$$

We would now like to find $p(x'_t, y'_t, \psi)$

$$p(x'_t, y'_t, \psi) = \frac{1}{(2\pi)^2 R(0)} \exp\left\{-\frac{(x' - P\cos\psi)^2 + (y' - P\sin\psi)^2}{2R(0)}\right\}$$

$$p(E, \varphi, \psi) = \frac{E}{(2\pi)^2 R(0)} \exp\left\{-\frac{(E\cos\varphi - P\cos\psi)^2 + (E\sin\varphi - P\sin\psi)^2}{2R(0)}\right\}$$

$$= \frac{E}{(2\pi)^2 R(0)} e^{-\frac{E^2}{2R(0)}} e^{-\frac{P^2}{2R(0)}} e^{\frac{EP\cos(\varphi - \psi)}{R(0)}}$$

Now, $$p(E, \psi) = \int_0^{2\pi} p(E, \varphi, \psi)\, d\varphi = \frac{E}{(2\pi)^2 R(0)} e^{-\frac{E^2+P^2}{2R(0)}} \int_0^{2\pi} e^{\frac{EP\cos(\varphi-\psi)}{R(0)}} d\varphi$$

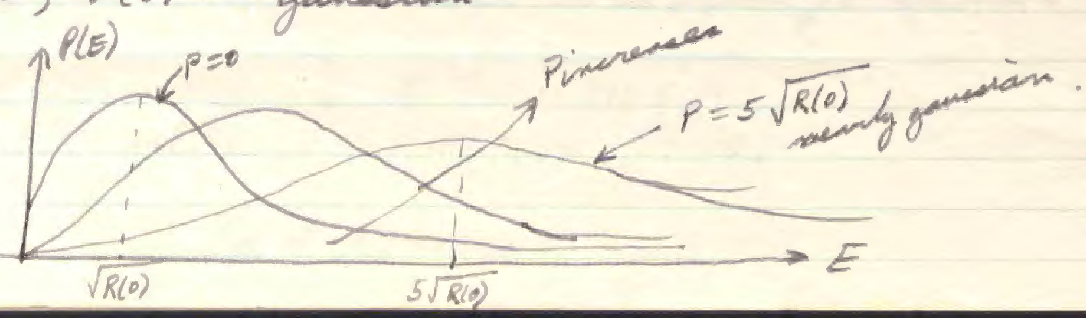
$$\boxed{p(E, \psi) = \frac{E}{(2\pi) R(0)} e^{-\frac{E^2+P^2}{2R(0)}} I_0\left(\frac{EP}{R(0)}\right)} \Bigg\} \text{ from } \boxed{\frac{1}{2\pi}\int_0^{2\pi} e^{-(a\cos\varphi + b\sin\varphi)} d\varphi = I_0\left(\sqrt{a^2+b^2}\right)}$$

$$\boxed{p(E) = \frac{E}{R(0)} e^{-\frac{E^2+P^2}{2R(0)}} I_0\left(\frac{EP}{R(0)}\right)}$$

///// Rices distribution ////

As $P \to 0$, $p(E) \to$ Rayleigh distribution

$P \to \infty$, $p(E) \to$ gaussian



Shot noise:

Poisson distribution:

Consider a time interval of T seconds, divided into M discrete intervals of duration ΔT; suppose the average rate of occurence of events (e.g., pulses of current due to random electrons) is $\bar{N}$ per second, and the probability of an event in any interval ΔT is

$$P = \bar{N}\Delta T$$

Then the probability of K events in the interval T is

$$P_M(K,T) = \binom{M}{K} P^K (1-P)^{M-K} \leftrightarrow M_M(j\nu) = \sum_{K=0}^{M} P_M(K,T)e^{j\nu K}$$

Now, let $M \to \infty$, $\Delta T \to 0$ so that $MP = \bar{N}T$, i.e., so that $M\Delta T = T$.

Assume K can be only 0 or 1; then

$$M_M(j\nu) = \sum_{K=0}^{M} \binom{M}{K}(Pe^{j\nu})^K(1-P)^{M-K} = (Pe^{j\nu} + 1 - P)^M \qquad (?)$$

$$= \left[1 - \frac{\bar{N}T}{M}(1-e^{+j\nu})\right]^M \xrightarrow[M\to\infty]{} e^{-\bar{N}T(1-e^{j\nu})}$$

$$M(j\nu) = e^{-\bar{N}T}\left[1 + \bar{N}Te^{j\nu} + \frac{(\bar{N}T)^2}{2!}e^{2j\nu} + \dots\right]$$

$$\Rightarrow \quad P(K,T) = \frac{e^{-\bar{N}T}(\bar{N}T)^K}{K!} \quad \text{— Poisson distribution}$$

Differential derivation of Poisson distribution:

Consider a process in continuous time where discrete events occur individually at random. Simultaneous events have zero probability. We want $P(K,t)$, the probability of K events in a time interval t:

$$P(1,dt) = a\,dt \quad \& \quad P(0,dt) = 1 - a\,dt$$

$$P_K(t+dt) = P_K(t)\left[1 - a\,dt\right] + P_{K-1}(t)\,a\,dt$$

or $$\frac{P_K(t+dt)-P_K(t)}{dt} + aP_K(t) = aP_{K-1}(t)$$

$$\boxed{\frac{d}{dt}P_K(t) + aP_K(t) = aP_{K-1}(t)}$$

We now need to find $P_0(t)$:

$$P_0(t+dt) = P_0(t)[1-a\,dt] \rightarrow \frac{d}{dt}P_0(t) + aP_0(t) = 0$$

Assuming that ~~$P_0(t)$~~ $P_0(0)=1$, $t \geq 0$, we have

$$\underline{P_0(t) = e^{-at}}. \qquad \text{Then } \frac{d}{dt}P_1(t) + aP_1(t) = ae^{-at}$$

or $$P_1(t) = (at)e^{-at}$$

etc.: so $$\underline{P_K(t) = \frac{(at)^K e^{-at}}{K!}}$$ same as before.

This is the probability that in a time t, K events will occur; thus $\sum_{K=0}^{\infty} P_K(t) = 1$ for all t.

Now, is this also the probability density for the time t required for K events? No.

The probability distribution for the time τ between events is given by

$$P(\tau)\,d\tau = P_0(\tau)\,a\,d\tau \rightarrow \underline{P(\tau) = ae^{-a\tau}}$$

Shot noise:

Let $i_e(t-t_K)$ be the current due to an electron hitting the plate of a tube (e.g.) at time t_K. Then

$$i(t) = \sum_{K=1}^{n} i_e(t-t_K) \quad ;$$

$$\langle i(t)\rangle = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} i(t)\,dt = \lim_{T\to\infty} \frac{N}{2T}\int i_e(t)\,dt$$

$$\langle i(t)\rangle = \bar{n}\int i_e(t)\,dt = \bar{n}e$$

where K = avg # of electrons in an interval $2T$
$\bar{n}$ = avg # of electrons per second
e = charge on one electron

$$E[i(t)] = \iint\cdots\int \sum_{K=0}^{\infty}\sum_{n=1}^{K} i_e(t-t_n)\,P(t_1,\dots,t_K,K)\,dt_1\,dt_2\cdots dt_K$$

$$= \sum_{K=0}^{\infty} Ke\,T^{K-1}\left(\frac{1}{T}\right)^K P(K,T) = \sum_{K=0}^{\infty} \frac{Ke}{T}\frac{(\bar{n}T)^K}{K!}e^{-\bar{n}T}$$

$$= \frac{e}{T}\sum_{K=0}^{\infty} K\,P(K,T) = \frac{e}{T}(\bar{n}T) = \bar{n}e = E[i(t)]$$

Thus the time & ensemble averages of $i(t)$ are both $\bar{n}e$. The correlation function for $i(t)$ is:

$$E[i(t)\,i(t+\tau)] = \sum_{K=0}^{\infty}\iint\cdots\int\sum_{i=0}^{K}\sum_{j=0}^{K} i_e(t-t_i)\,i_e(t+\tau-t_j)\,P(t_1,\dots,t_K,K)\,dt_1\cdots dt_K$$

$$= \sum_{K=0}^{\infty} P(K,T)\left(\frac{1}{T}\right)^K\iint\cdots\int i_e(t-t_K)\,i_e(t+\tau-t_j)\,dt_1\cdots dt_K \quad \text{since } P(t_1,\dots,t_K,K) = \left(\frac{1}{T}\right)^K P(K,T)$$

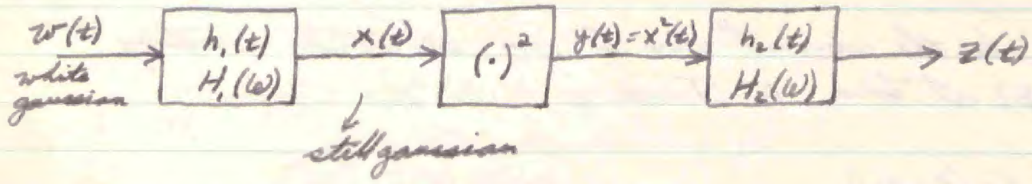
$$= \sum_{K=0}^{\infty} P(K,T)\begin{cases}\dfrac{e^2}{T^2}, & j\neq K\\[2mm] \dfrac{R(\tau)}{T}, & j=K\end{cases}$$

$$= \sum_{K=0}^{\infty} P(K,T)\underbrace{\left[K\frac{R(\tau)}{T} + (K^2-K)\frac{e^2}{T^2}\right]}$$

Now, using $K=\bar{n}T$, $\bar{n}^2 \gg \bar{n}$, we see ~~that for one electron~~,

$$R_i(\tau) = \bar{n}R(\tau) + (\bar{n}e)^2$$

Gaussian noise into a square-law detector:



$$x(t) = \int h_1(t-\tau)\, w(\tau)\, d\tau$$

$$y(t) = x^2(t) = \iint h_1(t-\tau) h_1(t-\mu)\, w(\tau)\, w(\mu)\, d\tau\, d\mu$$

$$z(\rho) = \iiint h_1(t-\tau) h_1(t-\mu)\, h_2(\rho - t)\, w(\tau)\, w(\mu)\, d\tau\, d\mu\, dt$$

Now change variables so $\xi = \rho - t$, $u = \rho - \mu$, $v = \rho - \tau$

Then

$$z(\rho) = \iiint h_1(v-\xi) h_1(u-\xi) h_2(\xi)\, w(\rho - v)\, w(\rho - u)\, du\, dv\, d\xi$$

Performing the ξ-integration gives as a definite function

$$\Lambda(u,v) = \int h_1(v-\xi)\, h_1(u-\xi)\, h_2(\xi)\, d\xi$$

$$= \sum_k \lambda_k\, \varphi_k(u)\, \varphi_k(v)$$

assuming we can so represent $\Lambda(u,v)$

Then

$$z(\rho) = \iint \sum_k \lambda_k\, \varphi_k(u)\, \varphi_k(v)\, w(\rho-u)\, w(\rho-v)\, du\, dv$$

$$= \sum_k \lambda_k \left[\int \varphi_k(v)\, w(\rho - v)\, dv \right]^2 = \sum_k \lambda_k\, W_k^2(\rho).$$

<u>Special case</u> :

$$z = \sum_{k=1}^{2N} x_k^2 \quad ; \quad E(x_k) = m \quad ; \quad E(x_k x_j) = \delta_{ij}\,(1+m^2)$$

The characteristic function for x_k^2 is

$$M_{x_k^2}(jv) = E\left(e^{jvx_k^2}\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{jvx_k^2}\, e^{-\frac{(x_k-m)^2}{2}}\, dx_k$$

$$M_{x_k^2}(jv) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(1-2jv)x_k^2 - 2mx_k + m^2}{2}}\, dx_k = \cancel{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x_k^2 - 2mx_k}{2\left[\frac{1}{1-2jv}\right]}}}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(1-2jv)}{2}\left[x_k^2 - 2\frac{m}{1-2jv}x_k + \frac{m^2}{(1-2jv)^2} - \frac{m^2}{(1-2jv)^2} + \frac{m^2}{1-2jv}\right]\right\} dx_k$$

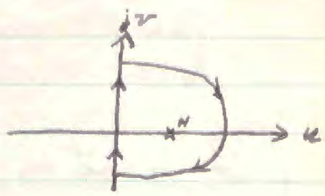
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\left(\frac{1-2jv}{2}\right)\left[\left(x_k - \frac{m}{1-2jv}\right)^2 - \frac{2jvm^2}{(1-2jv)^2}\right]}\, dx_k = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(1-2jv)\left(x_k - \frac{m}{1-2jv}\right)^2}{2}}\, e^{\frac{jvm^2}{1-2jv}}\, dx_k$$

$$= \frac{1}{\sqrt{1-2jv}}\, e^{\frac{jvm^2}{1-2jv}}$$

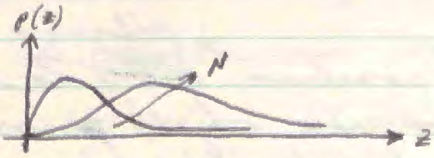
$$\& \quad M_z(jv) = \frac{1}{(1-2jv)^N}\, e^{\frac{jvNm^2}{1-2jv}}$$

For the special case $m=0$, $\quad M_z(jv) = \dfrac{1}{(1-2jv)^N}$

$$\& \quad P_0(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{(1-2jv)^N}\, e^{-jvz}\, dv = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \frac{1}{(1-2s)^N}\, e^{-sz}\, ds$$



$$P_0(z) = \frac{e^{-z/2}\, z^{N-1}}{2^N\,(N-1)!}$$



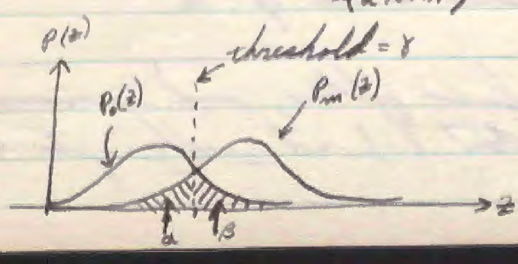
For large N, $P_0(z) \to$ gaussian $\left[\begin{smallmatrix}\text{var } 4N \\ \text{mean } 2N\end{smallmatrix}\right]$

For $m \neq 0$, we cannot use contour integration due to the essential singularity of the exponential term. But, it can be evaluated as

$$P_m(z) = \frac{z^{\frac{N-1}{2}}}{2(2Nm^2)^{N-1}}\, e^{-\frac{z+4N^2m^4}{2}}\, I_{N-1}\left(2Nm^2\sqrt{z}\right)$$

what γ is optimum?

what is "optimum"?



This is an example of a decision mechanism for deciding which of two processes ($m_x = m$ or $m_x = 0$) is generating the observed z. If $z > \gamma$, guess m; $z < \gamma$ guess 0.

$\alpha = P_r\{\text{missed detection}\}$; $\beta = P_r\{\text{false alarm}\}$

Decisions:

The types of decisions can be imagined along a spectrum; at one end we have the "discrete" decisions ~~and~~ (6.574) and at the other, the "smooth" decisions (6.571). In "discrete" decisions, the signal shape is well defined, perhaps known; we desire to detect the signal: is it present in the noise? ~~or which~~

In "smooth" decisions, the signal is not well-defined, we are interested in filtering & smoothing signal + noise to obtain a "best" estimate of the signal at some time past, present, or future.

We will start with the 6.571 problem of statistical linear filters & go to the 6.574 problems of discrete signals & "yes-no" decisions.

Linear filtering of signal plus noise:

Consider the following system: $x(t) = s(t) + n(t) \longrightarrow \boxed{\begin{matrix} h(t) \\ H(\omega) \end{matrix}} \longrightarrow s_o(t)$ (filter)

$$\left.\begin{aligned} s(t) &= \text{signal} \\ n(t) &= \text{random noise, zero mean} \end{aligned}\right\} \; s + n \text{ are uncorrelated. } \underline{R_{sn}(\tau) = 0}$$

$s_o(t)$ = actual output of filter

$s_d(t)$ = desired " " " $= f[s(t)]$

(<u>Comments on linearity, mean square error, + "goodness"</u>:)

We will constrain the filter ~~[illegible]~~ to be linear. We will also assume that we know only the correlation functions of the processes involved (i.e., first & second moments). If minimal mean-square error is to be our criterion for "goodness" of filtering, it is meaningless to ask if a non-linear filter would be better than our linear filter. This because only first & second moments of s & n are known & higher orders of filter response will enter into mean-square error only through higher order moments of s & n. We would have to know higher moments to tell if a non-linear filter would be "better".

Thus, correlation functions (second moments), linear filters, and minimum mean-square error all fit together as a consistent (if not really optimal) description of statistical filtering.

The Wiener-Hopf equation:

The work that follows merely indicates a trend, the purpose of which is to point up the important points, and is not to be considered a useful mathematical technique.

Suppose $s_d(t) = s(t+\alpha)$; i.e., we want to predict

$$\mathcal{E} = E\{[s_d(t) - s_o(t)]^2\} = E\{[\int h(u)[s(t-u) + n(t-u)]du - s(t+\alpha)]^2\}$$

$$= \iint h(u)h(v)R_s(u-v)\,du\,dv + \iint h(u)h(v)R_n(u-v)\,du\,dv + R_s(0) - 2\int h(u)R_s(u+\alpha)\,du$$

where $s(t)$ & $n(t)$ are uncorrelated: $R_{sn}(\tau) \equiv 0$

We now want to choose $h(t)$ to minimize $\mathcal{E}$. We will do so by using the calculus of variations:

Let

$$h(u) \to h(u) + \epsilon\,\eta(u) \text{ in the above expression for } \mathcal{E}.$$

Substituting, differentiating re: ϵ & setting $\epsilon = 0$ gives:

$$0 = \left.\frac{\partial \mathcal{E}}{\partial \epsilon}\right|_{\epsilon=0} = -2\int \eta(u)R_s(u+\alpha)\,du + 2\iint \eta(u)h(v)\left[R_s(u-v) + R_n(u-v)\right]du\,dv$$

This expression will be identically zero for all $\eta(\cdot)$ if

$$\boxed{R_s(u+\alpha) = \int h(v)\left[R_s(u-v) + R_n(u-v)\right]dv \quad \text{for all } u \geq 0}$$

This is the Wiener-Hopf equation.

The requirement of ~~rel~~ realizability on $h(\cdot)$ places the same requirement on $\eta(\cdot)$ & the final equation will not be valid for $u < 0$ where $\eta(u) \equiv 0$.

Minimum mean-square error by frequency domain solution:

We found the following expression for $\mathcal{E}$ on the last page

$$\mathcal{E} = \iint [R_s(t-\rho) + R_n(t-\rho)]\, h(t)\, h(\rho)\, dt\, d\rho - 2\int R_s(t+\alpha)\, h(t)\, dt + R_s(0)$$

Going to the transform domain, we get by transforming first on t, then ρ:

$$\iint [S_s(\omega) + S_n(\omega)] e^{-j\omega\rho} H(\omega)\, d\omega\, h(\rho)\, d\rho = \int [S_s(\omega) + S_n(\omega)] H(\omega) H^*(\omega)\, d\omega$$

so $$\mathcal{E} = \frac{1}{2\pi}\int [S_s(\omega) + S_n(\omega)] H^*(\omega) H(\omega)\, d\omega - 2\frac{1}{2\pi}\int S_s(\omega) H^*(\omega) e^{j\omega\alpha} d\omega + \frac{1}{2\pi}\int S_s(\omega)\, d\omega$$

$$= \frac{1}{2\pi}\int S_s(\omega)\left[1 - H^*(\omega) e^{j\omega\alpha} - H(\omega) e^{-j\omega\alpha} + |H(\omega)|^2\right] d\omega + \frac{1}{2\pi}\int S_n(\omega)\, |H(\omega)|^2\, d\omega$$

$$= \frac{1}{2\pi}\int S_s(\omega)\left[1 - 2G(\omega)\cos(\omega\alpha - \varphi(\omega)) + G^2(\omega)\right] d\omega + \frac{1}{2\pi}\int S_n(\omega)\, G^2(\omega)\, d\omega$$

where $H(\omega) = G(\omega) e^{j\varphi(\omega)}$; $G(\cdot)$ & $\varphi(\cdot)$ are real functions.

Now, we can minimize $\mathcal{E}$, ignoring realizability, by simply minimizing the sum of the two integrands ~~at~~ for each value of ω. This since both integrands are ≥ 0 for all ω.

In particular, we can minimize over $\varphi(\omega)$ readily as it occurs only in the cosine term. This is negative, so the integrand will be a min ~~|H|~~ over φ when $\cos[\omega\alpha - \varphi(\omega)] \equiv 1$ or $\boxed{\varphi(\omega) = \alpha\omega}$

Plugging this value of $\cos[\ ] = 1$ into the equation gives us

$$\gamma(G) = S_s(\omega)[1 - G(\omega)]^2 + S_n(\omega)\, G^2(\omega)$$

To minimize over G, we simply treat G as a number (ω is a parameter) and differentiate γ re: G:

$$0 = \frac{\partial\gamma}{\partial G} = -2S_s(\omega)[1 - G(\omega)] + 2S_n(\omega)\, G(\omega) = 0$$

or $$G(\omega) = \frac{S_s(\omega)}{S_s(\omega) + S_n(\omega)}$$

This should be the same sol we get in time domain if we disregard realizability there too.

& $$\boxed{H(\omega) = \frac{S_s(\omega)\, e^{j\omega\alpha}}{S_s(\omega) + S_n(\omega)} \quad \text{optimal disregarding realizability}}$$

Now, we plug this solution back into $\mathcal{E}$ & evaluate:

$$\mathcal{E}_{min} = \frac{1}{2\pi}\int\left\{S_s(\omega)\frac{S_n^2(\omega)}{[S_s(\omega)+S_n(\omega)]^2} + S_n(\omega)\frac{S_s^2(\omega)}{[S_s(\omega)+S_n(\omega)]^2}\right\}d\omega$$

$$\mathcal{E}_{min} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{S_s(\omega)S_n(\omega)}{S_s(\omega)+S_n(\omega)}\,d\omega$$ = minimum mean-square error for non-realizable linear filter.

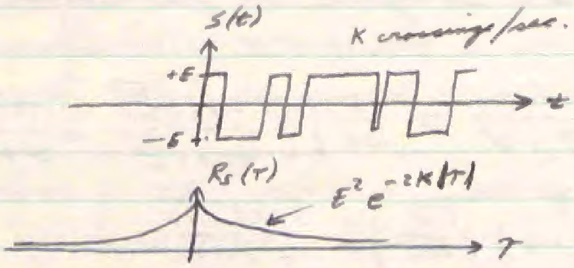
Example: Suppose $S_n(\omega) \equiv 0$. Then

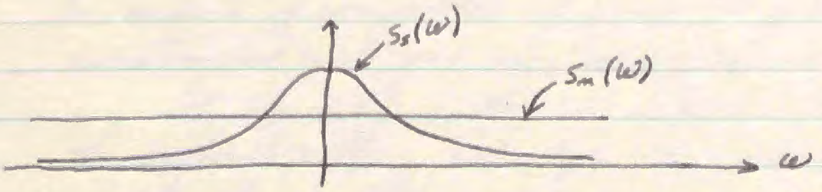
$$H(\omega) = \frac{S_s}{S_s+S_n}e^{j\omega\alpha} = e^{j\omega\alpha}$$

If $\alpha > 0$, this corresponds to a predictor (not realizable) } both have
$\alpha < 0$, " " " delay line (realizable) } zero m.s.e.

Example:

$S_n(\omega) = N$
$S_s(\omega) = \frac{1}{1+\omega^2}$



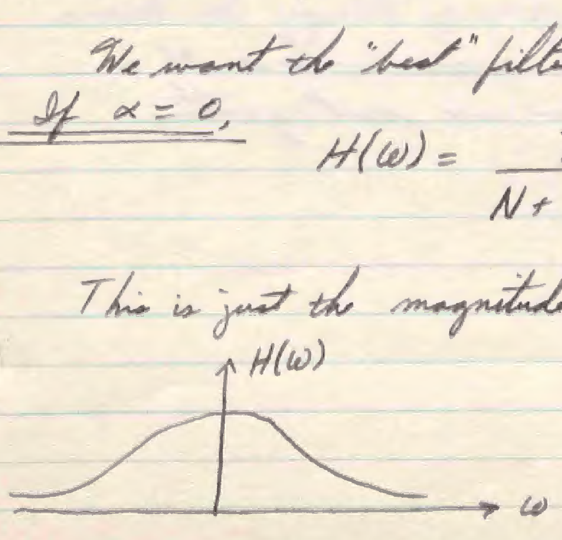


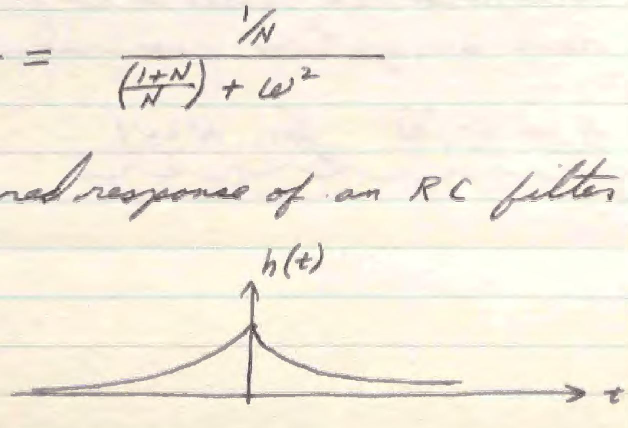
We want the "best" filter to separate the signal from the noise.

If $\alpha = 0$,

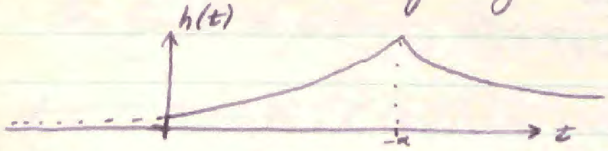
$$H(\omega) = \frac{\frac{1}{1+\omega^2}}{N+\frac{1}{1+\omega^2}} = \frac{1/N}{\left(\frac{1+N}{N}\right)+\omega^2}$$

This is just the magnitude squared response of an RC filter:





Now, note that the error is independent of the delay. If we make α be very negative, the filter response becomes

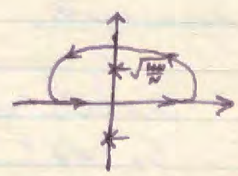


We can build a filter whose response for $t>0$ is $h(t)$ but which is realizable (response $=0$ for $t<0$). If α is sufficiently negative, the contribution of this dropped part of $h(t)$ will be very small.

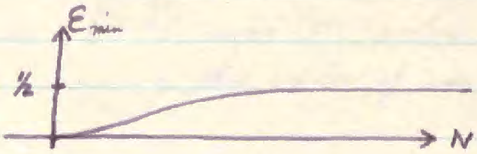
Thus, we can get very close to the minimum mean square error with a realizable filter if we are willing to wait a long time (this is usually milliseconds) for our answers. E_{min} is then the "irreducible error". We can approach this error but never go below it – with any filter so long as mean square error is our error criterion.

We will now compute the minimum m.s.e. for this problem:

$$E_{min} = \frac{1}{2\pi}\int \frac{d\omega}{\frac{1+N}{N}+\omega^2} = \frac{1}{2\pi}\int\left[\frac{-j\frac{1}{2}\sqrt{\frac{N}{1+N}}}{\omega-j\sqrt{\frac{1+N}{N}}} + \frac{j\frac{1}{2}\sqrt{\frac{N}{1+N}}}{\omega+j\sqrt{\frac{1+N}{N}}}\right]d\omega$$



$$= \frac{2\pi j}{2\pi}\left(-j\frac{1}{2}\sqrt{\frac{N}{1+N}}\right) = \frac{1}{2}\sqrt{\frac{N}{1+N}}$$

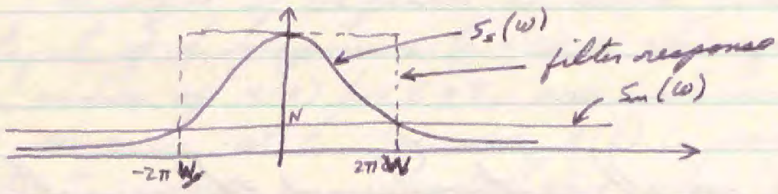


We are generally most interested in the cases where N is small; otherwise, the noise is so great we are not getting anything near acceptable communication:

$$E_{min} \to \frac{1}{2}\sqrt{N} \quad \text{for } N \ll 1$$

Naive solution:

Suppose we relied on intuition to build what seems to be a "good" filter: Let us build a simple realizable filter that passes all frequencies where $S_s(\omega) > S_n(\omega)$ and cuts off very rapidly beyond that point:



Assume N is so small that virtually all of the signal power is passed: Then

$$\mathcal{E} = \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} N\,d\omega = 2WN$$

At the point of crossing, $\left(\frac{1}{2\pi W}\right)^2 \approx N$ or $2WN = \frac{N}{\sqrt{N}\,\pi} = \underline{\underline{\frac{1}{\pi}\sqrt{N} \approx \mathcal{E}_{min}}}$

Obviously the constant multiplier is wrong due to our lousy approximations, but note that for small N as originally assumed, $\mathcal{E}_{min}$ goes as $\sqrt{N}$

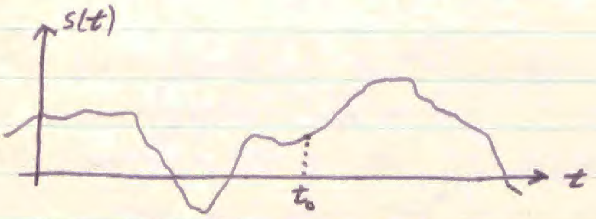
The point made here is that there is nothing of value to be gained by building the "optimal filter." We are interested only in orders of magnitude of error & just any old filter reasonably constructed will give very nearly the same error.

At the other extreme where N is large, it pays to build the optimal filter, but here the error is still so large as to render the system useless.

Pure prediction of a signal with a realizable filter:

No noise, $S_n(\omega) = 0$

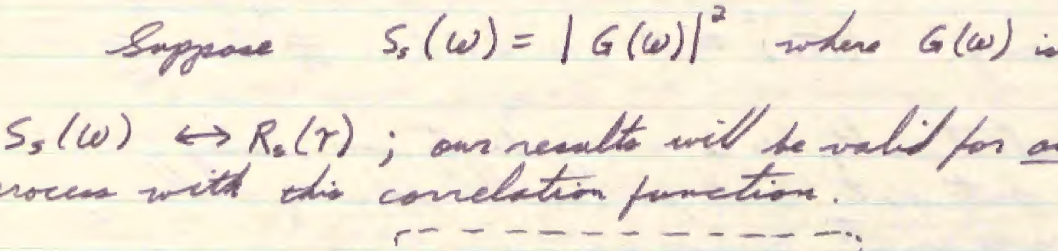
We want to know what the value of the signal will be at $t = t_0 + \alpha$, where we have observed $s(t)$ for $t < t_0$. $(\alpha > 0)$



The signal at t_0 depends on all past values of the signal. Thus, we have the infinite-dimensional equivalent of a simple two-dimensional linear approximation relating two random variables. (Or, if we have sampled past values, a multi-dimensional equivalent.)

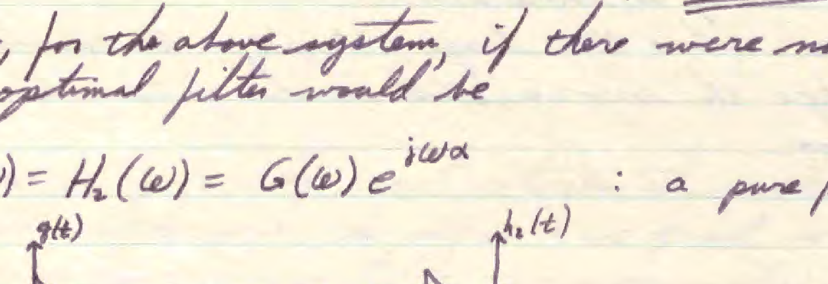
Suppose $\quad S_s(\omega) = |G(\omega)|^2$ where $G(\omega)$ is realizable ($\neq 0$ in RHP)

$S_s(\omega) \leftrightarrow R_s(\tau)$; our results will be valid for <u>any</u> random process with this correlation function.

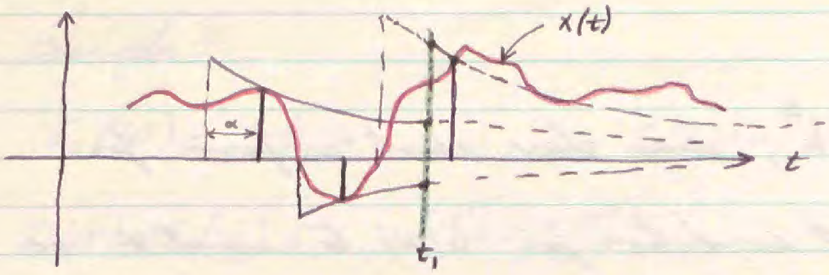


Now, for the above system, if there were no realizability constraints, the optimal filter would be

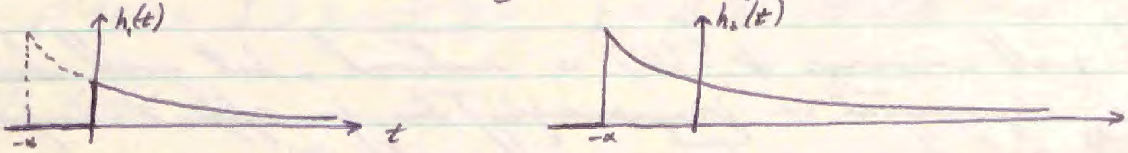
$H_1(\omega) = H_2(\omega) = G(\omega)\, e^{j\omega\alpha}$: a pure predictor



← e.g.



The output of the filter at some time t_1 is given by the integral of the responses to $x(t)$ in the region $-\infty < t < t_1 + \alpha$. But the values of $x(t)$ for $t > t_1$ are unknown and are independent of the value at t_1 because $x(t)$ is white gaussian. Thus, the best way to consider all these values is to ignore them; that is, make $H_1(\omega)$ be the realizable part of $H_2(\omega)$:



$H_1(\omega)$ is as good as $H_2(\omega)$ for considering the past values of $x(t)$. It merely ignores future values:

Thus the best realizable filter in the mean-square-error sense is

$$H(\omega) = \frac{1}{G(\omega)} \int_0^\infty \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} G(p)\, e^{jp\alpha} e^{jpt}\, dp \right] e^{-j\omega t}\, dt$$

Paley-Wiener problem:

Given a function $S(\omega) = |G(\omega)|^2$, does there exist a form for $G(\omega)$ which is realizable?

A necessary & sufficient condition for $G(\omega)$ to exist & be realizable is:

If $\int S(\omega)\,d\omega < \infty$, then there exists a $G(\omega)$ such that $|G(\omega)|^2 = S(\omega)$ & $g(t) \equiv 0$ for $t < 0$, if

$$\int \frac{|\ln S(\omega)|}{1+\omega^2}\,d\omega < \infty .$$

This integral will converge if $S(\omega)$ is rational, but will blow up if $S(\omega)$ is exponential in frequency. This in effect requires that $S(\omega)$ not decrease too rapidly with ω.

This is a very sharp (critical) requirement which is not very applicable to physical systems. It diverges for situations such as corners, functions identically zero over a finite range of frequency, & behavior at extremely large frequencies. These phenomena never occur or are of interest in practical situations.

Prediction with an unrealizable filter:

Recalling the result of page 52, if we do not require realizability, the optimal predictor in the presence of noise is

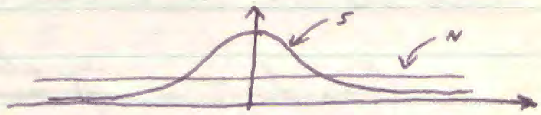
$$H(\omega) = \frac{S_s(\omega)}{S_s(\omega) + S_n(\omega)}\, e^{j\omega\alpha}$$

Note: If $S_s(\omega) = |G(\omega)|^2 = G(j\omega)G^*(j\omega) = G(\omega)G(-\omega)$

then the realizable optimum filter can be expressed:

$$H(\omega) = \frac{1}{G(\omega)} \int_0^{\infty} e^{-j\omega t} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{S_s(\omega')}{G(-\omega')} e^{j\omega'\alpha} e^{j\omega' t}\, d\omega'\, dt$$

<u>Example</u>:



$$S_s(\omega) = \frac{1}{1+\omega^2} \quad ; \quad S_N(\omega) = a^2$$

$$S_s(\omega) + S_N(\omega) = \frac{(a^2+1) + a^2\omega^2}{1+\omega^2}$$

$$|G(j\omega)|^2 = S_s(\omega) + S_N(\omega) = G(j\omega)\,G(-j\omega)$$

$$\to \boxed{\frac{1}{G(j\omega)}} \to \boxed{H_2'(\omega)} \to$$

Now $G(j\omega) = \mathcal{L}\mathcal{F}^{-1}\left[S_s(\omega) + S_N(\omega)\right] =$ LHP Part $\{S_s(\omega) + S_N(\omega)\} = \dfrac{\sqrt{a^2+1} + aj\omega}{1+j\omega}$

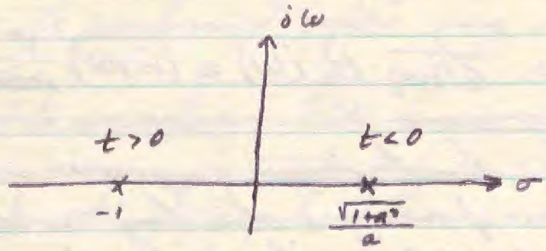
$$H_2'(\omega) = \mathcal{L}\mathcal{F}^{-1}\left\{\frac{S_s(\omega)}{G(-j\omega)} e^{j\omega\alpha}\right\} = \mathcal{L}\mathcal{F}^{-1}\{H_2(\omega)\}$$

The total transmission is

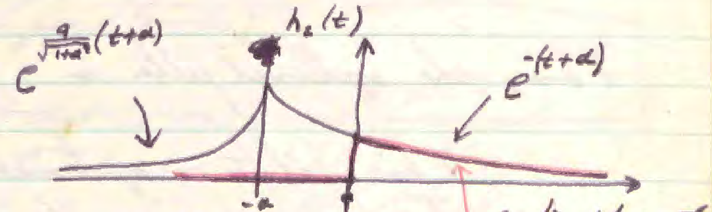
$$\mathcal{L}\mathcal{F}^{-1}\left\{\frac{S_s(\omega)}{|G(j\omega)|^2} e^{j\omega\alpha}\right\}$$

$$\mathcal{L}\mathcal{F}^{-1}\{S_s(\omega)\} = \frac{1}{1+j\omega}$$

$$H_2(\omega) = \frac{S_s(\omega)\,e^{j\omega\alpha}}{G(-j\omega)} = \frac{\frac{1}{(1+j\omega)(1-j\omega)}}{\frac{\sqrt{a^2+1}+aj\omega}{1-j\omega}} e^{j\omega\alpha} = \frac{e^{j\omega\alpha}}{(1+j\omega)(\sqrt{1+a^2} - j\omega a)}$$



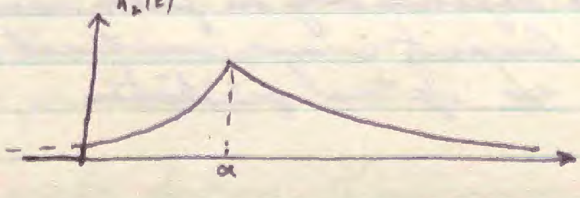
$\rightarrow$



Unrealizable filter $\boxed{h_2(t)}$

Realizable part $\boxed{h_2'(t)}$

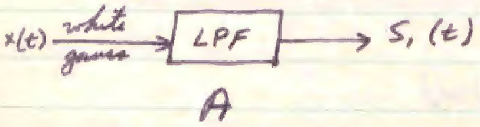
For past prediction, we can approach the "irreducible error" with a realizable filter:



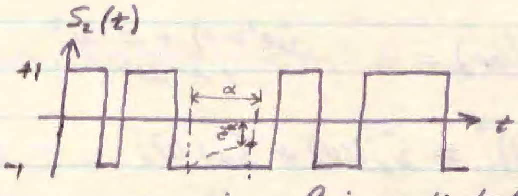
This filter is formally realizable, but would be difficult to build out of simple elements.

<u>Discrete vs. continuous systems & error criteria:</u>

Consider two ~~systems~~ signals, both having a power spectrum given by $S(\omega) = \frac{1}{1+\omega^2}$:

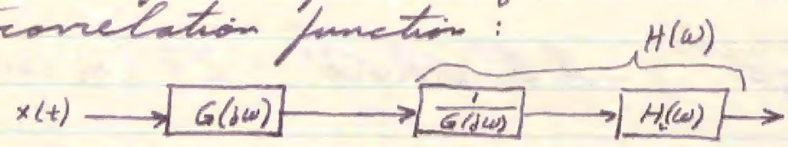


&

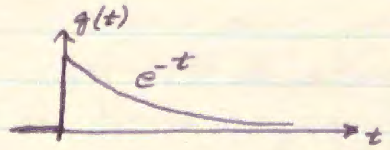


crossings Poisson distributed

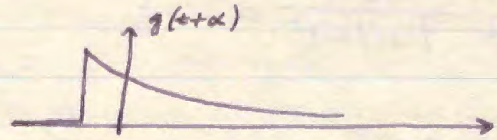
By the theory we have just been studying, our "optimal" filter is optimum for both processes as they have the same autocorrelation function:

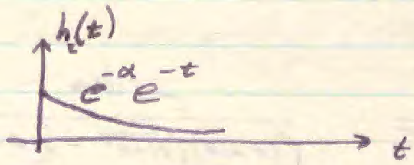


Here, $G(j\omega) = \frac{1}{1+j\omega}$, so $H_2(\omega) = \mathcal{L}\mathcal{F}^{-1}\{G(j\omega)e^{j\omega\alpha}\}$



$\longrightarrow$





$\longleftrightarrow$ $H_2(\omega) = \frac{e^{-\alpha}}{1+j\omega}$

$$\boxed{\text{Thus, the optimal filter is } H(\omega) = \frac{1}{G(j\omega)} H_2(\omega) = (1+j\omega)\frac{e^{-\alpha}}{1+j\omega} = e^{-\alpha}}$$

This filter is optimal in the least-mean-square-error sense. This is a reasonable criterion if we are dealing with a process such as A above. The L.M.S.E. criterion reduces gross errors but tolerates minor errors. It is good for smooth, symmetrical uni-modal curves, such as a gaussian.

For a true gaussian process, L.M.S.E. gives the best & only reasonable answer as it depends only on first & second moments anyway. Further, linearity is no constraint, so linear filtering for L.M.S.E. is <u>the</u> optimal prediction.

For processes such as process B above, LMSE is not a valid error criterion. It goes to pot for multi-modal or highly skewed distributions. For such cases, we must find some way to reduce (or weight) our solution space to allowed solutions.

For gaussian-like processes, we are dealing with filter theory; for discrete-like ~~~~ processes, we are dealing with decision theory.

Decision Theory:

In communications, coding is the dominant field of study; in radar, detection & decision dominate.

We will discuss decisions in the framework of radar-like systems, using applications of decision theory as a vehicle for study of the theory itself.

Assumptions:

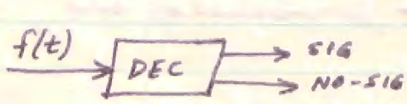
(0) We will consider only the output of the receiving antenna. We have no access to anything except the received waveform. There is no feedback from us to nature — this is strictly the "open-loop" situation of detection (to be performed before any eventual feedback).

(1) The waveform $f(t)$ received is all that is relevant. (no feedback)
(2) $f(t)$ is important only in the interval $0 < t < T$.
(3) In this interval, $f(t)$ is a sample function of one of two possible random processes: "signal-present" or "signal absent."
(4) The output of the decider is bi-valued: one output corresponds to "signal present", the other to "signal absent."
(5) The problem is repeated continually, each time with new data.
(6) Evaluation of our decision process is by average performance, error probabilities, etc.

→ The decision-maker destroys all information except that useful to my final purpose. This purpose must be stated before the decision-process is designed.

Detecting a determined signal in the presence of noise:

Suppose we have a situation in which the input is a given signal $f(t)$ ~~signal~~ $(0<t<T)$, plus noise or noise alone. We want to design a decision ~~rule~~ box which will give one of two outputs at time T: "signal" or "no-signal".



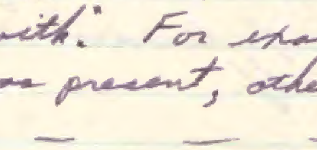
$$f(t) = \begin{Bmatrix} s(t)+n(t) \\ n(t) \end{Bmatrix}, \quad s(t) \text{ is known exactly for } 0 \le t \le T$$

Our decision rule will be of the form $F(f,s,T) \gtrless \gamma$, where γ is a threshold to be determined and the symbol $(\gtrless)$ means "compared with". For example, if $F(f,s,T) > \gamma$ we may decide a signal was present, otherwise, ~~we~~ it was not present.

One decision rule might be:

$$\int_0^T [f(t)-s(t)]^2 dt \gtrless \int_0^T f^2(t)\,dt \quad \rightleftarrows \quad \int_0^T f(t)\,s(t)\,dt \gtrless \tfrac{1}{2}\int_0^T s^2(t)\,dt$$

If $s(t)$ is



then this decision rule says $\int_0^T E f(t)\,dt \gtrless \frac{1}{2}E^2 T$

or $\frac{1}{T}\int_0^T f(t)\,dt \gtrless E/2$

i.e., compare the actual average value of the received signal with the average of the signal alone.

Is $\int_0^T s^2(t)[f(t)-s(t)]^2 dt \gtrless \int_0^T f^2(t)\,dt$ a better rule than the above? We cannot answer this question as we have not yet specified a criterion for "better".

Max $\frac{S}{N}$:
$$\int_{-\infty}^{t} x(\tau)\,h(t-\tau)\,d\tau \gtrless \gamma$$

~~where~~ $\cancel{\int h(\tau) R_n(t-\tau)\,d\tau = s(T-t)}$,

where $\int_{-\infty}^{\infty} R_n(t-\tau)\,h(\tau)\,d\tau = s(T-t), \quad t \geq 0$

Max likelihood :
$$\int_{0}^{\infty} x(\tau)\,h(\tau)\,d\tau \gtrless \gamma$$

w/ $\int_{0}^{\infty} R_n(t-\tau)\,h(\tau)\,d\tau = s(t)$

$\cancel{\int R_n(t-\tau)\,h_1(t-\tau)\,d\tau}$

or $\int_{-\infty}^{t} x(\tau)\,h_1(t-\tau)\,d\tau \gtrless \gamma$

$$\int_{-\infty}^{t} R_n(u-\tau)\,h_1(t-\tau)\,d\tau = s(u)$$

or if $t-\tau = \tau'$
$$\int_{0}^{\infty} h_1(\tau')\,R_n(u+\tau'-t)\,d\tau' = s(u)$$

$t-u=u'$

or $\int_{0}^{\infty} h_1(\tau')\,R_n(-u'+\tau') = s(t-u')$

Max $\frac{S}{N}\Big|$: at time t $\quad \int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau = \xi(t)$

$$x(t) \longrightarrow \boxed{h(\cdot)} \longrightarrow \xi(t)$$

where $\int_0^{\infty} R_n(\alpha-\tau)\, h(\tau)\, d\tau = s(t-\alpha)$, $\quad \alpha \geq 0$

<u>Max likelihood at time t:</u>

$$\int_{-\infty}^{t} x(\tau)\, h(t-\tau)\, d\tau = \xi(t)$$

where $\int_{-\infty}^{t} R_n(\alpha-t+\tau)\, h(t-\tau)\, d\tau = s(\alpha) \qquad \alpha \geq 0$

$$-\alpha+t \to \alpha$$

$$\int_{-\infty}^{t} R_n(\alpha-\tau)\, h(t-\tau)\, d\tau$$

Minimization of S/N with a linear filter (colored noise):

Our decision system is of the form

$$f(t) \rightarrow \underbrace{\boxed{h(t)}}_{\text{filter}} \xrightarrow{f_0(t)} \boxed{\text{SAMPLE}} \xrightarrow{f_0(T)} \boxed{\text{compare}} \rightarrow \text{decision}$$

Assume we know (1) $s(t)$ completely ; $s(t) = 0$ outside $0 \le t \le T$
(2) $R_n(\tau)$ + $S_n(\omega)$

Let $s_0(t)$ = output of filter when signal only is present
$n_0(t)$ = " " " " noise " " "

Now define the signal to noise ratio

$$\alpha \equiv \frac{s_0^2(T)}{E[n_0^2(T)]}$$

We will now select $h(t)$ to maximize α:

$$s_0(T) = \int_0^T s(t)\, h(T-t)\, dt$$

$$n_0(T) = \int_0^T n(t)\, h(T-t)\, dt$$

$$E\{n_0^2(T)\} = \iint_0^T E[n(t)n(\tau)\, h(T-t)\, h(T-\tau)\, dt\, d\tau = \iint_0^T R_n(t-\tau)\, h(T-t)\, h(T-\tau)\, dt\, d\tau.$$

$$\alpha = \frac{s_0^2(T)}{E\{n_0^2(T)\}} = \frac{\iint_0^T h(t)h(\tau)s(T-t)s(T-\tau)\,dt\,d\tau}{\iint_0^T R(t-\tau)\,h(t)\,h(\tau)\,dt\,d\tau.}$$

We now want to maximize α over $h(\cdot)$. We hold the denominator constant and maximize the numerator. This corresponds to a continuously varying scale factor on $h(\cdot)$ such that the denominator remains constant. Thus, $h(\cdot)$ will be determined only to within a scale factor, which ~~is unete~~ can always be tossed into the threshold anyway.

Now let $I = s_o^2(T) + \lambda E[n_o^2(T)]$

$$I + \epsilon I = \iint_0^T [h(t) + \epsilon \eta(t)][h(\tau) + \epsilon \eta(\tau)][s(T-t)\, s(T-\tau) + \lambda R_n(t-\tau)]\, dt\, d\tau$$

$\frac{d}{d\epsilon}(I + \epsilon I) = 0$ gives

$$0 = \iint_0^T \eta(t) h(\tau)[s(T-t)s(T-\tau) + \lambda R(t-\tau)]\, dt\, d\tau$$

Since this expression must hold for all $\eta(\cdot)$ in the region $0 \le \cdot \le T$, we have

$$0 = \int_0^T h(\tau)[s(T-\tau)s(T-t) + \lambda R_n(t-\tau)]\, d\tau \qquad 0 \le t \le T$$

$$= s(T-t) \underbrace{\int_0^T h(\tau) s(T-\tau)\, d\tau}_{\text{independent of } t} + \lambda \int_0^T h(\tau) R_n(t-\tau)\, d\tau \qquad 0 \le t \le T$$

or, equivalently,

$$0 = s(T-t) + \lambda' \int_0^T h(\tau) R_n(t-\tau)\, d\tau \qquad 0 \le t \le T$$

Recall however, that we can determine $h(\cdot)$ only to within a scale factor. Thus we can ignore the value of λ' and write

$$\boxed{\int_0^T h(\tau) R_n(t-\tau)\, d\tau = s(T-t) \qquad 0 \le t \le T}$$

This equation defines $h(\cdot)$ as a linear realizable filter which will maximize the signal to noise ratio before we make our threshold comparison.

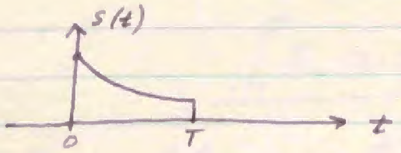
Solution to S/N maximizing filter for special signals:

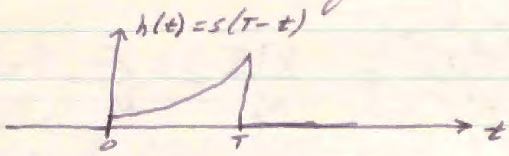
(1) If the noise $n(t)$ is white, $R_n(\tau) = N_0\,\delta(\tau)$. The equation defining $h(\cdot)$ on page 64 then becomes

$$N_0 \int_0^T h(\tau)\,\delta(t-\tau)\,d\tau = s(T-t)$$

or $h(t) = \frac{1}{N_0} s(T-t) \rightarrow s(T-t)$ within scale factor.

This is called the North filter or matched filter.





Our decision rule here is

$$f_0(T) = \int_0^T f(t)\,h(T-t)\,dt \gtrless \gamma$$

or $\int_0^T f(t)\,s(t)\,dt \gtrless \gamma$ since $s(t) = h(T-t)$

This is the same rule as we got by considering mean square errors.

• Evaluation of best S/N ratio:

$$\alpha = \frac{\left[\int_0^T s(T-t)\,h(t)\,dt\right]^2}{N_0 \int_0^T h^2(t)\,dt}$$

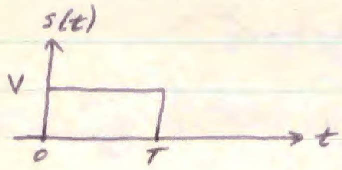
$$\frac{N_0}{E}\alpha = \frac{N_0\,\alpha}{\int_0^T s^2(t)\,dt} = \frac{\left[\int_0^T s(T-t)\,h(t)\,dt\right]^2}{\left[\int_0^T h^2(t)\,dt\right]\left[\int_0^T s^2(T-t)\,dt\right]} \le 1$$ by Schwartz inequality.

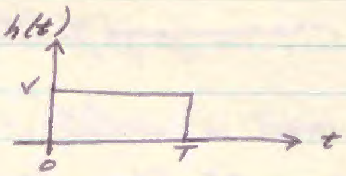
The equality holds when $h(t) = s(T-t)$: another way for determining the optimal form for $h(\cdot)$.

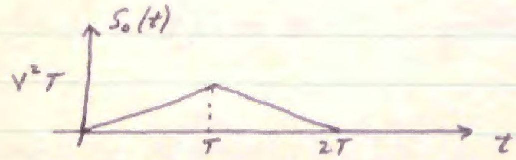
Thus:

$$\boxed{\alpha_{max} = \frac{E}{N_0}}$$

(see over next page)

If 

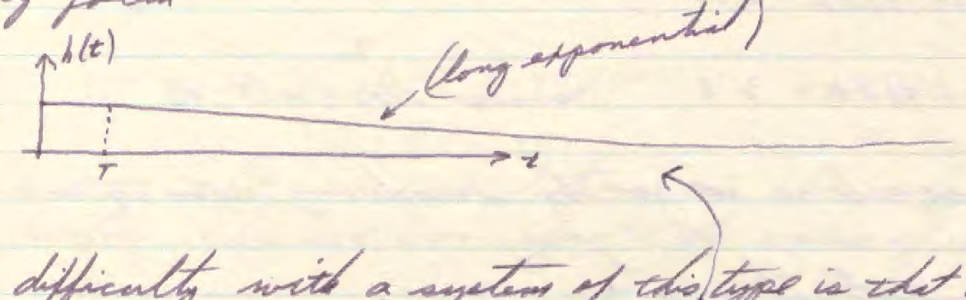
then 



The noise is $N_0 V^2 T$

so $\alpha = \frac{(V^2T)^2}{N_0 V^2 T} = \frac{V^2 T}{N_0} = \frac{E}{N_0}$

To physically construct a filter such as this, we could turn the system "on" at $t=0$ and let $h(t)$ take on the following form:



The difficulty with a system of this type is that it must be started from neutral for each time interval.

The difference between matched filter detection (discussed here) and correlation type detection is that the matched filter is used only once while the other can be used repeatedly.

Maximum S/N in infinite interval for detection:

If we consider the preceeding problem, but let the time interval be $-T \le t \le T$ and sample at $t=0$, our system can be redrawn as:

$$f(t)=s(t)+n(t) \longrightarrow \boxed{h(t)} \xrightarrow{f_0(t)} \boxed{\begin{matrix}\text{SAMPLE}\\ t=0\end{matrix}} \xrightarrow{f_0(0)} \boxed{\text{COMPARE}} \longrightarrow$$

If we now rework the problem, letting $T \to \infty$, we get

$$\int_{-\infty}^{\infty} R_n(t-\tau)\,h(\tau)\,d\tau = s(-t)$$

{ add condition $t \geq 0$ for realizable filter

Now, let

$$R_n(\tau) \longleftrightarrow \Phi_n(\omega)$$
$$h(\tau) \longleftrightarrow H(\omega)$$
$$s(t) \longleftrightarrow S(\omega)$$
$$s(-t) \longleftrightarrow S^*(\omega)$$

The above equation in the transform domain becomes

$$\Phi_n(\omega) \cdot H(\omega) = S^*(\omega)$$

or $$H(\omega) = \frac{S^*(\omega)}{\Phi_n(\omega)}$$

Now, if the noise is similar to the signal (i.e., sample functions look like the signal in some sense),

$$\Phi_n(\omega) = S^*(\omega)\,S(\omega)$$

& then $$H(\omega) = \frac{1}{S(\omega)}$$

This system gives a spike output whenever the signal $s(t)$ occurs.

Derivation of decision rule from sampling input signal:

Suppose the input to our system is $x(t)$ and we sample $x(t)$ at times $t_1, \ldots, t_K$:

$$\bar{x} = [x(t_1), x(t_2), \ldots, x(t_K)]$$

Similarly, let $\bar{s} = [s(t_1), s(t_2), \ldots, s(t_K)]$

$$\bar{n} = [n(t_1), n(t_2), \ldots, n(t_K)]$$

Assume that $P(\bar{n})$ is known.
" " P_s is the a priori probability that $s(t)$ is present.

We now concentrate on finding the a posteriori probability ~~of~~ that $\bar{s}$ is present given that $\bar{x}$ has occured, $P(\bar{s}|\bar{x})$. We want to boil all our information down to this one number. This is an optimal job of data reduction in that it retains all useful information + rejects all irrelevant information about whether $\bar{s}$ was present or not.

It carries no information about other aspects of the situation. Thus, we must decide what we want before we design our system.

Our decision rule here is: compare ~~Px~~ $P(\bar{s}|\bar{x})$ with a threshold somewhere on the real line $(0,1)$. This partitions the line segment into two regions, "signal present" and "signal not present".

Let $P(\bar{x}|\bar{n})$ be the probability of $\bar{x}$ given noise only.

$$P(\bar{s}|\bar{x}) = \frac{P(\bar{x}|\bar{s})\,P(\bar{s})}{P(\bar{x})} = \frac{P(\bar{x}|\bar{s})\,P_s}{P(\bar{x}|\bar{s})P_s + P(\bar{x}|\bar{n})\,[1-P_s]}$$

$$= \frac{\Lambda \frac{P_s}{1-P_s}}{1+\Lambda\frac{P_s}{1-P_s}}$$

where $\Lambda \equiv \dfrac{P(\bar{x}|\bar{s})}{P(\bar{x}|\bar{n})} \equiv$ <u>liklihood ratio</u>

$$\Lambda = \frac{\rho_n(\bar{x}-\bar{s})}{\rho_n(\bar{x})}$$

Note that setting a threshold for $P(\bar{s}|\bar{x})$ is equivalent to setting a different threshold for Λ since Λ is a monotonic function of $P(\bar{s}|\bar{x})$:

$$\underline{\Lambda = \frac{\rho_n(\bar{x}-\bar{s})}{\rho_n(\bar{x})} \gtrless \gamma}$$

If we now assume $R_n(\tau)$ is gaussian and we know $R_n(\tau)$, let

$$R \equiv \begin{bmatrix} R(0) & R(t_1-t_2) & \cdots & R(t_1-t_K) \\ R(t_1-t_2) & R(0) & \cdots & R(t_2-t_K) \\ \vdots & \vdots & & \vdots \\ R(t_1-t_K) & R(t_2-t_K) & \cdots & R(0) \end{bmatrix}$$

Then

$$\rho_n(\bar{x}) = \frac{1}{(2\pi)^{K/2}|R|^{1/2}}\, e^{-\frac{\bar{x}R^{-1}\bar{x}_t}{2}}$$

so

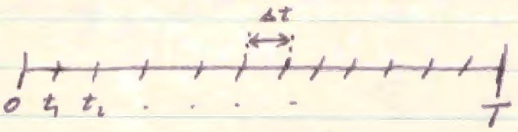
$$\Lambda = e^{+\frac{\bar{x}R^{-1}\bar{x}_t}{2} - \frac{(\bar{x}-\bar{s})R^{-1}(\bar{x}-\bar{s})_t}{2}}$$

$$= e^{\bar{x}R^{-1}\bar{s}_t}\, \underbrace{e^{-\frac{\bar{s}R^{-1}\bar{s}_t}{2}}}_{\text{independent of }\bar{x}} \gtrless \gamma$$

The exponential function is monotonic, so we can use the equivalent decision rule

$$\bar{x} R^{-1} \bar{s}_t \gtrless \gamma$$

We will now pass to an infinite number of samples as a limit & try to find the corresponding decision rule:



K intervals

$$t_i = i\Delta t$$

$$\Delta t = T/K$$

Define a column vector $\bar{h}_t$ by $\bar{h}_t \Delta t = R^{-1} \bar{s}_t$

so $\bar{x} R^{-1} \bar{s}_t = \bar{x} \bar{h}_t \Delta t \gtrless \gamma$

If we now pass to the limit $K \to \infty$, the matrix multiplication goes over to an integral:

$$\int_0^T x(t)\, h(t)\, dt \gtrless \gamma$$

We can now find the corresponding equation defining $h(t)$ by:

$$\bar{h}_t \Delta t = R^{-1} \bar{s}_t$$

$$R \bar{h}_t \Delta t = \bar{s}_t$$

or, in the limit,

$$\int_0^T R_n(t-\tau)\, h(t)\, dt = s(\tau)$$

Decisions in phase space:

In the last section, we found $\bar{x}$, a collection of K samples from $x(t)$ in a given range. If we now look at our decision making problem in K-space, the space of all possible $\bar{x}$, we find that a decision rule consists of labeling each possible point in K-space with a decision.

In our simple case of detection, this consists of partioning the space into two regions corresponding to "signal" and "no-signal".

Note that we have said nothing about the method used to find a decision rule, threshold, optimal filters, etc. Yet we still have something to talk about: How do we divide the K-space into two regions V_s and V_N corresponding to "signal" and "no signal" respectively.

Neyman-Pearson treatment:

This is the oldest classical treatment of decisions in this way. ~~We assign to each point~~ Each $\bar{x}$ corresponds to a point in K-space. We assign to each point the probabilities $P(\bar{x}|s)$ and $P(\bar{x}|n)$. ~~We arbitrarily ch~~

There are two possible errors in this simple system: (1) failure to detect a signal that is present, and (2) to decide that a signal is present when none actually is present (false alarm).

In the Neyman-Pearson method, we arbitrarily decide on some value for $P_F = \Pr\{\text{false alarm}\}$. We now try to find a partition rule (decision rule) which will maximize $P_D = \Pr\{\text{detection}\} = 1 - P_M$ where $P_M = \Pr\{\text{failure to detect}\}$.

Let $$P_F = \int_{V_s} P(\bar{x}|\bar{n})\, d\bar{x} = \alpha.$$

$$P_D = \int_{V_s} P(\bar{x}|s)\, d\bar{x}$$

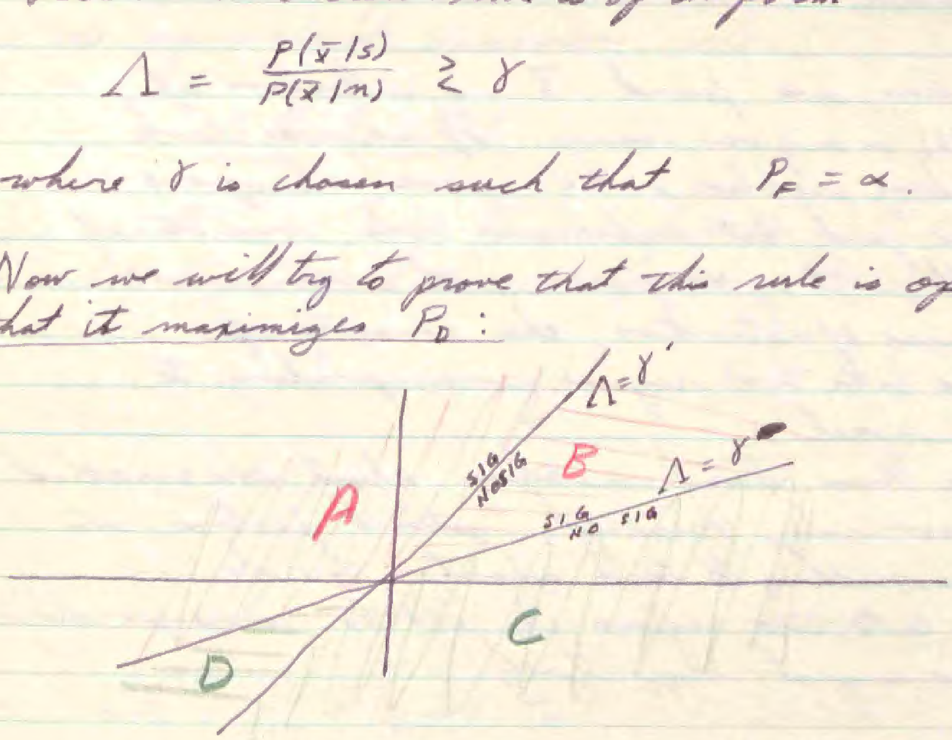
$$P_M = \int_{V_N} P(\bar{x}|s)\, d\bar{x}$$

Assume the decision rule is of the form

$$\Lambda = \frac{P(\bar{x}|s)}{P(\bar{x}|n)} \gtrless \gamma$$

where γ is chosen such that $P_F = \alpha$.

Now we will try to prove that this rule is optimal in the sense that it maximizes P_D:



For our rule $\Lambda \gtrless \gamma$,

$$P_F = \int_{A \cup B} P(\bar{x}|n)\, d\bar{x} \qquad \& \qquad P_D = \int_{A \cup B} P(\bar{x}|s)\, d\bar{x}$$

For the rule $\Lambda = \gamma'$,

$$P_F' = \int_{A \cup D} P(\bar{x}|n)\, d\bar{x} \le \alpha, \quad \text{and} \quad P_D' = \int_{A \cup D} P(\bar{x}|s)\, d\bar{x}$$

Now we want to show that $P_D' \le P_D$

First, $P_F' \le P_F \Rightarrow \int_{A \cup D} P(\bar{x}|n)\, d\bar{x} \le \int_{A \cup B} P(\bar{x}|n)\, d\bar{x}$

or $\int_D P(\bar{x}|n)\, d\bar{x} \le \int_B P(\bar{x}|n)\, d\bar{x}$

Now, in B, $\frac{P(\bar{x}|s)}{P(\bar{x}|n)} > \gamma$ or $P(\bar{x}|n) < \frac{P(\bar{x}|s)}{\gamma}$

and in D, $\frac{P(\bar{x}|s)}{P(\bar{x}|n)} < \gamma$ or $P(\bar{x}|n) > \frac{P(\bar{x}|s)}{\gamma}$

Combining these last three relations, we see that

$$\int_B \frac{P(\bar{x}|s)}{\gamma} d\bar{x} \geq \int_D \frac{P(\bar{x}|s)}{\gamma} d\bar{x}$$

or $$\int_{A \cup B} P(\bar{x}|s) \, d\bar{x} \geq \int_{A \cup D} P(\bar{x}|s) \, d\bar{x}$$

so $P_D' \leq P_D$ which was to be shown, so that for any given P_F, ~~any value of γ other than~~ a rule of this form is optimal.

Bayes decision rule (costs):

Assume that we know ~~a priori~~ the a priori probability, P_s, ~~of~~ that a signal is present. Let

C_F = cost of false alarm
C_M = cost of missing a target

$$E(C) = (1-P_s) P_F C_F + P_s P_M C_M \quad , \quad P_M = 1 - P_D$$

$$= C_F(1-P_s) \int_{V_s} P(\bar{x}|n)\,d\bar{x} + P_s C_M \int_{V_N} P(\bar{x}|s)\,d\bar{x}$$

$$\overset{?}{=} C_F(1-P_s) \int_{V_s+V_N} P(\bar{x}|n) \left\langle \begin{array}{c} \frac{C_M P_s}{C_F(1-P_s)} \frac{P(\bar{x}|s)}{P(\bar{x}|n)} \\ \\ 1 \end{array} \right\rangle d\bar{x} \quad > 0$$

We want to minimize this $E(C)$ average cost for each $\bar{x}$: This will be done if we partition κ-space such that

$$V_n = \left\{ \bar{x} \;\middle|\; \frac{C_M P_s}{C_F(1-P_s)} \frac{P(\bar{x}|s)}{P(\bar{x}|n)} < 1 \right\}$$

Thus, our decision rule is

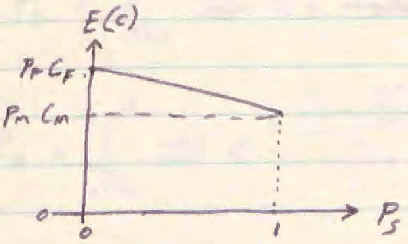
$$\frac{C_M P_s}{C_F(1-P_s)} \frac{P(\bar{x}|s)}{P(\bar{x}|n)} \underset{\text{no sig}}{\overset{\text{sig}}{\gtrless}} 1$$

The difficulties encountered in actually trying to use this model is the inability to pin numbers on P_s, C_F, and C_M.

Minimax decisions:

We can characterize each decision rule such as that found on the last page by finding the maximum value of $E(c)$ over all values of P_S. That rule can then be characterized by that value of P_S. We assume that for any decision rule, the worst will prevail (a somewhat unjustified pessimistic viewpoint) and try to bound our loss: We try to find that rule which gives the least upper bound to our loss

$$\text{minimax loss} = \min_{Z} \max_{P_S} E\{C(P_S)\}$$



The fundamental theorem of the minimax approach is:

《 That rule which achieves the minimax solution is the same for maximin $E(c)$ and minimax $E(c)$, and the resulting bound is the same for both cases.

This implies that this bound is independent of who moves first, me or nature. My least upper bound is natures highest lower bound. 》

We have previously found the minimum of $E(c)$ over all possible decision rules on the last page. We now want to find the maximum of this min $E(c)$ over P_S:

$$\max_{P_S} \left[P_F(P_S)\, C_F (1-P_S) + P_M(P_S)\, C_m\, P_S \right] = \, ?$$

$P_F(P_S)$ and $P_M(P_S)$ both increase with P_S from zero to one as P_S varies from zero to one. Thus both terms peak somewhere and the sum has a peak. Thus there exists a P_S which maximizes the min $E(C)$. If we work the math out, we find that when $P_F C_F = P_S C_M$, the maximin $E(c)$ is the same for all values of P_S.

Review of detection rules:

White noise: $\int_0^T x(t)s(t)\,dt \underset{NS}{\overset{S}{\gtrless}} \gamma$

Colored noise: $\int_0^T x(t)\,h(t)\,dt \underset{NS}{\overset{S}{\gtrless}} \gamma$

where $h(\tau)$ satisfies $\int_0^T R_n(t,\tau)\,h(\tau)\,d\tau = s(\tau)$, $0 < t < T$

Some applications of the hair:

Assume white gaussian noise with zero mean:

Let $L \equiv \int_0^T x(t)\,s(t)\,dt$

$P(L|n) \equiv P_r\{L \,|\, \text{only noise present}\}$

$E(L|n) = 0$, $E(L^2|n) = \iint_{0\,0}^{T\,T} E[x(t)\,x(\tau)]\,s(t)\,s(\tau)\,dt\,d\tau$

$E(L^2|n) = \iint_0^T N_0\,\delta(t-\tau)\,s(t)\,s(\tau)\,dt\,d\tau = N_0 \int_0^T s^2(t)\,dt \; \underline{= N_0 E}$

Thus, since $P(L|n)$ must be gaussian (we have nothing but gaussian!):

$$\boxed{P(L|n) = \frac{1}{\sqrt{2\pi}\sqrt{N_0 E}}\, e^{-\frac{L^2}{2N_0 E}}}$$

$E(L|s) = \int_0^T E[s(t)+n(t)]\,s(t)\,dt = \int_0^T s^2(t)\,dt = E$

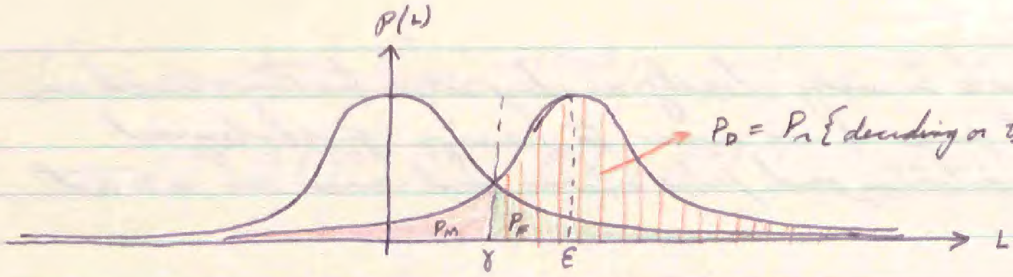
$E(L^2|s) = \iint_0^T E\{[n(t)+s(t)][n(\tau)+s(\tau)]\}\,s(t)\,s(\tau)\,dt\,d\tau = N_0 E + E^2$

or $\text{var}(L|s) = N_0 E$

so

$$\boxed{P(L|s) = \frac{1}{\sqrt{2\pi}\sqrt{N_0 E}}\, e^{-\frac{(L-E)^2}{2N_0 E}}}$$

Same as $P(L|n)$ except that the mean is non-zero.



$P_D = P_r\{$deciding or *thinking* we saw something$\}$

$$P_D = \int_\gamma^\infty p(L|s)\,dL \quad , \quad P_F = \int_\gamma^\infty p(L|n)\,dL$$

Let $\frac{L}{\sqrt{N_0 E}} = z$, then:

$$P_F = \int_{\gamma/\sqrt{N_0 E}}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}\,dz = \operatorname{erf}\left[\frac{\gamma}{\sqrt{N_0 E}}\right]$$

$$P_D = \int_{\frac{\gamma}{\sqrt{N_0 E}}}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{\left(z-\sqrt{\frac{E}{N_0}}\right)^2}{2}}\,dz = \operatorname{erf}\left[\frac{\gamma - E}{\sqrt{N_0 E}}\right]$$

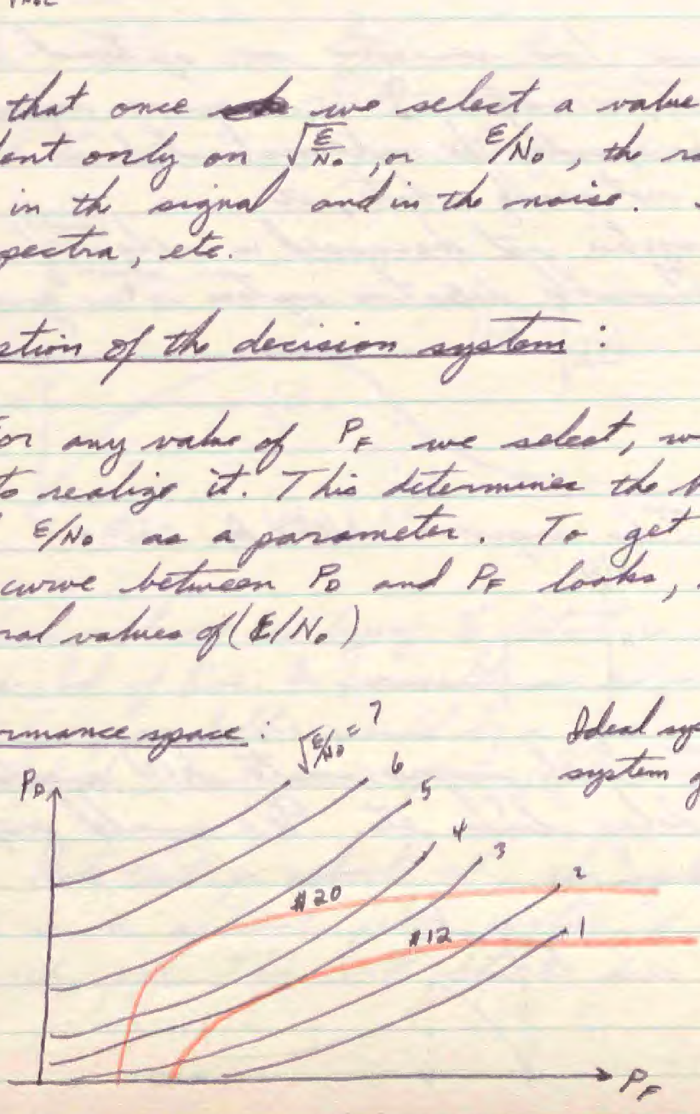
<u>tabulated functions</u>

Notice that once ~~we~~ we select a value for P_F, P_D is dependent only on $\sqrt{\frac{E}{N_0}}$, or E/N_0, the ratio of energy or power in the signal and in the noise. It is independent of the spectra, etc.

<u>Description of the decision system</u>:

For any value of P_F we select, we can find the ~~x~~ value of γ needed to realize it. This determines the best we can do for P_D, ~~for~~ with E/N_0 as a parameter. To get a picture of how the trading curve between P_D and P_F looks, we can plot P_D vs P_F for several values of (E/N_0)

<u>Performance space</u>:



Ideal system for any E/N_0; any other system gives lesser P_D for given P_F

typical cost curve: cost = const.

(over)

We might be able to very closely realize a situation such as depicted on the last page if we were to assign psychological costs & rewards to detection and false alarm and miss in a test of hearing, etc.

For colored noise, the curves remain the same, but the S/N parameter is changed to Q defined as

$$Q = \begin{cases} \sqrt{E/N_0} & \text{for white noise} \\ \int_0^T h(t)s(t)\,dt & \text{for colored noise} \end{cases}$$

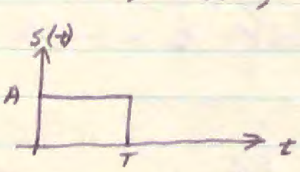
<u>Sensitivity of rules to perturbations from "optimal"</u>:

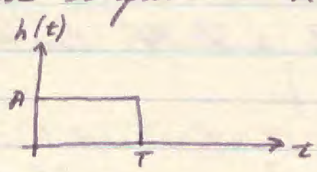
We would like to know how sensitive our optimal solution is to small (or not so small) variations in various aspects of the system: $s(t)$ will not be exactly the mathematical form we use to represent it, $n(t)$ is not really white, etc.

The answer is that the system just discussed is relatively insensitive. Just about any plausible detection system is almost as good as our optimal system.

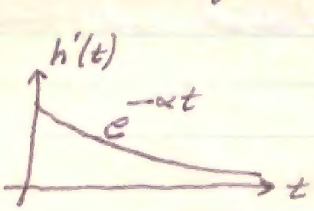
<u>Example (matched filter)</u>:

Suppose $s(t) = A$, $0 \le t \le T$. This requires $h(t) = A$, $0 \le t \le T$

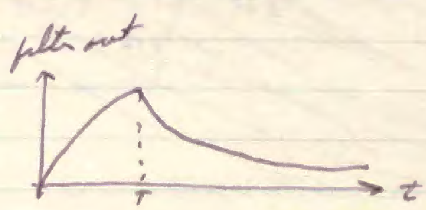




To actually build a filter to closely approximate $h(t)$ would be quite expensive, involving delay lines, etc. We would prefer to use a simple RC filter:



→



$$\int_0^T h'(T-t)s(t)\,dt = \int_0^T s(t)e^{-\alpha(T-t)}\,dt = \int_0^T Ae^{-\alpha(T-t)}\,dt$$

$$= e^{-\alpha T}\frac{A}{\alpha}\left(e^{\alpha T}-1\right) = \frac{A}{\alpha}\left(1-e^{-\alpha T}\right)$$ for signal

$$\int |H'(\omega)|^2 N(\omega)\,d\omega = \int |H'(\omega)|^2 N_0\,\delta(\omega)\,d\omega = N_0\int |H'(\omega)|^2\,d\omega$$

$$= \int |h'(t)|^2\,dt = N_0\int_0^\infty e^{-2\alpha t}\,dt = N_0/2\alpha$$ = mean square noise

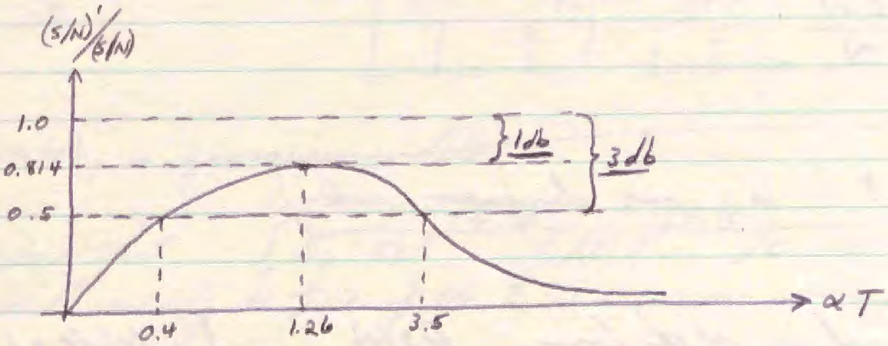
So, for our approximate RC-filter,

$$\boxed{(S/N)' = \frac{2A^2(1-e^{-\alpha T})^2}{\alpha N_0}}$$

For the "optimal" matched filter,

$$\boxed{(S/N) = \frac{A^2 T}{N_0}} \quad \text{so} \quad \frac{(S/N)'}{(S/N)} = \frac{2}{\alpha T}\left(1-e^{-\alpha T}\right)^2$$

We want to choose α for the maximum value of this ratio:



→ The signal we are trying to detect is a DC level existing for a period of time. As a result, we should expect that some sort of low pass filter would be a good one. The above curve shows that the exact form of the filter isn't too important.

Note that for optimal αT, this system will give the same results as an "optimal" filter system if we can only raise our S/N ratio by 1 db. This will often be less expensive than building a fancy filter.

Signals with variable parameters:

Let a be a parameter associated with the input signal we are trying to detect. We can then write the signal as $s(t,a)$ with $P(a)$ known apriori. We want a binary output, "sig" or "no sig", but we do not have a binary input.

$$P(a) = P_r\{ s(t,a) \mid \text{some signal is present}\}$$

$$P(s) = P_r\{\text{any signal being present}\}$$

$$P(s|x) = \frac{P(s)\int P(x|s,a)P(a)\,da}{P(s)\int P(x|s,a)P(a)\,da + P(x|n)P(n)}$$

$$= \frac{\frac{P(s)}{P(n)P(x|n)}\int P(x|s,a)P(a)\,da}{1 + \frac{P(s)}{P(n)P(x|n)}\int P(x|s,a)P(a)\,da} = P(s_1 \text{ or } s_2 \text{ or} \cdots)$$

s_i mutually exclusive.

Let $\Lambda(a) = \dfrac{P(x|s,a)}{P(x|n)}$

$\boxed{\int \Lambda(a)P(a)\,da \gtrless \gamma \quad \text{is our decision rule}}$

Example:

Suppose the signal is of the form $As(t)$, $\int_0^T s^2(t)\,dt = 1$

$\Lambda(A) = \dfrac{P(x|A,s)}{P(x|n)}$ & the rule is $\int \Lambda(A)P(A)\,dA \gtrless \gamma$.

$$\Lambda(A) = e^{-\frac{1}{2N_0}\int_0^T [x(t)-As(t)]^2 dt + \frac{1}{2N_0}\int_0^T x^2(t)\,dt}$$

for white gaussian noise

$$\rightarrow L' = \int e^{\frac{A}{N_0}\int_0^T x(t)s(t)\,dt - \frac{A^2}{2N_0}} P(A)\,dA \gtrless \gamma$$

Note that the received signal $x(t)$ enters into the decision only through the form $\int_0^T x(t)s(t)dt$. If we can show that this exponential is monotonic non-decreasing with L, the decision is equivalent to comparing L to a threshold.

$\frac{dL'}{dL}$ will be greater or less than zero depending on $P(A)$.

Thus, we see that there will be some specific situations where the decision rule here will be the same as for the single signal case.

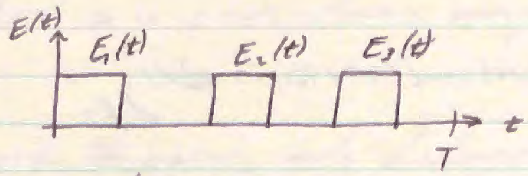
In general, if we have several signals,

$$\sum \Lambda_i P(s_i) = \sum \frac{P(x|s_i)}{P(x|n)} P(s_i) \gtrless \gamma$$

Rotating target model for radar:

Suppose $s(t) = \sum_{i=1}^{m} E_i(t) \cos[\omega t + \varphi_i(t) + \theta_i]$

$P(\theta_i) = \frac{1}{2\pi}$



Noise, $n(t)$, is gaussian white.

We require that $\int_0^T E_k(t) e^{j\varphi_k(t)} E_\ell(t) e^{j\varphi_\ell(t)} dt = 0$.

This will surely be true if the $E_i(t)$ are non-overlapping as shown above.

Decision rule:

$$\frac{1}{(2\pi)^m} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{P(x|s,\theta)}{P(x|n)} d\bar{\theta} \gtrless \gamma$$

$$= \frac{1}{(2\pi)^m} \int_0^{2\pi} \cdots \int_0^{2\pi} \frac{e^{-\frac{1}{2N_0}\int_0^T [x(t) - s(t,\bar{\theta})]^2 dt}}{e^{-\frac{1}{2N_0}\int_0^T x^2(t) dt}} d\bar{\theta}$$

$$\frac{1}{(2\pi)^n}\int_0^{2\pi}\cdots\int_0^{2\pi} e^{\frac{1}{N_0}\int_0^T x(t)s(t,\bar\theta)dt}\, e^{-\frac{1}{2N_0}\int_0^T s^2(t,\bar\theta)dt}\, d\bar\theta$$

When we perform $\int_0^T s^2(t,\bar\theta)dt$, the cross terms of $s^2(t,\bar\theta)$ go out due to our assumed orthogonality. We assume the diagonal terms are independent of θ_i because the $\cos(\cdot)$ is a narrow band function:

$$\int_0^T E^2\cos^2(\omega_0 t+\varphi_i(t)+\theta_i)\,dt = \int_0^T \frac{E^2}{2}dt = \frac{E^2 T}{2} \quad \text{if } \underline{T \gg \frac{1}{\omega_0}}$$

Thus our decision rule is reduced to

$$\frac{1}{(2\pi)^m}\int_0^{2\pi}\cdots\int_0^{2\pi} e^{\frac{1}{N_0}\int_0^T x(t)\sum E_i(t)\cos(\omega_0 t+\varphi_i(t)+\theta_i)dt}\, d\bar\theta \ \geq \gamma$$

Bringing the summation outside the integral on t gives

$$\frac{1}{(2\pi)^n}\int\cdots\int e^{\frac{1}{N_0}\sum\int x(t)E_i(t)\cos(\omega_0 t+\varphi_i(t)+\theta_i)dt}\, d\bar\theta \ \geq \gamma$$

$$\prod_i \frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{N_0}\int_0^T x(t)E_i(t)\cos(\omega_0 t+\varphi_i(t)+\theta_i)dt}\, d\theta_i \quad \geq \gamma$$

$$\prod_i \frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{N_0}\int_0^T [x(t)E_i\cancel{(t)}\cos[\omega_0 t+\varphi_i(t)]\cos\theta_i - x(t)E_i(t)\sin[\omega_0 t+\varphi_i(t)]\sin\theta_i]dt}\, d\theta_i$$

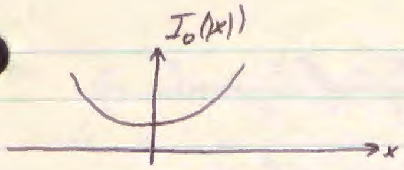
Now we know $\frac{1}{2\pi}\int_0^{2\pi} e^{a\cos\theta - b\sin\theta}\,d\theta = I_0\left(\sqrt{a^2+b^2}\right)$

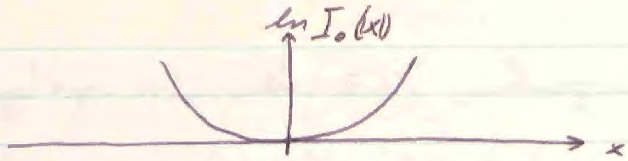
so

$$\prod_i I_0\left\{\frac{1}{N_0}\left|\int_0^T x(t)E_i(t)e^{j[\omega_0 t+\varphi_i(t)]}dt\right|\right\} \geq \gamma$$

or, to get rid of the product,

$$\boxed{\sum_i \ln I_0\left\{\frac{1}{N_0}\left|\int_0^T x(t)E_i(t)e^{j[\omega_0 t+\varphi_i(t)]}dt\right|\right\} \geq \gamma}$$

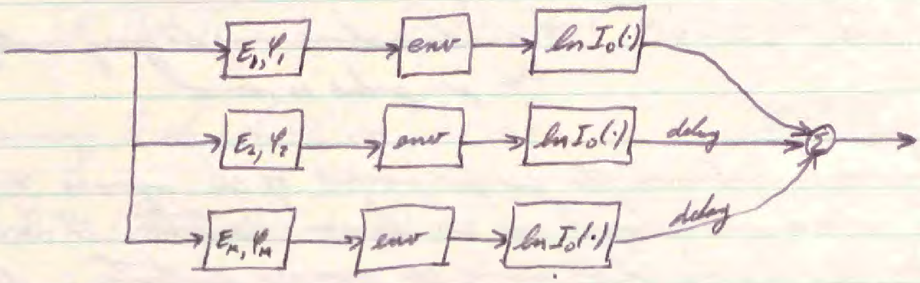




If we have a filter whose impulse response is $E_i(t)\cos[\omega_0 t + \varphi_i(t)]$ which is narrow band, the argument

$$\int_0^T x(t)\, E_i(t)\cos[\omega_0 t + \varphi_i(t)]\,dt$$

is the envelope of the output. This suggests how we might construct a detection device as outlined above:



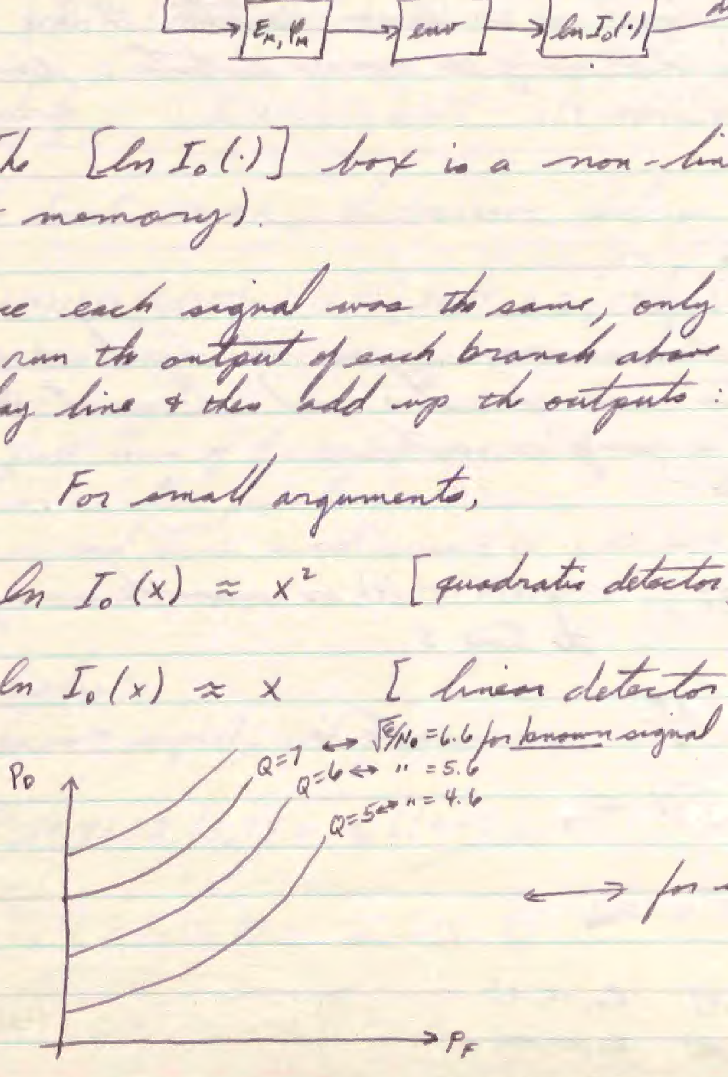
The $[\ln I_0(\cdot)]$ box is a non-linear amplitude transformation (no memory).

Since each signal was the same, only delayed ~~to~~ in time, we can run the output of each branch above into an appropriate delay line & then add up the outputs:

For small arguments,

$\ln I_0(x) \approx x^2$ [quadratic detector] ~~of~~ if (S/N) is small per pulse

$\ln I_0(x) \approx x$ [linear detector] if (S/N) is large per pulse.



$\longleftrightarrow$ for single signal

For multiple signals, $Q = E/N_0$ for each signal, M orthogonal signals

Q_e = the equivalent E/N_0 for the known signal case that gives the same curve P_D vs P_F

$$\sqrt{Q_e} = \frac{\left(1+\frac{\sqrt{M}}{2}\right)Q}{\sqrt{1+Q}}$$

If Q is large, $\sqrt{Q_e} \approx \sqrt{QM}$

If Q is small, $\sqrt{Q_e} \approx \frac{1}{2} Q\sqrt{M}$ $\Rightarrow$ random phase may be very costly if (E/N) per pulse is small.

There is little difference between the performance of the square & linear detectors in the region where we should use the linear detector. There is, however, a big difference if we should be using the square detector. Hence, we use the square detector all the time.

Detection of noise in noise:

As a general case of our past discussions, suppose the signal we are looking for is noise. We have two noise generators and we want to know if one particular generator is going or not.

Our decision rule will be:

$$\frac{P\{x \mid R_1(t,T)\ ;\ m_1(t)\}}{P\{x \mid R_0(t,T)\ ;\ m_0(t)\}} \gtrless \gamma\ ;$$ $m_0(t)$ + $m_1(t)$ are ensemble averages at the time t.

If we sample at times $t_n = n\frac{\Delta t}{T}$, our rule is

$$\frac{\frac{1}{(2\pi)^{\frac{T}{\Delta t}}|R_1|^{1/2}}\, e^{-(\bar{x}-\bar{m}_1)[R_1]^{-1}(\bar{x}-\bar{m}_1)_t}}{\frac{1}{(2\pi)^{T/\Delta t}|R_0|^{1/2}}\, e^{-(\bar{x}-m_0)[R_0]^{-1}(\bar{x}-\bar{m}_0)_t}}$$

$$R_0 = \text{matrix}\begin{bmatrix} R_0(0,0) & R_0(\Delta t,0) & \cdots \\ R_0(0,\Delta t) & R_0(\Delta t,\Delta t) & \cdots \\ - & - & - \quad - \end{bmatrix}$$

By ~~so~~ virtue of the monotonicity of the exponential, an equivalent rule is:

$$-(\bar{x}-m_1)[R_1]^{-1}(\bar{x}-m_1)_t + (\bar{x}-\bar{m}_0)[R_0]^{-1}(\bar{x}-\bar{m}_0)_t \gtrless \gamma$$

or $$\underline{\bar{x}\left([R_0]^{-1}-[R_1]^{-1}\right)\bar{x}_t + 2\bar{x}\left[[R_1]^{-1}\bar{m}_{1_t} - [R_0]^{-1}\bar{m}_{0_t}\right] \gtrless \gamma}$$

Define $[r_1]$ as a square matrix such that $[r_1](\Delta t)^2 = [R_1]^{-1}$

and $[r_0] \equiv \frac{1}{(\Delta t)^2}[R_0]^{-1}$

Now $[R_1][R_1]^{-1} = I = [R_1][r_1](\Delta t)^2$ or $[R_1][r_1]\Delta t = \frac{1}{\Delta t} I$

Now, passing to the limit as $\Delta t \to 0$, $\underline{\sum R_{ij}\, r_{jk}\Delta t = \left(\tfrac{1}{\Delta t}\right)\delta_{ik}}$

$$\boxed{\begin{aligned} &\int_0^T R_1(t,\tau)\, r_1(\tau,u)\, d\tau = \delta(t-u) \\ &\qquad\qquad\qquad\qquad \boxed{0<t,u<T} \\ &\text{\& similarly,} \quad \int_0^T R_0(t,\tau)\, r_0(\tau,u)\, d\tau = \delta(t-u) \end{aligned}}$$

These equations define $r_1(\tau,u)$ and $r_2(\tau,u)$.

In the limit $\Delta t \to 0$, the decision rule is

$$\boxed{\iint_0^T x(t)x(\tau)\left[r_0(t,\tau) - r_1(t,\tau)\right]dt\, d\tau + 2\int_0^T x(t)\left\{\int_0^T\left[r_1(t,\tau)m_1(\tau) - r_0(t,\tau)m_0(\tau)\right]d\tau\right\}dt \gtrless \gamma}$$

The first term of this expression will vanish if both noises have the same correlation functions.
The second term will vanish if both have zero mean.

Examples:

(1) known signal, white noise

$$R_0(t,\tau) = R_1(t,\tau) = N_0\,\delta(t-\tau)$$

$$r_0(t,\tau) = r_1(t,\tau) = \frac{1}{N_0}\,\delta(t-\tau)$$

$$m_1(t) = s(t), \quad m_0(t) = 0.$$

Our rule is then: $2\int_0^T x(t)\left\{\int_0^T r_1(t,\tau)m_1(\tau)d\tau\right\}dt \gtrless \gamma$

or $2\int_0^T x(t)\left\{\int_0^T \frac{1}{N_0}\delta(t-\tau)s(\tau)d\tau\right\}dt = \frac{2}{N_0}\int_0^T x(t)s(t)dt \gtrless \gamma$

or $\int_0^T x(t)s(t) \gtrless \gamma$

(2) White noise, ~~too with~~ white noise with mean:

Same as above (1), except $m_0(t) \neq 0$
Rule becomes

$$\int_0^T x(t)\left[s(t) - m_0(t)\right]dt \gtrless \gamma$$

(3) Known signal, colored noise:

$R_0(t,\tau) = R_1(t,\tau) = R(t,\tau)$
$r_1(t,\tau) = r_0(t,\tau) = r(t,\tau)$
$m_0(t) = 0, \quad m_1(t) = s(t)$

Rule is $2\int_0^T x(t)\left\{\int_0^T r(t,\tau)s(\tau)d\tau\right\}dt \gtrless \gamma$

or $\int_0^T x(t)h(t)dt \gtrless \gamma$ where $h(t) \equiv \int_0^T r(t,\tau)s(\tau)d\tau$

(4) White noise, signal known except for amplitude A, $p(A) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{A^2}{2\sigma^2}}$:
$m_1(t) = m_2(t) = 0$

$R_0(t,\tau) = N_0\delta(t-\tau) \Rightarrow r_0(t,\tau) = \frac{1}{N_0}\delta(t-\tau)$

$R_1(t,\tau) = E\left\{\left[n(t)+As(t)\right]\left[n(\tau)+As(\tau)\right]\right\} = N_0\delta(t-\tau) + \sigma^2 s(t)s(\tau)$

Guess that $r_1(t,\tau) = \frac{1}{N_0}\delta(t-\tau) + K s(t)s(\tau)$

Now convolve $R_1 * r_1$:

$$\delta(t-u) = \int R_1(t,\tau)\, r_1(\tau,u)\, d\tau$$

$$= \delta(t-u) + \frac{\sigma^2}{N_0} s(t)s(u) + K N_0\, s(t)s(u) + \sigma^2 K s(t) s(u)$$

Thus $1 = 1$, $\quad 0 = \frac{\sigma^2}{N_0} + K N_0 + \sigma^2 K \quad \Rightarrow \quad K = -\frac{\sigma^2/N_0}{\sigma^2 + N_0}$

Then $\iint x(t)x(\tau)\left[\cancel{\tfrac{1}{N_0}\delta(t-\tau)} + K\, s(t)s(\tau)\right] dt\, d\tau \gtrless \gamma$

or $\iint x(t)x(\tau)s(t)s(\tau)\, dt\, d\tau \gtrless \gamma$

$$\left[\int_0^T x(t)s(t)\,dt\right]^2 \gtrless \gamma$$

<u>A paradox of perfect decisions on noise-in-noise</u>:

Suppose $R(t,\tau) = R_1(t,\tau) = K R_0(t,\tau)$, $\quad K \neq 1$; $\quad m_0(t) = m_1(t) = 0$.

$r_1(t,\tau) = \frac{1}{K} r_0(t,\tau) = r(t,\tau)$

The decision rule is then $\quad L = \iint x(t)x(\tau)\left[\frac{1}{K} - 1\right] r(t,\tau)\, dt\, d\tau \gtrless \gamma$

$$E(L|0) = \iint R_0(t,\tau)\left[\frac{1-K}{K}\right]\frac{1}{K} r_0(t,\tau)\, dt\, d\tau = \frac{1-K}{K^2}\,\delta(0) = \infty$$

$$E(L|1) = \iint R_1(t,\tau)\, \frac{1-K}{K}\, r_1(t,\tau)\, dt\, d\tau = \frac{1-K}{K}\,\delta(0) = \infty$$

Similarly, $\quad E(L^2|1,0) \to \infty \quad$ & $\quad \boxed{\frac{E[L|0]}{\sqrt{\mathrm{var}[L|0]}} \to \infty}$

This implies that we can make <u>perfect decisions</u> (no error) since the ratio of mean to variance goes to infinity.

Recall that we can write $\quad R(t,\tau) = \sum_0^\infty \lambda_n \varphi_n(t)\varphi_n(\tau)$

where $\sum_{n=0}^\infty \varphi_n(t)\varphi_n(\tau) = \delta(t-\tau)$ for the orthonormal functions $\varphi_n(\cdot)$

Thus $\quad r(t,\tau) = \sum \frac{1}{\lambda_n} \varphi_n(t)\varphi_n(\tau)$

We require $\sum_n \lambda_n < B$ for R to be well behaved.

But this implies $\sum_n \frac{1}{\lambda_n} > A$, so r is not well behaved.

Now, our rule is

$$\iint x(t)x(\tau) \sum \frac{1}{\lambda_n} \varphi_n(t)\varphi_n(\tau)\,dt\,d\tau \gtrless \gamma$$

or $$\sum \frac{\iint x(t)x(t)\varphi_n(t)\varphi_n(\tau)\,dt\,d\tau}{\lambda_n} = \sum \frac{x_n^2}{\lambda_n} \gtrless \gamma$$

where $x(t) = \sum_{n=0}^{\infty} x_n \varphi_n(t)$

and the x_n are statistically independent random variables

Thus we have an infinite number of statistically independent samples & can make a perfect decision.

This is done as follows: Put $x(t)$ through a filter to make it white. This gives an infinite number of samples to make our decision on.

That is, if $\underset{\text{white}}{w(t)} \longrightarrow \boxed{H(\omega)} \longrightarrow$ $S(\omega) \leftrightarrow x(t)$ $\boxed{S(\omega) = 1 \cdot |H(\omega)|^2}$

then $x(t) \longrightarrow \boxed{\frac{1}{H(\omega)}} \longrightarrow w(t)$: white $S_w(\omega) = S(\omega)\left|\frac{1}{H(\omega)}\right|^2 = |H(\omega)|^2\left|\frac{1}{H(\omega)}\right|^2 = 1.$

In reality, we of course don't know $S(\omega)$ for truly infinite frequencies. To treat this problem, we would have to include in our model the fact that $S(\omega)$ is not really known for $\omega \to \infty$, or that we don't really know $f(t)$ to 67 decimal places for all values of time.

In general, problems are not so sensitive to the assumption that we know everything exactly as we put into the model. This is a singular situation.

The model is least sensitive when trying to separate two very unlike signals (e.g., known & random).

Detection of and differentiation between two possible signals in noise:

Here we have three possibilities: noise and (1) no signal, (2) $s_1(t)$, or (3) $s_2(t)$. We want to detect which signal is present if either. We would like to find the a posteriori probabilities

$P(s_1|x)$, $P(s_2|x)$, and $P(n|x)$,

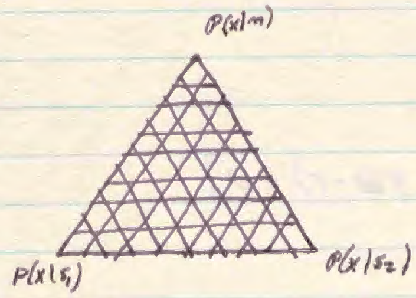
as these three numbers summarize all we can know about this situation. Using Bayes' rule,

$$P(s_1|x) = \frac{P(x|s_1)P(s_1)}{P(x|s_1)P(s_1) + P(x|s_2)P(s_2) + P(x|n)P(n)}$$

Similarly, we can find $P(s_2|x)$ and $P(n|x)$ from the a priori probabilities $P(x|s_1)$, $P(x|s_2)$, and $P(x|n)$. But the sum of these three numbers is unity, so we can use the likelihood ratios

$$\left.\begin{aligned} &\Lambda_1 = \frac{P(x|s_1)}{P(x|n)} \\ \text{and} \quad &\Lambda_2 = \frac{P(x|s_2)}{P(x|n)} \end{aligned}\right\}$$ These two numbers also fully summarize all the information useful to this detection system.

Here errors are not completely specified. If we guess s_1 and that is wrong, is s_2 or n correct? Are all possible errors equally costly.



linear scales

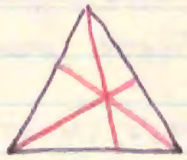
Any point on this triangle defines three numbers whose sum is unity.

Let $\Lambda_3 = \frac{\Lambda_1}{\Lambda_2}$

Now use three rules

$$\Lambda_1 \gtrless \gamma_1 \quad , \quad \Lambda_2 \gtrless \gamma_2 \quad , \quad \Lambda_3 \gtrless \gamma_3 \quad \text{where } \gamma_1 + \gamma_2 + \gamma_3 = 1$$

This gives the partitions



It can be shown that (1) a unique decision results
(2) this rule is appropriate for linear costs.

But - there is no simple way of saying what is the best rule. A common system is to assume all errors equally ~~likely~~ costly and all signals equally likely. We could also assume ~~--~~ all signals are present whose γ is exceeded.

<u>Example</u>:

This model is applicable to a radar situation. We have a continuum of possible signals in the time & frequency domains. This is useful as a model for range & doppler shift in radar.

$$\Lambda = \frac{P[x(t) | s(t, \tau, \omega)]}{P[x(t) | n]}$$

$$\sim e^{\text{envelope} \int_0^T x(t) s(t, \tau, \omega) dt}$$

$$s(t, \tau, \omega) = E[at - \tau] \cos [\omega_0 t + \omega t + \varphi(t-\tau) + \theta]$$

For a narrow band signal, $a \approx 1$.

$$\Lambda \sim e^{\frac{1}{N_0} \left| \int_0^T x(t) E(t-\tau) e^{i\omega_0 t} e^{i\omega t} e^{i\varphi(t-\tau)} dt \right|}$$

Assume $\quad x(t) = n(t) + \mathrm{Re}\{u(t-\tau')e^{j\omega' t}e^{j\omega_0 t}\}$

where $\quad u(t) = E(t)\, e^{j\varphi(t)}$

Now we neglect the noise:

$$P(\tau, \omega) \sim e^{\frac{1}{N_0}\left|\int_0^T \mathrm{Re}\{u(t-\tau')e^{j\omega' t}e^{j\omega_0 t}\} u(t-\tau)e^{j\omega_0 t}e^{j\omega t}dt\right|}$$

$$\sim e^{\frac{1}{2N_0}\left|\int_0^T u^*(t-\tau')e^{j\omega' t}u(t-\tau)e^{j\omega t}dt\right|}$$

since $\omega \ll \omega_0$.

$$P(\tau, \omega) \sim e^{\frac{1}{2N_0}\left|\int_0^T u^*(t)u(t-\tau)e^{j\omega t}dt\right|}$$

Where τ & ω are now difference from where the target really is.

or $\underline{P(\tau, \omega) \sim e^{\frac{1}{2N_0}|\theta(\tau, \omega)|}}$

where $\theta(\tau, \omega) = \int_0^T u^*(t)u(t-\tau)e^{j\omega t}dt = \begin{cases}\text{radar uncertainty function} \\ \text{or Woodward's " "}\end{cases}$