THE VECTOR SPACE OF RANDOM VARIABLES.

Now, the set of real random variables defined on a probability space with the field of real numbers constitute a vector space with $X + Y$ the vector addition of two random variables X, Y and aX the multiplication a vector (r.v X) by a scalar a (easy to check). However, the vector spaces of r.vs are somewhat more complicated then R^n or C^n. Recall that R^n and C^n are all finite dimensional spaces so that it is possible to specify every vector as a linear combination of a number of basis vectors. In the case, of r.v.s, unless the sample space is finite the vector space becomes infinite dimensional.

INNER PRODUCT

An additional property that a vector space over a scalar field may have is the existense of an inner product. This is a function mapping vector pairs to the scalar fields.
An inner product which is denoted as $< \cdot, \cdot >$ must satisfy the following properties:

1. Existence: $< X, Y >$ is defined for all vectors X,Y
2. Linearity: $< X, \alpha Y + \beta X >= \alpha < X, Y > + \beta < X, Z >$
3. Hermitian symmetry: $< X, Y >=< Y, X >^H$
4. Positive difiniteness: $< X, X >> 0 \; if \; X \neq 0$

* A vector space with an inner product is called inner product space

* R^n with the well known dot product is an inner product space

* For the vector space of random variables a valid inner product is

$$\boxed{< X, Y > \triangleq E\{X^* Y\}} = Cov(X, Y) \text{ (for zero mean r.vs)}$$

as it can be easily verified (assuming of course that $E\{|X|^2\} < \infty$ for all r.vs X in the space).

NORM or DISTANCE

Recall that the notion of length or distance or norm of a vector is define as

$$||X|| =< X, X >^{1/2}$$

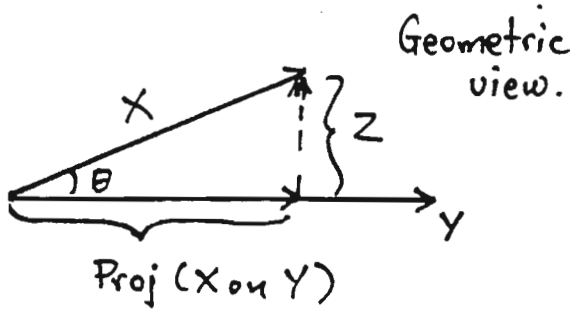
and satisfies the following properties:

1. $||aX|| = |a|\ ||X||$
2. $||X|| > 0$ unless $X = 0$ where $||0|| = 0$
3. $| < X, Y > | \leq ||X||\ ||Y||$(Cauchy-Schwarz inequality)
4. $||X + Y|| \leq ||X|| + ||Y||$ (Triangle inequality)

PROJECTION

Given two vectors X and Y, the projection of X on Y is given by:

$$Proj(X\ on\ Y) =< X, \frac{Y}{||Y||} > \frac{Y}{||Y|}$$



For vector spaces over R only we can write (by inspection from the geometric view)

$$||X||cos\theta =< X, \frac{Y}{||Y||} >=< X, Y > \frac{1}{||Y||} \Rightarrow < X, Y >= ||X||\ ||Y||cos\theta$$

(Note that proj(X,Y) is nothing else but the vector in the form of Y which is closest to X in the norm sense)

More general in the vector space of random variables (assume zero mean r.v for the moment)

$$< X, Y > = E\{X^* Y\} = cov(X, Y)$$
$$||X||^2 = E\{X^* X\} = var(X)$$

And from the familiar relation $\rho = \frac{cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

$$\Rightarrow \boxed{< X, Y >= \rho \cdot ||X||\, ||Y||} \Rightarrow < X, \frac{Y}{||Y||} >= \rho \cdot ||X||$$

(i.e., $cos\theta$ is the correlation coefficient in vector spaces over R)

ORTHOGONALITY

Two vectors X and Y are called orthogonal iff $< XY >= 0$. In vector spaces of random variables this is equivalent to $\rho = 0$. Looking back into the $proj(X\ on\ Y)$ we see that (zero mean r.v.s)

$$\begin{aligned} E\{ZY^*\} &= E\{(X - proj(X\ on\ Y))Y^*\} = E[(X - < X, \frac{Y}{||Y||} > \frac{Y}{||Y||})Y^*] \\ &= E\{(X - \rho\frac{||X||}{||Y||}Y)Y^*\} = E\{XY^*\} - \rho\frac{||X||}{||Y||}E\{YY^*\} \\ &= \rho\sigma_X\sigma_Y - \rho\frac{\sigma_X}{\sigma_Y}\sigma_Y^2 = 0 \Rightarrow \\ E\{ZY^*\} &= 0 \Rightarrow Z, Y\ orthogonal \end{aligned}$$

MEAN SQUARE ESTIMATION

Let us condider the problem where we are given a random variable Y, and we are asked to find the best approximation to Y using a particular model. Obviously, the solution to the problem will depend on the criterion of optimality chosen. We would like to consider this proplem using the criterion of mimizing the mean square error of the difference between Y and its approximation. (Mean square estimation of Minimum mean square error model (MSE)). The results are the basis for a great number of applications in signal processing.

A) Consider the MSE problem where we want to approximate (estimate) a random variable Y by a constant c:

Choose c so that to minimize $||Y-c||^2$

Now, $||Y-c||^2 = E\{|Y-c|^2\} = E\{|Y|^2\} + |c|^2 - 2Re[c^* E\{Y\}]$

and by placing

$$\frac{d}{dc}||Y-c||^2 = 0 \Rightarrow c^* - [E\{Y\}]^* = 0 \Rightarrow c = E\{Y\}$$

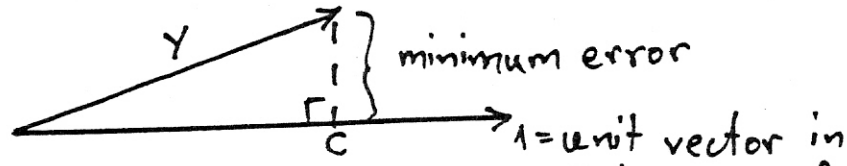
In other words, the best constant approximation of a random variable is its mean

Also, $min||Y-c||^2 = E\{|Y-E\{Y\}|^2 = Var\{Y\} = \sigma_y^2$

Geometric interpretation

Let consider a constant as a r.v. which maps the entire sample space to a constant value. Then the entire collection of such r.v.s constitute a 1-D subspace of the vector space of r.v.s

Then geometrically the best estimate is the projection of Y on the subspace

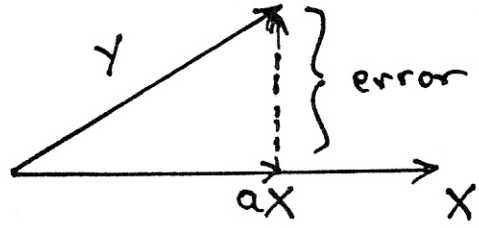


$$c =< 1, Y > 1 = E\{Y \cdot 1\} = E\{Y\}$$

Note that the error $Y - c$ such that

$$\boxed{< Y - c, 1 >= 0}$$

B) Consider now the problem where we want to approximate Y with a scaled version of another r.v. X i.e., aX. Then



$$\begin{aligned}
&minimizing \quad ||Y - aX||^2 \Rightarrow error\ orthogonal\ to\ aX \\
&\Rightarrow < Y - aX, aX >= 0 \Rightarrow E\{(Y - aX)aX^*\} = 0 \\
&\Rightarrow aE\{X^*Y\} - a^2E\{|X|^2\} = 0 \Rightarrow E\{X^*Y\} = a \cdot E\{|X|^2\} \\
&\Rightarrow a = \frac{E\{X^*Y\}}{E\{|X|^2\}} \overset{zero\ mean\ process}{\Rightarrow} a = \frac{Cov(X,Y)}{Var(X)} = \rho\frac{\sigma_Y}{\sigma_X}
\end{aligned}$$

and

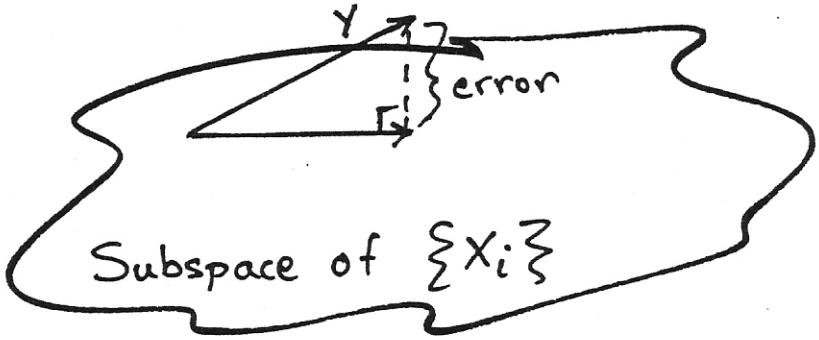
$$min||Y - aX||^2 = \sigma_Y^2 + \rho^2\sigma_Y^2 \quad - \sigma_X^2 2aRe(a) \overset{real\ a}{=} \sigma_Y^2(1 - \rho^2)$$

C) The most general form of this type of problem is to approximate the r.v. Y by a linear combination of n r.v.s $X_1, X_2, \ldots X_n$, i.e., to choose $a_1, a_2, \ldots, a_n$ so that.

$$||Y - (a_1X_1 + \ldots + a_nX_n)||^2$$

is minimized (Linear mean square estimation problem (LMSE)).

To solve this problem we note that the set of r.v of the form $a_1X_1 + \ldots + a_nX_n$ constitute a subspace of the vector space of r.v.s. Geometrically, the solution on of minimizing the magnitude squared of the error quantity $(Y - \sum_{i=1}^{n} a_iX_i)$ is obtained when the error becomes orthogonal to the approximation $\sum_{i=1}^{n} a_iX_i$. i.e.,



Thus, $min||Y - (a_1X_1 + \ldots + a_nX_n)||$ is equivalent to

$$\boxed{< Y - \sum_{i=1}^{n} a_iX_i, \sum_{i=1}^{n} a_iX_i >= 0}$$

This is the projection theorem for the LMSE problem

$$\Rightarrow E\{(Y - \sum_{i=1}^{n} a_iX_i) \sum_{j=1}^{n} a_j^* X_j^*\} = 0 \Rightarrow \sum_{j=1}^{n} a_j^* [E\{X_j^* \cdot \{Y - \sum_{i=1}^{n} a_iX_i\}] = 0$$

Since this must be true for all $\{a_j\}_{j=1,\ldots n}$

$$\Rightarrow E\{X_j^* \cdot (Y - \sum_{i=1}^{n} a_iX_i)\} = 0 \Rightarrow E\{X_j^* Y\} = \sum_{i=1}^{n} a_i E\{X_j^* X_i\}, \qquad j = 1, 2, \ldots, n$$

For X,Y jointly WSS r.vs

$$\boxed{\sum_{i=1}^{n} a_i R_x(i-j) = R_{xy}(-j) \quad j=1,\ldots,n} \quad \textit{Wiener - Hopf equations}$$

or

$$\begin{bmatrix} R_x(0) & R_x(1) & \ldots & R_x(n-1) \\ R_X(-1) & R_x(0) & \ldots & . \\ R_x(-2) & R_x(-1) & \ldots & . \\ \vdots & \vdots & \vdots & \vdots \\ R_x(-n+1) & \ldots & \ldots & R_x(-1)R_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} R_{xy}(-1) \\ R_{xy}(-2) \\ \vdots \\ R_{xy}(-n) \end{bmatrix}$$

Normal equations

or

$$\mathbf{R_x \cdot a = r_{xy}}$$

Thus, $\boxed{\mathbf{a = R_x^{-1} r_{xy}}}$

It is easy to show that $Min||Y - \sum_{i=1}^{n} a_i X_i|| = \sigma_y^2 - \mathbf{r_{xy}^{*T} R_x^{-1} r_{xy}} = \sigma_y^2 - \underline{r}_{xy}^{*} \cdot \underline{a}$

D) Let us consider now the more general problem where instead of a linear combination of r.vs, we employ a more general model. For example we might seek a function $g(X)$ so that $||Y - g(X)||^2$ is minimized. Recall that

$$E\{|Y - g(X)|^2\} = E\{E\{|Y - g(X)|^2/X\}$$

Therefore, the problem becomes equivalent to minimizing $E\{|Y - g(X)|^2/X = x\}$. Then, from the results of case A, we conclude that the absolute mean square error is minimized when

$$\boxed{g(X) = E\{Y/X\}}$$

WIENER FILTERING PROBLEMS

We observe a possibly non-stationary process $X(n)$ over an interval I. We want to find the LMMSE of a process $d(n)$ (desired response) based on a linear combination of the r.v.s that constitute $X(n)$. The process $X(n)$ is related to some other "signal" process $S(n)$ .....

PROBLEM

1) Filtering of signal in noise: $x(n) = s(n) + n(n)$, $d(n) = s(n)$

2) Prediction of signal in noise: $x(n) = s(n) + n(n)$, $d(n) = s(n+p)$, $p > 0$

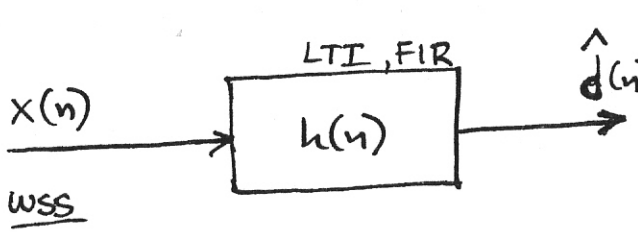
3) Smoothing of signd in noise: $x(n) = s(n) + n(n)$, $d(n) = s(n-q)$, $q > 0$

4) Linear prediction : $x(n) = s(n-1)$, $d(n) = s(n)$

5) General nonlinear problem: $x(n) = G(s(n), n(n))$, $d(n) = s(n)$

6) Deconvolution : $x(n) = s(n) * h(n) + n(n)$, $d(n) = s(n)$

WIENER-HOPF EQUATION – STATIONARY SIGNALS



$$\hat{d}(n) = \sum_{\ell=0}^{P-1} h(\ell)\, x(n-\ell) \quad , \quad e(n) = d(n) - \hat{d}(n)$$

Then, minimization of $E\{|d(n)-\hat{d}(n)|^2\}$ results in

$$\left[\sum_{\ell=0}^{P-1} h(\ell) \cdot R_x(i-\ell) = R_{xd}(i) \right] \quad , \quad i = 0, 1, \ldots, P-1.$$

where $R_x(m) \triangleq E\{x^*(n)\, x(n+m)\}$, $R_{xd}(m) \triangleq E\{x^*(n)\, d(n+m)\}$

Minimum mean square error: $\sigma_e^2 = \sigma_d^2 - \sum_{\ell=0}^{P-1} h(\ell) \cdot R_{xd}(\ell) = R_d(0) - \sum_{\ell=0}^{P-1} h(\ell) R_{xd}(\ell)$

Solution:

$$\begin{bmatrix} R_x(0) & R_x(-1) & \cdots & R_x(1-p) \\ R_x(1) & R_x(0) & \ddots & \\ \vdots & & \ddots & R_x(-1) \\ R_x(p-1) & \cdots & & R_x(0) \end{bmatrix} \cdot \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(p-1) \end{bmatrix} = \begin{bmatrix} R_{xd}(0) \\ \vdots \\ R_{xd}(p-1) \end{bmatrix}$$

WIENER-HOPF EQUATION - NON-CAUSAL CASE

$$\hat{d}(n) = \sum_{l=-\infty}^{\infty} h(l)\, x(n-l) \Rightarrow \sum_{l=-\infty}^{\infty} h(l)\, R_x(i-l) = R_{xd}(i) \Rightarrow$$

$$\Rightarrow \quad h(i) * R_x(i) = R_{xd}(i)$$

$$\Rightarrow \quad \boxed{H(z) = \frac{S_{dx}(z)}{S_x(z)}} \qquad \left\{ \text{based on definition } R_{xd}(m) = E\{x^*(n)\, d(n+m)\} \right\}$$

The non-causal equation provides insight into the operation of the wiener filter.

Causal case: $(h(l)=0,\ l<0)$ Assuming that $S_x(z)$ can be written as $S_x(z) = K H_{ca}(z) H_{ca}^*(1/z^*)$ where $H_{ca}(z)$ is a causal stable system with causal stable inverse.

Then, $$H(z) = \frac{1}{K H_{ca}(z)} \left[\frac{S_{dx}(z)}{H_{ca}^*(1/z^*)} \right]_+$$ where $[\ ]_+$ corresponds only to the causal part of the respon

SPECTRAL FACTORIZATION

Given a random process $x(n)$ with continuous PSD $S_x(\omega)$:

If $\left[\int_{-\pi}^{\pi} |\ln S_x(\omega)|\, d\omega < \infty\right]$ then we can write. $\left[S_x(z) = K \cdot H_{ca}(z)\, H_{ca}^*(1/z^*)\right]$

where, K is a positive constant, $H_{ca}(z)$ is a causal stable system with causal stable inverse.

- If $S_x(z)$ is a rational polynomial, $H_{ca}(z)$ is minimum phase, $H_{ca}^*(1/z^*)$ maximum phase systems.
- The above condition is known as the Paley-Wiener condition.

Modeling of $x(n)$:

"innovations process"

$w(n)$ → $H_{ca}(z)$ → $x(n)$: "regular process"

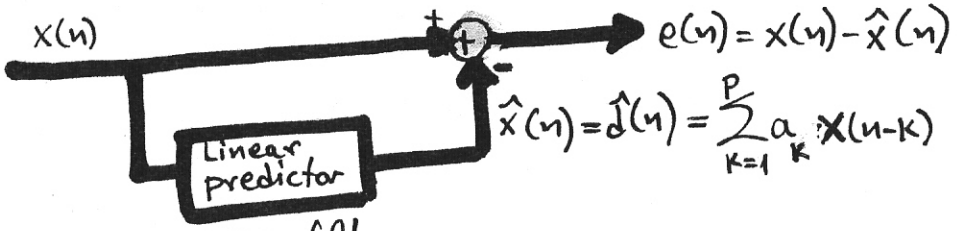
white noise
$S_w(z) = K$

$$S_x(z) = S_w(z) \cdot H_{ca}(z) \cdot H_{ca}^*(1/z^*)$$

Innovations Representation of a r. process.

$x(n)$ → $H_{ca}^{-1}(z)$ → $w(n)$: innovations process (white noise)

LINEAR PREDICTION (FORWARD)



Objective: predict value of $x(n)$ at time n based on P past observations at $n-1, n-2, \ldots n-P$

$e(n)$: forward prediction error $= x(n) - \sum_{k=1}^{P} a_k x(n-k)$

$d(n) = x(n)$

Minimization of $E\{|d(n)-\hat{d}(n)|^2\}$

results in

① $\left[\sum_{k=1}^{P} a_k R_x(i-k) = R_x(i)\right]$, $i=1,2,\ldots,P$

② $\sigma_\varepsilon^2 = R_x(0) - \sum_{k=1}^{P} a_k R_x(k)$

① + ② → "Normal Equations"

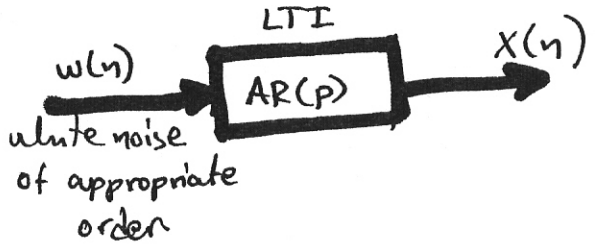
Whitening property of of prediction filter.

Calculation of $\{a_k\}$ is based on the Correlation properties of $x(n)$. The prediction filter estimates $x(n)$ by utilizing all correlation information between $x(n)$ and past observations. Thus, the minimum mean square error corresponds to a white sequence. $e(n)$

CAUSAL AR modeling of a r. process

Given a r. pr. $x(n)$ generated from an AR(p) stable model.
(autoregressive)

$$\left[x(n) + \sum_{k=1}^{p} a_k x(n-k) = w(n) \right] \quad : \quad$$



* Multiply both sides by $x^*(n-\ell)$ and take expectation. Then

$$E\{x(n)x^*(n-\ell)\} + \sum_{k=1}^{p} a_k E\{x(n-k)x^*(n-\ell)\} = E\{w(n)x^*(n-\ell)\} \Rightarrow R_x(\ell) + \sum_{k=1}^{p} a_k R_x(\ell-k) = \begin{cases} \sigma_w^2, & \ell=0 \\ 0, & \ell>0 \end{cases}$$

This is nothing else but the normal equations

*