SOLVED EXERCISES

Index

CHAPTER I

I.8.1.-

$$\underline{\underline{R}}_{n+1} = \beta.\underline{\underline{R}}_n + \alpha.\underline{X}_n.\underline{X}_n^H$$

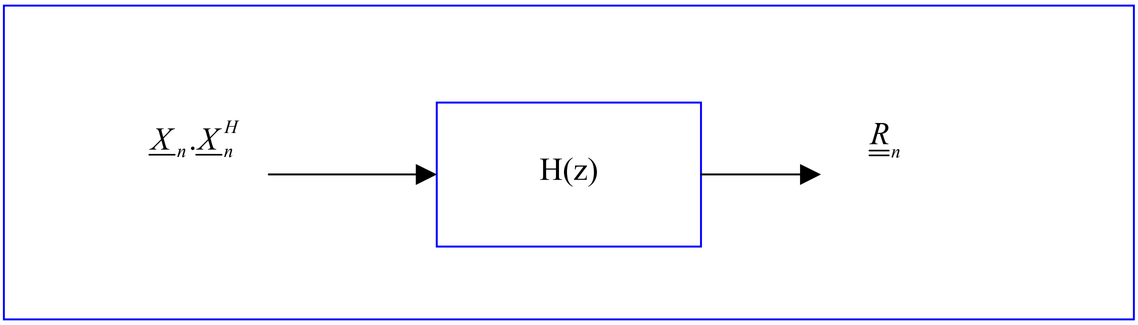
a) Each factor $\underline{X}_m.\underline{X}_m^H$ contributes to the estimate with a coefficient equal to $\alpha.\beta^{n-m}$. This is easy to check just by considering the IIR filter with numerator equal to α and denominator equal to $1-\beta z^{-1}$. The effective length is defined when the coefficient goes down to a given value, let us say 1/e, in consequence the effective length will be defined as:

$$\beta^N = 1/e \qquad \text{or} \qquad N = -\frac{1}{Ln(\beta)}$$

when β is close to one

$$-Ln(\beta) \cong (1-\beta) \qquad \text{and} \qquad N \approx \frac{1}{1-\beta}$$

b) $H(z) = \dfrac{\alpha}{1-\beta z^{-1}}$



c) $H(z) = \alpha.\sum_{q=0}^{\infty} \beta^q.z^{-q}$ and $\underline{\underline{R}}_n = \sum_{q=0}^{\infty} \alpha.\beta^q.\underline{X}_{n-q}.\underline{X}_{n-q}^H$

d) $E(\underline{\underline{R}}_{n+1}) = \beta.E(\underline{\underline{R}}_n) + \alpha.E(\underline{X}_{n+1}.\underline{X}_{n+1}^H)$

when the random processes forming the components of the original vectors are stationary, the expected value of the input dot product will be the actual autocorrelation matrix $\underline{\underline{R}}$ and the expected values of the first two terms will coincide.

Thus,

$$E(\underline{\underline{R}}_n) = \frac{\alpha}{(1-\beta)}.E(\underline{X}_{n+1}.\underline{X}_{n+1}^H) = \frac{\alpha}{(1-\beta)}.\underline{\underline{R}}$$

Obviously, to remove bias and ensuring that the expected value of the estimate tends to the actual value α has to be equal to 1-β.

e)

The inverse lemma is:

If $\underline{\underline{R}} = (\underline{\underline{A}} + \underline{\underline{B}}.\underline{\underline{C}}.\underline{\underline{D}})$ then $\underline{\underline{R}}^{-1} = \underline{\underline{A}}^{-1} - \underline{\underline{A}}^{-1}.\underline{\underline{B}}.(\underline{\underline{D}}.\underline{\underline{A}}^{-1}.\underline{\underline{B}} + \underline{\underline{C}}^{-1})^{-1}.\underline{\underline{D}}.\underline{\underline{A}}^{-1}$

A particular case is the Woodbury's identity

$$(\underline{\underline{A}} + \delta.\underline{a}.\underline{a}^H)^{-1} = \underline{\underline{A}}^{-1} - \frac{\delta}{1+\delta.\underline{a}^H.\underline{\underline{A}}^{-1}.\underline{a}}.\underline{\underline{A}}^{-1}.\underline{a}.\underline{a}^H.\underline{\underline{A}}^{-1}$$

Using the second one (much easy to use in our case) we get:

$$\underline{\underline{R}}_{n+1}^{-1} = \frac{1}{\beta}\left(\underline{\underline{R}}_n^{-1} - \frac{\alpha^2.\underline{\underline{R}}_n^{-1}.\underline{X}_{n+1}.\underline{X}_{n+1}^H.\underline{\underline{R}}_n^{-1}}{\beta + \alpha^2\underline{X}_{n+1}^H.\underline{\underline{R}}_n^{-1}.\underline{X}_{n+1}}\right)$$

I.8.2.-

a) $\underline{S}_0.\underline{S}_0^H$ This is a rank-one matrix (all its columns are proportional each other). Being rank one it has only an eigenvalue different from zero. The eigenvector associated to this non cero eigenvalue is precisely $\underline{S}_0$ normalized, i.e.

$$\left(\underline{S}_o.\underline{S}_o^H\right).\frac{\underline{S}_o}{\|\underline{S}_o\|} = \|\underline{S}_0\|^2.\frac{\underline{S}_o}{\|\underline{S}_o\|}$$

The magnitude square of the vector is the eigenvalue.

$\lambda_2, \ldots, \lambda_Q$ are cero and the eigenvector are any set of orthonormal vectors each other and with $\underline{S}_o$

b) Any set of orthonormal vectors are valid eigenvectors of the identity matrix. The eigenvalues are all equal to one.

c) The largest eigenvalue and the eigenvector associated to it is given by the rank one contribution.

$$\left(\alpha\underline{S}.\underline{S}^H + \sigma^2.\underline{\underline{I}}\right).\frac{\underline{S}}{\|\underline{S}\|} = \left(\alpha\|\underline{S}\|^2 + \sigma^2\right).\frac{\underline{S}}{\|\underline{S}\|}$$

the rest of eigenvectors have to be orthogonal to this largest eigenvector. In consequence when multiplying any of them by the rank one part the dot product will be cero and:

$$\left(\alpha \underline{S}.\underline{S}^H + \sigma^2.\underline{\underline{I}}\right).\underline{e}_q = \sigma^2.\underline{\underline{I}}.\underline{e}_q = \sigma^2.\underline{e}_q ; \forall q = 2, Q$$

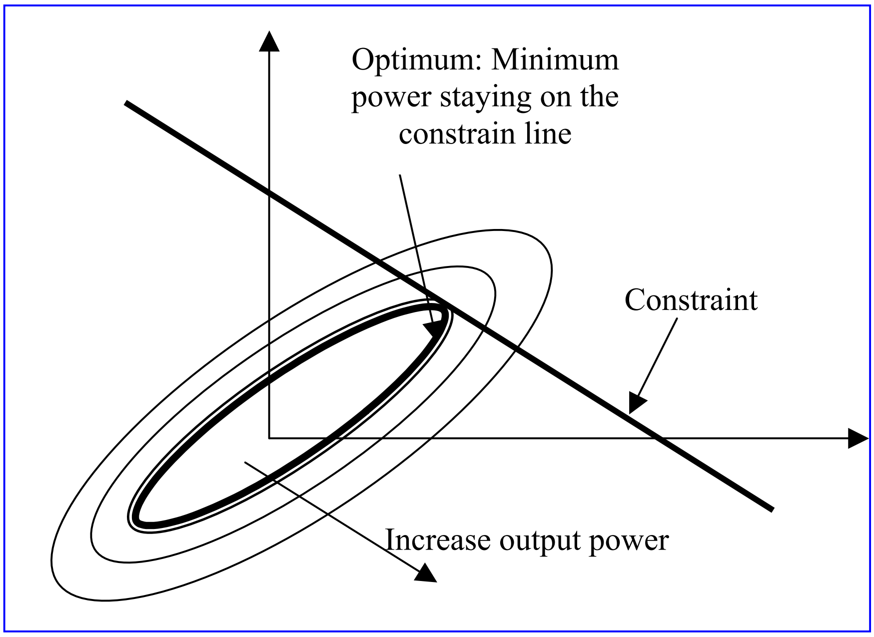
All the eigenvalues are equal to the contribution of the identity matrix. Only the first one has the contribution of the rank one matrix in addition.

I.8.3 .-

The d.c. response equal to 0 dB. Is set as $\underline{A}^H.\underline{1} = 1$, where vector $\underline{1}$ contains 1 at every component, i.e. $\sum_{q=0}^{Q-1} a(q).1 = 1$

The output filter power is given by $\underline{A}^H.\underline{\underline{R}}.\underline{A}$, where matrix R is the autocorrelation matrix of the input signal, i.e. $\sum_{q=0}^{Q-1}\sum_{p=0}^{Q-1} a^*(q).a(p).r(q,p)$

In particular, for q equal 2, the d.c. constraint is a line in the 2-D space where the two coefficients stay, and the power is defined by ellipses centered at the origin, which size is proportional to the output power of the filter under design.



If the input is white noise its autocorrelation matrix is the identity and the design evolves to:

$$\underline{A}^H.\underline{1} = 1 \quad \underline{A}^H.\underline{A}\Big|_{MIN}$$

The Lagrangian (i.e. the objective minus a multiplier per every constraint) is $\Upsilon = \underline{A}^H.\underline{A} - \lambda.\left(\underline{A}^H.\underline{1} - 1\right)$. Taking derivatives with respect the real and the imaginary parts of the filter coefficients (equivalent to take derivatives with respect $\underline{A}^H$) and setting them to cero, results on

$\underline{A} - \lambda.\underline{1} = \underline{0}$ and $\underline{A} = \lambda.\underline{1}$

The Lagrange multiplier is found from the constraint equation

$$\lambda = \frac{1}{\underline{1}^H.\underline{1}} = \underline{1}/Q = \underline{1}/2$$

b) since $\underline{A}^H.\underline{X}_n = \underline{A}^H.\left(A.\underline{1} + \underline{w}_n\right) = A + \underline{A}^H.\underline{w}_n$.

Signal power A^2

Noise power $E\left(\left|\underline{A}^H.\underline{w}_n\right|^2\right) = E\left(\underline{A}^H.\sigma^2.\underline{\underline{I}}.\underline{A}\right) = \sigma^2.\underline{A}^H.\underline{A} = \sigma^2$

$SNR = 2.\dfrac{A^2}{\sigma^2}$ A gain of 3 dB.

c) When the noise is not white, the constraint is maintained but the objective to be minimized changes. The solution to the new problem is given by

$$\underline{A} = \frac{\underline{\underline{R}}_w^{-1}.\underline{1}}{\underline{1}^H.\underline{\underline{R}}_w^{-1}.\underline{1}}$$ and $\underline{A}^H.\underline{X}_n = \underline{A}^H.\left(A.\underline{1} + \underline{w}_n\right) = A + \underline{A}^H.\underline{w}_n$ still holds

Signal power A^2

Noise Power $E\left(\left|\underline{A}^H.\underline{w}_n\right|^2\right) = E\left(\underline{A}^H.\underline{\underline{R}}_w.\underline{A}\right) = \dfrac{1}{\underline{1}^H.\underline{\underline{R}}_w^{-1}.\underline{1}}$

$SNR = A^2.\left(\underline{1}^H.\underline{\underline{R}}_w^{-1}.\underline{1}\right)$

For the given acf matrix,

$$\underline{\underline{R}}_w^{-1} = \frac{1}{\sigma^4 - \gamma^2}.\begin{pmatrix} \sigma^2 & -\gamma \\ -\gamma & \sigma^2 \end{pmatrix} \quad ; \quad \underline{1}^H.\underline{\underline{R}}_w^{-1} = \frac{\left[\sigma^2 - \gamma \quad \sigma^2 - \gamma\right]}{\sigma^4 - \gamma^2} = \left[\frac{1}{\sigma^2 + \gamma} \quad \frac{1}{\sigma^2 + \gamma}\right] \quad ;$$

$$\underline{1}^H.\underline{\underline{R}}_w^{-1}\underline{1} = \frac{2}{\sigma^2 + \gamma}$$

finally

$$SNR = 2.\frac{A^2}{\sigma^2}.\left[\frac{1}{1 + \left(\dfrac{\gamma}{\sigma^2}\right)}\right]$$ ¡Increasing correlation decreases gain from the maximum of 3 dB. For the white noise case!

Just in case the white noise (sub-optimum) filter is used, the SNR is:

$$SNR = \frac{A^2}{\sigma^2}.\left[\frac{1}{1+\left(\frac{\gamma}{\sigma^2}\right)}\right]$$ with a permanent loss of 3 dB. With respect the optimum gain. This loss increases when increasing the filter order Q.

I.8.4 .-

Given a, we can select it as one of axis for the new orthogonal space. Once this is decided, the next vector, to be orthogonal to it, must lie in the subspace orthogonal to a which is defined by the projection operator as follows:

$$\underline{\underline{A}}_{\perp} = \left(\underline{\underline{I}} - \frac{\underline{a}.\underline{a}^H}{\underline{a}^H.\underline{a}}\right)$$ In consequence, the second vector is given by the projection of it in the subspace orthogonal to a

In summary, the orthogonal basis defined by vectors a and b is:

$$\underline{e}_1 = \underline{a}$$

$$\underline{e}_2 = \left(\underline{\underline{I}} - \frac{\underline{a}.\underline{a}^H}{\underline{a}^H.\underline{a}}\right).\underline{b}$$

Yes. This procedure is independent of the components forming each original vector.

Furthermore, given a set of Q vectors of dimension P greater than Q, the orthogonal projection is formed as:

Given $\underline{a}_1, \underline{a}_2, ..., \underline{a}_Q$

Define $\underline{\underline{\alpha}} = \left[\underline{a}_1, \underline{a}_2, ..., \underline{a}_Q\right]$

Then $\underline{\underline{A}}_{\perp} = \left[\underline{\underline{I}}_{PxP} - \underline{\underline{\alpha}}.\left(\underline{\underline{\alpha}}^H.\underline{\underline{\alpha}}\right)^{-1}.\underline{\underline{\alpha}}^H\right]$

I.8.5 .-

The IRR equation is $y(n) + \sum_{p=1}^{P} a(p).y(n-p) = \sum_{q=0}^{Q} b(q).x(n-q)$, it can be written in vector form as: $\underline{y}_n^H.\underline{a} = \underline{x}_n^H.\underline{b}$ where the first component of vector a is equal to one.

Now setting x(n)=δ(n) then y(n)=h(n), and arranging the vector equation in matrix forms

$$\begin{bmatrix} h(0 & 0 & 0 & .. & 0 \\ h(1) & h(0) & 0 & .. & 0 \\ h(2) & h(1) & h(0) & .. & 0 \\ .. & .. & .. & .. & \\ h(P-1) & h(P-2) & h(P-3) & .. & h(0) \\ h(P) & h(P-1) & h(P-2) & .. & h(1) \\ .. & .. & .. & .. & .. \end{bmatrix} . \begin{bmatrix} 1 \\ a(1) \\ .. \\ a(P) \end{bmatrix} = \begin{bmatrix} 1 & 0 & .. & 0 \\ 0 & 1 & .. & 0 \\ 0 & 0 & .. & 0 \\ 0 & 0 & .. & 1 \\ 0 & 0 & .. & 0 \\ 0 & 0 & .. & 0 \\ .. & .. & .. & .. \end{bmatrix} . \begin{bmatrix} b(0) \\ b(1) \\ .. \\ b(Q) \end{bmatrix}$$

From this formula matrix $\underline{\underline{H}}$ is easily identified.

I.8.6 .-

Being vector $\underline{P}=\left[p(0) \quad p(1) \quad .. \quad p(Q-1)\right]^T$ the idea is to design a filter of Q coefficients $\underline{A}$ such that the response to the filter is a given level (let us say 1 or 0 dB.) and to any other signal w(t) with acf equal to $\underline{\underline{R}}$ the power of its response is minimal. In the case of white noise, this traduces in the norm of the filter vector. In summary the formulation will be the following:

$$\underline{A}^H.\underline{P}=1$$
$$\underline{A}^H.\underline{A}\Big|_{\min}$$

(see I.8.3 for details) The optimum is given by $\underline{A}=\dfrac{\underline{P}}{\|\underline{P}\|^2}$

The signal to noise ratio is, for any filter design, equal to: $SNR=\dfrac{\left|\underline{A}^H.\underline{P}\right|^2}{\sigma^2.\|\underline{A}\|^2}$. Using the optimum filter, the resulting SNR is:

$$SNR_{\max}=\frac{1}{\sigma^2/\|\underline{P}\|^2}=\frac{Pulse\ Energy}{Spectral\ density\ of\ white\ noise}$$

I.8.7 .-

a) Yes, since is the quotient of the power at the filter output for signal and noise respectively.
b) Since $\underline{\underline{R}}_s=\underline{\underline{R}}-\underline{\underline{R}}_n$ (Whenever signal and noise are uncorrelated) then $SNR=\dfrac{\underline{a}^H.\underline{\underline{R}}.\underline{a}}{\underline{a}^H.\underline{\underline{R}}_n.\underline{a}}-1$ In consequence the use of the full correlation matrix is correct for design purposes.
c)

$$SNR=\frac{\underline{a}^H.\underline{\underline{R}}.\underline{a}}{\underline{a}^H.\underline{\underline{R}}_n.\underline{a}}=\frac{\underline{a}^H.\underline{\underline{R}}_n^{1/2}.\left(\underline{\underline{R}}_n^{-1/2}.\underline{\underline{R}}.\underline{a}\right)}{\underline{a}^H.\underline{\underline{R}}_n.\underline{a}}=$$
$$=\frac{\underline{u}^H.\underline{v}}{\underline{u}^H.\underline{u}}\leq\frac{\sqrt{\left(\underline{u}^H.\underline{u}\right)\left(\underline{v}^H.\underline{v}\right)}}{\underline{u}^H.\underline{u}}=\sqrt{\frac{\underline{v}^H.\underline{v}}{\underline{u}^H.\underline{u}}}$$

The equality holds when both vectors are co-linear, i.e. $\underline{u}=\rho^{-1}.\underline{v}$ which provides the design to obtain the maximum SNR. Thus, the design equation is:

$$\underline{\underline{R}}_n^{1/2}.\underline{a}=\frac{1}{\rho}.\underline{\underline{R}}_n^{-1/2}\underline{\underline{R}}.\underline{a} \quad or \quad \underline{\underline{R}}.\underline{a}=\rho.\underline{\underline{R}}_n.\underline{a}$$

which is the generalized eigen decomposition of the couple $\left[\underline{\underline{R}} \quad \underline{\underline{R}}_n\right]$. At the same time, the maximum of the SNR for this design is:

$SNR_{\max} = \rho$ Then, the maximum eigenvalue of the eigen-decomposition is the solution and the optimum filter is given by the maximum eigenvector.

d) Since $\left(\underline{\underline{R}}_s + \underline{\underline{R}}_n\right).\underline{a} = (\rho - 1).\underline{a}$ the objective still is the maximum eigenvalue

e) Easily concluded from the previous section

f) When $\underline{\underline{R}}_s$ is rank-one, i.e. $\underline{\underline{R}}_s = \underline{c}.\underline{c}^H$, the eigendecomposition reduces to $\underline{\underline{R}}_s.\underline{a} = \underline{c}.\left(\underline{c}^H.\underline{a}\right) = \rho.\underline{\underline{R}}_n.\underline{a}$

At this moment, note that any constant multiplying the optimum vector does not affect its optimum character because the SNR does not change.

As can be seen in the previous formula, the eigendecomposition is not longer necessary since the optimum vector, within a constant, can be obtained directly from the equation as:

$\underline{a} \propto \underline{\underline{R}}_n^{-1}.\underline{c}$

g) Just change $\underline{\underline{R}}_n$ by $\sigma^2.\underline{\underline{I}}$ in all sections.

I.8.8.-

Since $Trace\left(\underline{\underline{A}}\right) = \sum eigenvalues$ and all the eigenvalues of a positive definite matrix are positive removing terms in the sum always produce a lower bound of the trace operator.

I.8.9, I.8.10,I.8.11 .-

Use an Algebra text-book like reference [2].

II.13.1 .-

The characteristic function is defined as the Fourier transform of the pdf. A good candidate to estimate this function is to use the less committed pdf, i.e. uniform and to replace the continuous integral by a sum constrained just to the available samples. In summary the estimate of the characteristic function will be:

$\hat{\Phi}(w)=\sum_i \exp(-j.w.x(i))$ Where i covers the set of the N available samples

The expected value of the estimate is

$$E\left\{\hat{\Phi}(w)\right\}=E\left\{\sum_i \exp(-j.w.x(i))\right\}=\sum_i E\left\{\exp(-j.w.x(i))\right\}=$$

$$=\sum_i \int \Pr(x(i)).\exp(-j.w.x(i)).dx(i)=N.\Phi(w)$$

In consequence, to remove bias, the definite formulation of the estimate will be:

$$\hat{\Phi}(w)=\frac{1}{N}\sum_i \exp(-j.w.x(i))$$

Computing the variance,

$$E\left[\left|\Phi(w)-\hat{\Phi}(w)\right|^2\right]=E\left[\left|\hat{\Phi}(w)\right|^2\right]-\left|\Phi(w)\right|^2=$$

$$=\sum_i\sum_j \frac{1}{N^2}.E\left(e^{-jwx(i)}.e^{jwx(j)}\right)-\left|\Phi(w)\right|^2=\begin{vmatrix} when\ x(i) \\ and\ x(j)\ are \\ independent \end{vmatrix}=$$

$$=\sum_i\sum_j \frac{1}{N^2}.E\left(e^{-jwx(i)}\right)E\left(.e^{jwx(j)}\right)-\left|\Phi(w)\right|^2=$$

$$=\sum_i\sum_j \frac{1}{N^2}\left|\Phi(w)\right|^2-\left|\Phi(w)\right|^2=$$

$$=\frac{1}{N}+\frac{N^2-N}{N^2}.\left|\Phi(w)\right|^2-\left|\Phi(w)\right|^2=\frac{1-\left|\Phi(w)\right|^2}{N}$$

It is easy to check that the magnitude of the actual characteristic function is always below one. At the same time note that estimate is consistent since it tends to cero as the length of the data set tends to infinity.

Note also that the direct estimation of the pdf by the histogram with a given weight or window function p(x), is equivalent to the following estimate:

$\hat{\Phi}(w)=P(w).\sum_i \exp(-j.w.x(i))$ where P(w) is the Fourier transform of p(x).

II.13.2 .-

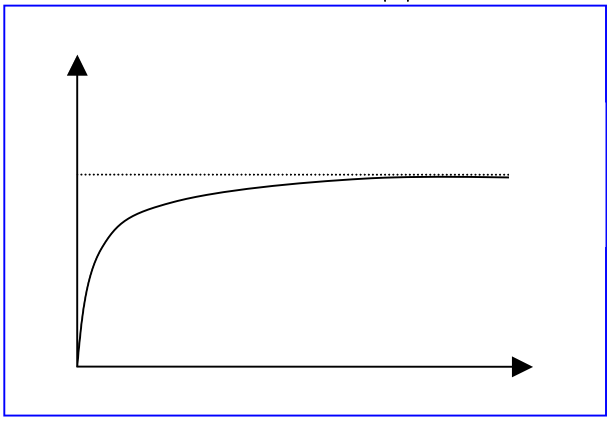
We are looking for a non-linear system which maps x to y such that the second pdf is un

$$\Pr(x)=\frac{\alpha}{2}.\exp(-\alpha.|x|) \qquad \Pr(y)=\frac{1}{2.Y_o}$$

$$\frac{dg}{dx}=\frac{\Pr(x)}{\Pr(y)\big|_{y=g(x)}}=2.Y_o.\frac{\alpha}{2}.\exp(-\alpha.|x|)$$

and

$$g(x)=Y_o.(1-\exp(-\alpha.|x|))$$



II.13.3.-

$\underline{\underline{R}}=\begin{pmatrix}1 & \alpha\\ \alpha & 1\end{pmatrix}$ then $\underline{\underline{R}}^{-1}=\frac{1}{1-\alpha^2}\begin{pmatrix}1 & -\alpha\\ -\alpha & 1\end{pmatrix}$ The ML estimate is $\mu_{ML}=\frac{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{X}_n}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}}$

performing a few computations $\mu_{ML}=\frac{\underline{1}^H.\underline{X}_n}{2}$

¡The estimate is identical to the arithmetic mean estimate, independently of α! In fact both estimates have variance $\sigma^2_{ML}=\frac{1+\alpha}{2}$. In other words, correlation between two consecutive samples does not alter the structure of the estimate but increases variance.

II.13.4 .-

Instantaneous estimate $\breve{\underline{\underline{R}}}_{xy}=\underline{X}_n.\underline{Y}_n^H$ averaged estimate $\hat{\underline{\underline{R}}}_{xy}=\frac{1}{N}\sum_n \underline{X}_n.\underline{Y}_n^H$

Both estimates are unbiased

The corresponding estimates of the cross-spectral density are

$$\breve{S}_{xy}(w)=\frac{\underline{S}^H.\underline{\underline{R}}_{xy}.\underline{S}}{\underline{S}^H.\underline{S}}=\frac{1}{Q}.\left(\underline{S}^H.\underline{X}_n\right).\left(\underline{Y}_n^H.\underline{S}\right)$$

$$\hat{S}_{xy}(w)=\frac{\underline{S}^H.\underline{\underline{R}}_{xy}.\underline{S}}{\underline{S}^H.\underline{S}}=\frac{1}{Q}.\frac{1}{N}\sum_n\left(\underline{S}^H.\underline{X}_n\right).\left(\underline{Y}_n^H.\underline{S}\right)$$

where Q is the length of the segments and N is the global (overlapped or not to reduce variance) number of available segments.

The use of the biased or unbiased estimate for $r_{xy}(n)$ is a relevant issue in this case since the Fourier transform is not longer a positive defined function. Windows data dependent or independent can be used for the cross-functions in the same framework that they were defined for auto-functions within this chapter.

I.3.5 .-

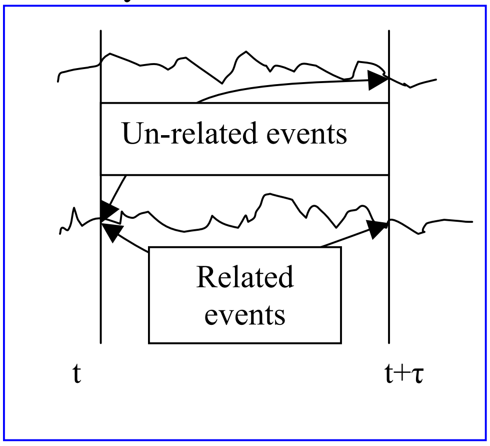
The definition of the acf function of a random process is $r_x(t,\tau) = E\left[x_{\xi 1}(t).x_{\xi 2}(t+\tau)\right]$ which in its integral form $\iint x_{\xi 1}(t).x_{\xi 2}(t+\tau).\Pr\left(\xi 1,\xi 2,t,t+\tau\right).d\xi 1.d\xi 2$.

The probability function can be re-written as a function of θ as

$$\Pr(\xi 1,\xi 2,t,t+\tau) = \Pr\left(\theta = \theta_1 / t, \theta = \theta_2 / t+\tau\right) =$$
$$= \Pr(\theta_1).\delta\left(\theta_2 - \theta_1\right)$$

This second line formulation obeys to the fact that the parameter defines a realization of the random process promoting that, unless we stay on the same realization, the expected value of the product defining the acf is cero. In summary the double integral reduces to a single integration as follows:

$$r_x(t,\tau) = \int x_\theta(t).x_\theta(t+\tau).\Pr(\theta).d\theta$$



II.13.6 .-

$$x(t) = a.\cos(w.t+\theta)$$
$$E(x(t)) = E_a(a).E_\theta(\cos(w.t+\theta))$$
$$E_\theta(\cos(w.t+\theta)) = \operatorname{Re}\left[\int \exp(jwt).\exp(j\theta).\Pr(\theta).d\theta\right] =$$
$$= \operatorname{Re}\left[\exp(jwt).\int \exp(j\theta).\Pr(\theta).d\theta\right]$$

The integral is the characteristic function of the phase evaluated at Ω=1. Since the phase is uniformly distributed between –π and π this integral is cero.

And $E(x(t)) = E_a(a).0 = 0 \quad \forall t$

In summary, the random process is stationary in mean either if the mean of the amplitude is cero or the characteristic function of the phase is equal to cero when evaluated at frequency equal to one ($\Phi\left(\Omega=1\right)=FT(\Pr(\theta))_{\Omega=1}=0$).

The acf function:

$$E\left[x(t).x(t+\tau)\right]=E_a(a^2).E_\theta\left(cos(wt+\theta).\cos(w(t+\tau)+\theta\right)=$$
$$=\sigma_a^2.E_\theta\left(\frac{cos(w\tau)}{2}+\frac{\cos(2wt+w\tau+2\theta)}{2}\right)=$$
$$=\sigma_a^2.\frac{cos(w\tau)}{2}$$

Since

$$E_\theta\left[\cos(2wt+w\tau+2\theta)\right]=$$
$$=\mathrm{Re}\left(\exp(j2wt+jw\tau).\Phi(\Omega=2)\right)$$

again being the phase distribution uniform the characteristic function evaluated at frequency 2 is cero.

Thus, the random process is stationary in its acf whenever the characteristic function of the phase is cero at frequency equal 2.

b) When the amplitude is deterministic, it is easy to check that

$$E(x)=0 \quad r_x(t,\tau)=\frac{a^2}{2}.\cos(w\tau)$$

When w is also a random variable, independent of the phase, remains one steep in computing the acf which is to perform the expected value with respect the frequency.

$$E_w(\cos(w\tau))=\int\cos(w\tau).\Pr(w).dw=$$
$$=\mathrm{Re}\left(\Phi_w(\Omega=\tau)\right)=\left|\begin{matrix}\textit{When the}\\ \textit{pdf function}\\ \textit{is even}\end{matrix}\right|=$$
$$=\int\exp(-w\tau).\Pr(w).dw$$

In summary, $r_x(\tau)=\frac{\sigma_a^2}{2}.\underset{Transform}{Fourier}\left(\Pr(w)\right)$

And the power spectral density $S_x(w)=\frac{\sigma_a^2}{2}.\Pr(w)$

This explains why the bandwidth of wideband FM coincides with the dynamic range of the instantaneous frequency.

II.13.7 .-

The MEM extrapolation of the acf is just to assume perfect AR modeling and to continue ahead with the Y-W equations, i.e.

$$\begin{bmatrix} 1 & 0.5 & 0.1 \\ 0.5 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \\ r_3 & 0.1 & 0.5 \\ .. & .. & .. \end{bmatrix} . \begin{bmatrix} 1 \\ a(1) \\ a(2) \end{bmatrix} = \begin{bmatrix} \sigma_2^2 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Levinson's algorithm

$a_1^1 = -0.5 \quad K_1 = -0.5 \quad \sigma_1^2 = 1.(1-0.25) = 0.75$

$\Delta_2 = 0.1 + 0.5(-0.5) = -0.15$

$K_2 = \dfrac{0.15}{0.75} = 0.2$

$a_1^2 = -0.5 + 0.2(-0.5) = -0.49$

$a_2^2 = K_2 = 0.2$

$\sigma_2^2 = 0.75(1-0.2^2) = 0.72$

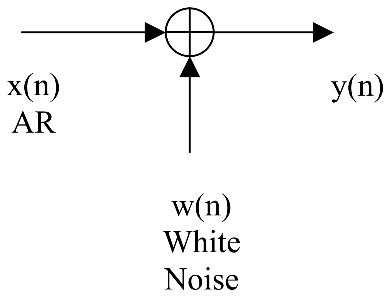
MEM extrapolation for r_3

$r_3^{MEM} + 0.1a_1^2 + 0.5a_2^2 = 0 \quad r_3^{MEM} = 0.039$

Matlab© routine for Levinson algorithm

```
%       File LEV.M
%
%   Given the set of acf values, the function return in vectors cor(.),
% par(.) and err(.) the predictor coefficients, the nq-1 parcors, and the nq
% prediction error power for each successive predictor order.
%
% M.A. Lagunas
%------------------------------------------------------------------------
function [coe,par,err,nq]=lev(r);
nq=length(r);
err(1)=r(1);
coe(1)=1;coe(2)=-r(2)/r(1);par(1)=coe(2);
err(2)=r(1)+coe(2)*r(2);
for ii=3:nq;
   delta=coe*r(ii:-1:2)';par(ii-1)=-delta/err(ii-1);
   err(ii)=(1-par(ii-1)*par(ii-1))*err(ii-1);
   coe(ii)=par(ii-1);
   for j=2:round((ii)/2);
      aus=coe(j);bus=coe(ii-1-j+2);
      coe(j)=aus+par(ii-1)*bus;coe(ii-1-j+2)=aus*par(ii-1)+bus;
   end;
end;
%------------------------------------------------------------------------
```

II.3.8 .-

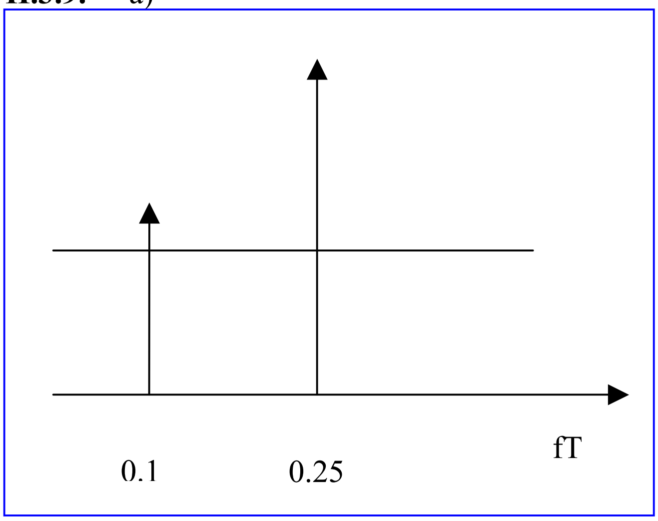


$$S_y(w) = \left| \begin{matrix} For\ \{x\} \\ indepependent \\ of\ \{w\} \end{matrix} \right| = S_x(w) + S_w(w) = \frac{\sigma_x^2}{A(z).A(z^{-1})} + \sigma_w^2$$

In summary $S_y(w) = \left. \dfrac{\sigma_x^2 + \sigma_w^2.A(z).A(z^{-1})}{A(z).A(z^{-1})} \right|_{z=\exp(jwT)}$ which results in an AMRA model.

In consequence speech recording, even assuming is pure AR, due to recording noise results in an ARMA model.

II.3.9.- a)



b) $s(n) = x(n) - w(n)$ when using the recursive equation for the deterministic and completely predicted signal s(n) results in a special ARMA model

$$\sum_{q=0}^{Q} a(q).x(n-q) = \sum_{q=0}^{Q} a(q).w(n-q)$$

and

$$x(n)=\frac{1}{a(0)}.\sum_{q=0}^{Q}a(q).w(n-q)-\frac{1}{a(0)}.\sum_{q=1}^{Q}a(q).x(n-q)$$

$$S_x(w)=\frac{A(z)}{A(z)}.\sigma_w^2=\sigma_w^2 \quad ????$$

c) The problem comes from the fact that s(n) is not a random process but a deterministic signal described by a recurrence equation. For this reason it is not longer stationary and has not a time invariant power spectral density profile. Modeling x(n) as an ARMA we impose stationarity and, in consequence, only the stationary part ot x(n) remains.

d) In order to set s(n) as a realization of a random process we need to introduce at least one random variable in its modeling. Assuming that the phase of each sinusoid is a valid candidate, the solution is

$$s(n)=A_1.\cos(2\pi.0.1.n+\theta_1)+A_2.\cos(2\pi.0.25.n+\theta_2)$$

where θ_1 and θ_2 are uniformly distributed r.v.

Of course, for this s(n) the recurrence equation is not longer valid.

II.13.10.-

A modulated signal takes the form o modulating waveform acting over a carrier waveform as $x(t)=a(t).\cos(wt)$. Usually the modulating wave is assumed to be strict sense stationary

$$E\left[a(t)\right]=0 \quad \forall t \quad and \quad E\left[a(t).a(t+\tau)\right]=r_a(\tau) \quad \forall t$$

thus

$$E\left[x(t).x(t+\tau)\right]=r_a(\tau).\cos(wt).\cos(w(t+\tau))$$

Note that the carrier, being deterministic, introduces the non-stationarity

$$r_x(t,\tau)=\frac{r_a(\tau)}{2}.\cos(w\tau)+\frac{r_a(\tau)}{2}.\cos(w(2t+\tau))$$

In accordance with this result, the power of the random process fluctuates at frequency 2w as well as any other moment of the correlation function. Since in engineering the average power is a key design parameter, it is decided just to keep the average term of this time-varying correlation. Formally, it is said that the random process {x} is ciclo-stationary (periodic) and we keep just the first coefficient of its Fourier series. In summary, we define the acf of {x} as:

$$r_x(\tau)\triangleright\frac{1}{T_p}.\int_{T_p}r_x(t,\tau).dt=\frac{r_a(\tau)}{2}.\cos(w\tau)$$

With respect the discrete o digital modulation case,

$$x(t)=Lim_{N\to\infty}\sum_{k=-N}^{N}a(k).p(t-k/r)$$ for $-N/r\le t\le N/r$ and being r the symbol velocity (bauds), a(.) the information symbols draw from a discrete constellation, and p(.) the pulse shape (either full or partial response).

Assuming that the symbol sequence are stationary, it can be defined is acf in the same manner than in the continuous wave modulation case seen previously.

$$E\left[x(t).x(t+\tau)\right] = \lim_{N\to\infty} \sum_{k=-N}^{N} \sum_{n=-N}^{N} E\left[a(k).a(n)\right].p(t-kT).p(t+\tau-nT)$$

with k=m and k-n=l

$$r_x(t,\tau) = \lim_{N\to\infty} \sum\sum r_a(l).p(t-mT).p(t+\tau-(m-l).T)$$

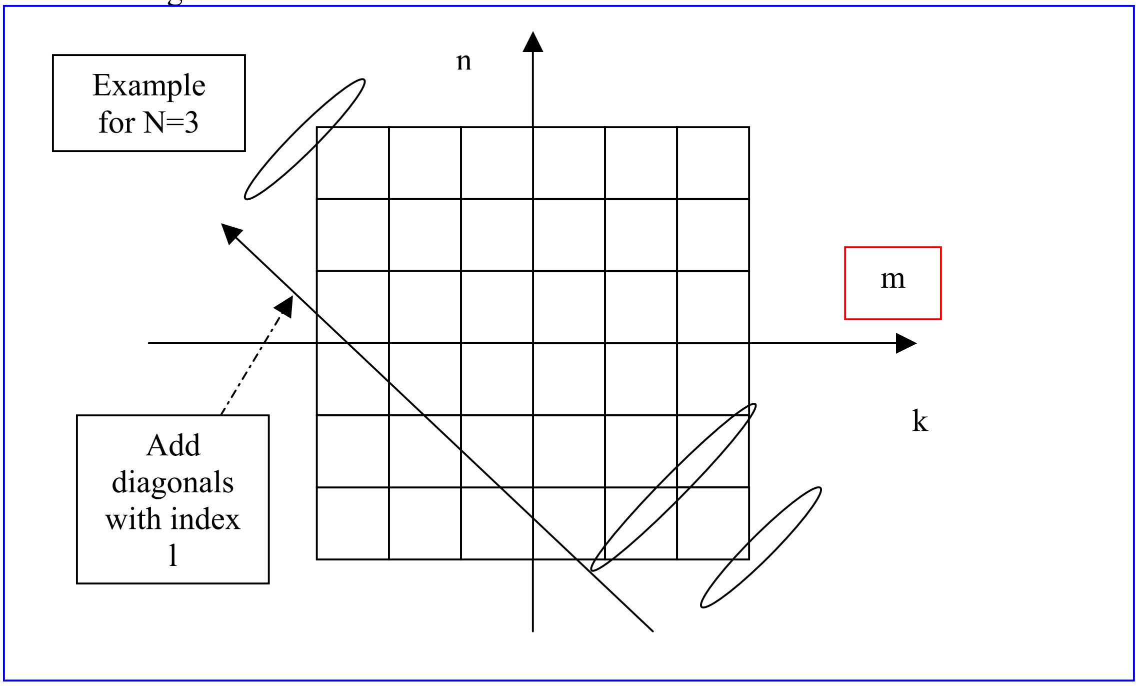
Note that this r.p. is also ciclo-stationary since $r_x(t,\tau) = r_x(t+T,\tau)$
And computing the first term of its Fourier series

$$r_x(\tau) \triangleright \lim_{N\to\infty} \frac{1}{(2N+1)T} . \int_{-NT}^{NT} r_x(t,\tau).dt$$

since N large and full-response pulses

$$\int_{-NT}^{NT} p(t-mT).p(t+\tau-(l+m)T).dt \triangleright r_{pp}(\tau-lT)$$ i.e. the acf for finite energy signals

and the change of variables is:



which means that for a given l there are $\left(2N+1-|l|\right)$ terms. In consequence:

$$r_x(\tau) = \lim_{N\to\infty} . \sum_{l=-N}^{N} \frac{(2N+1-|l|)}{(2N+1)T} r_a(l).r_{pp}(\tau-lT)$$

In summary:

$$r_x(\tau) = r. \sum_{l=-\infty}^{\infty} r_a(l).r_{pp}(\tau-lT)$$

Note that the acf defined before is just the convolution of the symbols' acf with the acf of the signaling pulse

In fact, the power spectral density is the product of the energy spectrum of the pulse by the FT of the acf of the information symbols sequence.

$$S_x(w) = r.S_{pp}(w).\sum_{q=-\infty}^{\infty} r_a(q).\exp(-jqw/r)$$

II.13.11.-

See II.13.13

II.13.12.-

$$\underline{X}_n = a.\underline{S} + \underline{w}_n$$
$$\Lambda\left(a,\underline{S},\sigma^2\right) = K_0 - Q.Ln\left(\sigma^2\right) - \frac{1}{\sigma^2}.\left[\left(\underline{X}_n - a.\underline{S}\right)^H.\left(\underline{X}_n - a.\underline{S}\right)\right]$$

a)

$$a_{ML} = \max_a \Lambda \,;\, \frac{\partial\Lambda}{\partial a^H} = 0 \,;\, 0 = \underline{S}^H.\left(\underline{X}_n - a.\underline{S}\right) \quad \text{and} \quad a_{ML} = \frac{\underline{S}^H.\underline{X}_n}{\underline{S}^H.\underline{S}} = \frac{1}{Q}\underline{S}^H.\underline{X}_n$$

--

$$\Lambda\left(\underline{S},\sigma^2\right) = K_0 - Q.Ln\left(\sigma^2\right) - \frac{1}{\sigma^2}\left[\underline{X}_n^H.\left(\underline{\underline{I}} - \frac{\underline{S}.\underline{S}^H}{Q}\right)^2.\underline{X}_n\right]$$

--

b)

since $\underline{\underline{P}} = \left(\underline{\underline{I}} - \frac{\underline{S}.\underline{S}^H}{Q}\right)$ and $\underline{\underline{P}} = \underline{\underline{P}}^H$ together with $\underline{\underline{P}}.\underline{\underline{P}}^H = \underline{\underline{P}}$ allows writing down the likelihood in a more compact form.

$$\Lambda\left(\underline{S},\sigma^2\right) = K_0 - Q.Ln\left(\sigma^2\right) - \frac{1}{\sigma^2}\left[\underline{X}_n^H.\underline{\underline{P}}.\underline{X}_n\right]$$

Now, because $Trace\left(\underline{\underline{u}}.\underline{\underline{v}}^H\right) = Trace\left(\underline{\underline{v}}^H.\underline{\underline{u}}\right)$ then

$$\underline{X}_n^H.\underline{\underline{P}}.\underline{X}_n = Trace\left[\underline{X}_n^H.\underline{\underline{P}}.\underline{X}_n\right] = Trace\left[\underline{\underline{P}}.\underline{X}_n.\underline{X}_n^H\right]$$

and

$$\Lambda\left(\underline{S},\sigma^2\right) = K_0 - Q.Ln\left(\sigma^2\right) - \frac{1}{\sigma^2}Trace\left[\underline{\underline{P}}.\underline{X}_n.\underline{X}_n^H\right]$$

c)

$$\Lambda_N\left(\underline{S},\sigma^2\right)=\left|\begin{matrix}Independent\\ sample\ vector\end{matrix}\right| Ln\left\{\prod_{n=0}^{N-1}\Pr\left(\underline{X}_n / \underline{S},\sigma^2\right)\right\}=\sum_{n=0}^{N-1}Ln\left\{\Pr\left(\underline{X}_n / \underline{S},\sigma^2\right)\right\}=\sum_{n=0}^{N-1}Ln\left\{\Lambda_n\right\}$$

and the log-likelihood results in

$$\Lambda_N\left(\underline{S},\sigma^2\right)=K_1-N.Q.Ln(\sigma^2)-\frac{N}{\sigma^2}.\sum_{n=0}^{N-1}Trace\left(\underline{\underline{P}}.\underline{X}_n.\underline{X}_n^H\right)=$$

$$=K_1-N.Q.Ln(\sigma^2)-\frac{N}{\sigma^2}.Trace\left(\underline{\underline{P}}.\frac{1}{N}\sum_{n=0}^{N-1}\underline{X}_n.\underline{X}_n^H\right)=$$

$$=K_1-N.Q.Ln(\sigma^2)-\frac{N}{\sigma^2}.Trace\left(\underline{\underline{P}}.\underline{\underline{R}}\right)$$

where $\underline{\underline{R}}$ is the estimated auto covariance matrix.

d)

$$E\left[\underline{X}_n.\underline{X}_n^H\right]=|a|^2.\underline{S}.\underline{S}^H+\sigma^2.\underline{\underline{I}}$$

e)

$$Trace\left(\underline{\underline{P}}.\underline{\underline{R}}\right)=Trace\left(\underline{\underline{R}}-\underline{S}.\underline{S}^H.\underline{\underline{R}}/Q\right)=Trace\left(\underline{\underline{R}}\right)-Trace\left(\frac{\underline{S}^H.\underline{\underline{R}}.\underline{S}}{Q}\right)=$$

$$=Trace\left(\underline{\underline{R}}\right)-\frac{1}{Q}Trace\left(\underline{S}^H.\underline{\underline{R}}.\underline{S}\right)=$$

$$=Trace\left(\underline{\underline{R}}\right)-\frac{1}{Q}\left(\underline{S}^H.\underline{\underline{R}}.\underline{S}\right)$$

f)

$$\sigma_{ML}^2=\max_{\sigma^2}\Lambda_N(\underline{S},\sigma^2)\quad;\quad\frac{d\Lambda_N}{d\sigma^2}=0$$

$$0=-\frac{N.Q}{\sigma^2}+\frac{N}{\sigma^4}.\left(Trace\left(\underline{\underline{R}}\right)-\frac{1}{Q}.\underline{S}^H.\underline{\underline{R}}.\underline{S}\right)$$

and the ML estimate of the noise power is $\sigma_{ML}^2=\frac{1}{Q}.\left(Trace\left(\underline{\underline{R}}\right)-\frac{1}{Q}.\underline{S}^H.\underline{\underline{R}}.\underline{S}\right)$

g)Using the acf of section (d) the new expression for the expected value of the noise covariance follows

$$E\left[\sigma_{ML}^2\right]=\frac{1}{Q}\left[\left(|a|^2Q+Q\sigma^2\right)-\left(\frac{1}{Q}\left\{|a|^2Q^2+\sigma^2Q\right\}\right)\right]$$

Trace($\underline{\underline{R}}$)

And

$$E\left[\sigma_{ML}^2\right]=\sigma^2\frac{Q-1}{Q}$$

h)

Using the estimates of the complex envelope and the noise power, the likelihood reduces to

----

$$\Lambda\left(\underline{S}\right) = K_1 - N.Q.\left[Ln\left(\sigma_{ML}^2\right) - 1\right]$$ and

$$\underline{S}_{ML} = \max \Lambda\left(\underline{S}\right) = \min\left[Ln\left(\sigma_{ML}^2\right)\right] = \min\left[Trace(\underline{\underline{R}}) - ...\right] =$$

$$\max\left[\frac{1}{Q}.\left(\underline{S}^H.\underline{\underline{R}}.\underline{S}\right)\right]$$ which is the maximum of the Periodogram (Welch procedure) of the available data.

II.13.13.-

a) $E\left[\hat{P}_x\right] = \frac{1}{2N-1}\left(2N-1\right).P_x = P_x$ unbiased

b) $\text{var}^2\left(\hat{P}_x\right) = E\left[\hat{P}_x^2\right] - \left(E\left[\hat{P}_x\right]\right)^2$ the second term is the previous section and the second term is:

$$E\left[\hat{P}_x^2\right] = E\left\{\frac{1}{(2N-1)^2}\sum_n\sum_m x^2(n).x^2(m)\right\} = \frac{1}{(2N-1)^2}\sum_n\sum_m\left(r_x^2(0) + 2.r_x^2(n-m)\right) =$$

$$= P_x^2 + \frac{1}{(2N-1)^2}\sum_n\sum_m\left(2.r_x^2(n-m)\right)$$

in consequence

$$\text{var}^2\left(\hat{P}_x\right) = \frac{1}{(2N-1)^2}\sum_n\sum_m\left(2.r_x^2(n-m)\right) = \begin{vmatrix} n-m=q \\ m=p \end{vmatrix} = \frac{1}{(2N-1)^2}\sum_{q=-N+1}^{N-1} r_x^2(q).\left(\left(2N-1\right) - |q|\right) =$$

$$= \frac{1}{(2N-1)}.\sum_{q=-N+1}^{N-1} r_x^2(q).\left(1 - \frac{|q|}{(2N-1)}\right)$$

c)

When the number of available data N tends to infinity

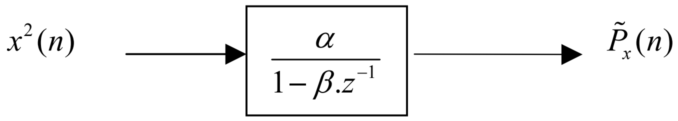
$$\text{var}^2\left(\hat{P}_x\right) = \frac{1}{(2N-1)}.\sum_{q=-N+1}^{N-1} r_x^2(q) = |Parseval| = \frac{1}{(2N-1)}.\left\{\frac{1}{T}\int_{-B}^{B} S_x^2(f).df\right\}$$

Now assuming that the spectral density is flat in the bandwidth [-B,B]

i.e. $S_x(f) = \frac{P_x}{2B}.\prod\left(\frac{f}{2B}\right)$ then $\text{var}^2(\hat{P}_x) \Rightarrow \frac{P_x^2}{2B.(2N-1)T}$

Note that the denominator is the product of the signal time duration by the signal frequency duration. This product is now as the degrees of freedom in the estimation procedure and they reduce always (with the proper estimation procedure) the variance of the estimate. A brief explanation for this fact follows: Since the random process has a bandwidth of 2B, the coherence time, i.e. the time we have to wait to label independent records is equal to the inverse (approx.) 1/2B sec. And, in consequence, the number of independent records, which dictates the possible variance reduction, is given by the quotient between the time support of the data and the coherence time. This is just the denominator of the asymptotic expression of the variance depicted above.

e)



and $E\left[\tilde{P}_x(n)\right]=\beta.E\left[\tilde{P}_x(n-1)\right]+\alpha.E\left[x^2(n)\right]$; assuming that the r.p. is stationary

$E\left[\tilde{P}_x(n)\right]=\beta.E\left[\tilde{P}_x(n)\right]+\alpha.P_x$; and $E\left[\tilde{P}_x(n)\right]=\dfrac{\alpha}{1-\beta}P_x$ from which is easy to deduce the adequate choice of the filter parameters to obtain an unbiased estimate.

f) when $\breve{P}_x(n)=\dfrac{1}{N}.\sum_{m=n-N+1}^{n}x^2(n)$ then $\breve{P}_x(n)=\dfrac{N-1}{N}.\breve{P}_x(n-1)+\dfrac{x^2(n)}{N}-\dfrac{x^2(n-N)}{N-1}$

thus, the estimate is similar to the previous one when $\alpha=\dfrac{1}{N}$ and $\beta=1-\dfrac{1}{N}$. The only difference is the last term which can be considered of low impact in the estimate the power at instant n mainly for N large. Also, this term does not longer exist for n<N and x(n) is cero for negative arguments.

II.13.14.-

It is a pdf since $\dfrac{1}{\pi}\int_{-\infty}^{\infty}\dfrac{1}{1+a^2}.da=\dfrac{1}{\pi}.\arg tg(a)\Big|_{-\infty}^{\infty}=1$, the mean is cero, but the variance

$\dfrac{1}{\pi}\int_{-\infty}^{\infty}\dfrac{a^2}{1+a^2}.da=\dfrac{1}{\pi}(1-\arg tg(a))\Big|_{-\infty}^{\infty}=\infty$!!is infinite!!

Nevertheless, the Cauchy distribution is preserved when adding Cauchy r.v. since it is invariant to the convolution operator.

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{1+a^2}.\frac{1}{1+\left(b-a\right)^2}.da\propto\frac{1}{\pi}.\frac{1}{1+b^2}$$

The characteristic function is $\dfrac{1}{\pi}\int\dfrac{\cos(wa)}{1+a^2}da=\Phi(w)$ to solve this integral, note that $\Phi(0)=1$ and $\dfrac{\partial^2\Phi}{\partial\omega^2}=-\Phi(w)$ being $\Phi(w)=e^{-|w|}$ the solution

II.13.15.-

Done within the corresponding section of this chapter.

II.13.16.-

Sentences are correct when applied to a finite record length but no to describe global features of the random process to which input and output records belong. Note that there is implicit the assumption of circular convolution in sentences that does not longer apply for the convolution of the input with the impulse response. In fact, it is easy to prove that the periodogram for a non white random process is always biased. This bias is introduced by the convolution of the actual spectral density with the lag-window transform. This lag-window is implicit in all the sentences, from (a) to (d).

II.13.17.-

$$\underline{z}=\begin{bmatrix} x \\ y \end{bmatrix} \quad E(\underline{z})=\begin{bmatrix} m_x \\ m_y \end{bmatrix} \quad \underline{\underline{R}}_z=\begin{pmatrix} r_{xx} & r_{xy} \\ r_{yx} & r_{yy} \end{pmatrix} \quad \underline{\underline{R}}_z^{-1}=\underline{\underline{G}}.\underline{\underline{D}}.\underline{\underline{G}}^H$$

a)

$$\Pr(x,y)=\Pr(\underline{z})=\frac{K_o}{\det(\underline{\underline{R}}_z)}.e^{-(\underline{z}-\underline{m}_z)^H.\underline{\underline{R}}_z^{-1}.(\underline{z}-\underline{m}_z)}$$

We are going to use the following diagonalization of the acf matrix:

$$\underline{\underline{R}}_z^{-1}=\begin{pmatrix} 1 & 0 \\ -r_{yy}^{-1}.r_{yx} & 1 \end{pmatrix}.\begin{pmatrix} \left(r_{xx}-r_{xy}.r_{yy}^{-1}.r_{yx}\right)^{-1} & 0 \\ 0 & r_{yy}^{-1} \end{pmatrix}.\begin{pmatrix} 1 & -r_{yy}^{-1}.r_{xy} \\ 0 & 1 \end{pmatrix}$$

b)

$$\begin{bmatrix} x1 \\ x2 \end{bmatrix}=\begin{pmatrix} 1 & -r_{yy}^{-1}.r_{xy} \\ 0 & 1 \end{pmatrix}.\begin{bmatrix} x-m_x \\ y-m_y \end{bmatrix}$$

$$x1=(x-m_x)-r_{yy}^{-1}.r_{xy}.(y-m_y)$$

$$x2=y-m_y$$

c)

$$(\underline{z}-\underline{m}_z)^H.\underline{\underline{R}}_z^{-1}.(\underline{z}-\underline{m}_z)=[x1 \quad x2]^H.\begin{pmatrix} \left(r_{xx}-r_{xy}.r_{yy}^{-1}.r_{yx}\right)^{-1} & 0 \\ 0 & r_{yy}^{-1} \end{pmatrix}.\begin{bmatrix} x1 \\ x2 \end{bmatrix}=$$

$$=(x1)^2.\left(r_{xx}-r_{xy}.r_{yy}^{-1}.r_{yx}\right)^{-1}+(x2)^2.r_{yy}^{-1}$$

d)

$$\Pr\left(x/y\right)=K_2.\exp\left(-(x1)^2.\left(r_{xx}-r_{xy}.r_{yy}^{-1}.r_{yx}\right)^{-1}\right)$$

Gaussian and it can be formulated as

$$\left[(\underline{x}-\underline{m}_x)-\underline{\underline{r}}_{yy}^{-1}.\underline{\underline{r}}_{xy}.(\underline{y}-\underline{m}_y)\right]^H.\left(r_{xx}-r_{xy}.r_{yy}^{-1}.r_{yx}\right)^{-1}.\left[(\underline{x}-\underline{m}_x)-\underline{\underline{r}}_{yy}^{-1}.\underline{\underline{r}}_{xy}.(\underline{y}-\underline{m}_y)\right]$$

Note that, in a more general approach, in the last formula both x and y have been considered vectors instead mere scalars. In fact the previous presentation is still valid in all the respects for this general case.

In summary, the conditional mean of x is given by:

$$\underline{\hat{x}} = \underline{m}_x + \underline{\underline{r}}_{yy}^{-1}.\underline{\underline{r}}_{xy}.\left(\underline{y} - \underline{m}_y\right)$$

This provides the optimum conditional mean estimate of a vector x from a data vector y and the corresponding means and auto and cross correlation matrices. This is also known as the Wiener filtering method.

e) The variance is also derived from the conditional distribution

$$\underline{\underline{C}}_{\hat{x}\hat{x}} = \left(\underline{\underline{r}}_{xx} - \underline{\underline{r}}_{xy}.\underline{\underline{r}}_{yy}^{-1}.\underline{\underline{r}}_{yx}\right)^{-1}$$

f)

$$ECM \equiv \int \left(x - \hat{x}\right)^2 .\Pr\left(x/y\right).dx =$$

$$= \int x^2 .\Pr\left(x/y\right).dx + \hat{x}^2 - 2.\int \left(x - \hat{x}\right).\Pr\left(x/y\right).dx$$

and to find the minimum ECM

$$\frac{\partial ECM}{\partial \hat{x}} = 2.\hat{x} - 2.\int x.\Pr\left(x/y\right).dx = 0$$

$$\hat{x}_{\substack{\min \\ ECM}} = \int x.\Pr\left(x/y\right).dx = E_x\left[\Pr\left(x/y\right)\right]$$

g)

Already done previously.

II.13.18.-

$$H(z) = 1 - b(1).z^{-1} \qquad r_x(m) = \begin{cases} 1 + b(1)^2 & m = 0 \\ -b(1) & m = \pm 1 \\ 0 & else \end{cases}$$

then

$S_x\left(z\right) = -z.b(1) + \left(1 + b(1)^2\right) - b(1).z^{-1} = \left(1 - b(1).z^{-1}\right).\left(1 - b(1).z\right)$ where the first term is minimum phase

The procedure will be the following:

- Given x(n) for n=0,N-1
- Estimate the acf function using the unbiased estimate when N>>1
- Use r(0) and r(1) form the polynomial and find out two roots homothetic, with respect the unit circle
- b(1) is the root inside the unit circle.

II.13.19.-

Descriptive of the role of non linear systems for linearizing power amplifiers.

II.13.20.-

$$\underline{x}_n = \begin{bmatrix} x(n) \\ x(n-1) \\ x(n-2) \end{bmatrix} \qquad \underline{x}_n = \left(a.e^{j.\theta}\right).e^{j.w_0.t}.\begin{pmatrix} 1 \\ e^{jw_0} \\ e^{j2w_0} \end{pmatrix} + \underline{w}_n \qquad \underline{\underline{C}} = E\left[\underline{w}_n.\underline{w}_n^H\right] = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix}$$

a) with $b = a.e^{j\theta}$,i.e. the low pass complex envelope,

$$\Pr\left(b/\underline{x}_n\right) = K_0.\Pr\left(\underline{x}_n/b\right).\Pr(b)$$

$$\Pr\left(\underline{x}_n/b\right) = K_1.\exp\left[-\left(\underline{x}_n - b.\underline{S}\right)^H.\underline{\underline{C}}^{-1}.\left(\underline{x}_n - b.\underline{S}\right)\right]$$

$$\Pr(b) = K_2.\exp\left[-\left|b-m_b\right|^2/\sigma_b^2\right]$$

b) Taking logarithms and removing constant irrelevant for the minimization procedure

$$\Lambda(b) = -\left(\underline{x}_n - b.\underline{S}\right)^H.\underline{\underline{C}}^{-1}.\left(\underline{x}_n - b.\underline{S}\right) - \frac{\left|b-m_b\right|^2}{\sigma_b^2}$$

$$b_{MAP} = \min_b \Lambda(b)$$

$$\frac{\partial\Lambda}{\partial b} = 0 = \underline{S}^H.\underline{\underline{C}}^{-1}.\left(\underline{x}_n - b.\underline{S}\right) - \frac{\left(b-m_b\right)}{\sigma_b^2}$$

$$b_{MAP} = \frac{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{x}_n + \frac{m_b}{\sigma_b^2}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S} + \frac{1}{\sigma_b^2}}$$

c)

$$\Lambda_{ML}(b) = -\left(\underline{x}_n - b.\underline{S}\right)^H.\underline{\underline{C}}^{-1}.\left(\underline{x}_n - b.\underline{S}\right)$$

$$b_{ML} = \min_b \Lambda_{ML}(b)$$

$$\frac{\partial\Lambda}{\partial b} = 0 = \underline{S}^H.\underline{\underline{C}}^{-1}.\left(\underline{x}_n - b.\underline{S}\right)$$

$$b_{MAP} = \frac{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{x}_n}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}} \qquad \begin{matrix} \sigma_b \to \infty & b_{MAP} \Rightarrow b_{ML} \\ \sigma_b \to 0 & b_{MAP} \Rightarrow m_b \end{matrix}$$

d)

$$E\left[b_{ML}\right] = \frac{\underline{S}^H.\underline{\underline{C}}^{-1}.E\left(\underline{x}_n\right)}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}} = \frac{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}.b}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}} = b \qquad \text{unbiased estimate}$$

$$\mathrm{var}^2\left(b_{ML}\right)=E\left[\left(b-b_{ML}\right).\left(b-b_{ML}\right)\right]=E\left[\left(b.\frac{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}-b_{ML}\right).\left(b-b_{ML}\right)\right]=$$

$$=\left[\left(b.\frac{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}-E\left[b_{ML}\right]\right).\left(b-E\left[\frac{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{x}_n}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}\right]\right)\right]=$$

$$=\frac{\underline{S}^H.\underline{\underline{C}}^{-1}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}.E\left[\left(\underline{x}_n-b.\underline{S}\right).\left(\underline{x}_n-b.\underline{S}\right)^H\right].\frac{\underline{\underline{C}}^{-1}.\underline{S}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}=$$

$$=\frac{\underline{S}^H.\underline{\underline{C}}^{-1}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}.E\left[\underline{w}_n.\underline{w}_n^H\right].\frac{\underline{\underline{C}}^{-1}.\underline{S}}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}=\frac{1}{\underline{S}^H.\underline{\underline{C}}^{-1}.\underline{S}}$$

It is easy to check that it coincides with the Cramer-Rao bound.

d) When the noise is white and the number of available samples is N

$$\mathrm{var}^2\left(b_{ML}\right)=\frac{1}{\left(\sigma^2\right)^{-1}.\underline{S}^H.\underline{S}}=\frac{\sigma^2}{N}\Rightarrow 0 \quad when \quad N\Rightarrow\infty$$

If there are two lines then

$$\underline{x}_n=b1.\underline{S}_1+b2.\underline{S}_2+\underline{w}_n=\left(\underline{S}_1 \quad \underline{S}_2\right).\begin{bmatrix} b1 \\ b2 \end{bmatrix}+\underline{w}_n=\underline{\underline{S}}.\underline{b}+\underline{w}_n$$

$$\Lambda\left(\underline{b}\right)=\left(\underline{x}_n-\underline{\underline{S}}.\underline{b}\right)^H.\underline{\underline{C}}^{-1}.\left(\underline{x}_n-\underline{\underline{S}}.\underline{b}\right) \quad ; \quad \underline{b}_{ML}=\left(\underline{\underline{S}}^H.\underline{\underline{C}}^{-1}.\underline{\underline{S}}\right)^{-1}.\underline{\underline{S}}^H.\underline{\underline{C}}^{-1}.\underline{x}_n$$

Given the two frequencies the complex envelopes are estimated accordingly with the last expression.

When both frequencies are unknown, the use of any spectral estimation procedure to locate their positions is not longer optimum. In fact the ML estimate of the locations of both frequencies is formulated as follows:

$$\underline{\underline{S}}_{ML}=\min_{\underline{S}_1,\underline{S}_2}\Lambda\left(\underline{\underline{S}}\right)$$

$$\Lambda\left(\underline{\underline{S}}\right)=\underline{x}_n^H.\left[\underline{\underline{I}}-\underline{\underline{C}}^{-1}.\underline{\underline{S}}.\underline{\underline{A}}.\underline{\underline{S}}^H\right].\underline{\underline{C}}^{-1}.\left[\underline{\underline{I}}-\underline{\underline{S}}.\underline{\underline{A}}.\underline{\underline{S}}^H.\underline{\underline{C}}^{-1}\right].\underline{x}_n$$

$$where \quad \underline{\underline{A}}=\left[\underline{\underline{S}}.\underline{\underline{C}}^{-1}.\underline{\underline{S}}^H\right]^{-1}$$

When N vectors of independent data are available

$$\Lambda\left(\underline{\underline{S}}\right)=Trace\left[\left[\underline{\underline{I}}-\underline{\underline{C}}^{-1}.\underline{\underline{S}}.\underline{\underline{A}}.\underline{\underline{S}}^H\right].\underline{\underline{C}}^{-1}.\left[\underline{\underline{I}}-\underline{\underline{S}}.\underline{\underline{A}}.\underline{\underline{S}}^H.\underline{\underline{C}}^{-1}\right].\underline{\underline{R}}\right]$$

$$where \quad \underline{\underline{R}}=\frac{1}{N}\sum_{n=0}^{N-1}\underline{x}_n.\underline{x}_n^H$$

In addition when the noise is white

$$\underline{\underline{S}}_{ML} = \min_{\underline{S}_1,\underline{S}_2} \left\{ Trace\left(\left[\underline{\underline{I}} - \underline{\underline{S}}.\left(\underline{\underline{S}}^H.\underline{\underline{S}} \right)^{-1}.\underline{\underline{S}}^H \right].\underline{\underline{R}} \right) \right\}$$

The expression looks easy but it entails a great complexity for searching the two frequencies that minimize the trace. (See course notes of Arrays for a detailed description of the problem and sub-optimum solutions).

II.13.21.-

a) $\mu_{ML} = \dfrac{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{x}_n}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}} \qquad \mathrm{var}^2\left(\mu_{ML}\right) = \dfrac{1}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}}$

b) $\tilde{\mu} = \dfrac{\underline{1}^H.\underline{x}_n}{\underline{1}^H.\underline{1}} \qquad \mathrm{var}^2\left(\tilde{\mu}\right) = \dfrac{\underline{1}^H.\underline{\underline{R}}.\underline{1}}{\underline{1}^H.\underline{1}}$

c)

$$0 \le \frac{1}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}} < \frac{1}{\lambda_{\max}^{-1}.\left(\underline{1}^H.\underline{1}\right)} = \frac{\lambda_{\max}}{Q} \quad \textit{when } Q \to \infty \textit{ tends to cero}$$

$$0 \le \frac{\underline{1}^H.\underline{\underline{R}}.\underline{1}}{\left(\underline{1}^H.\underline{1}\right)^2} < \frac{\lambda_{\max}}{\left(\underline{1}^H.\underline{1}\right)} = \frac{\lambda_{\max}}{Q} \quad \textit{when } Q \to \infty \textit{ tends to cero}$$

d)

$$\left(\underline{1}^H.\underline{\underline{R}}.\underline{1}\right).\left(\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}\right) \ge \left|\underline{1}^H.\underline{1}\right|^2 = Q^2$$

use $\begin{array}{l} u = \underline{\underline{R}}^{1/2}.\underline{1} \\ v = \underline{\underline{R}}^{-1/2}.\underline{1} \end{array}$

and

$$\frac{1}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}} \le \frac{\underline{1}^H.\underline{\underline{R}}.\underline{1}}{\left(\underline{1}^H.\underline{1}\right)^2}$$

they are equal only for u proportional to v, i.e. when the samples belong to a white random process.

e)

done

II.13.22.-

a.- $x(n) = -a(1).x(n-1) + v(n)$ multiply both sides by x(n+m) and taking expectations we get $r_{xx}(m) = -a(1).r_{xx}(m-1) + \sigma_v^2.\delta(m) \quad \forall m \ge 0$. Thus for any m greater than cero $a(1) = -\dfrac{r_{xx}(m)}{r_{xx}(m-1)}$

b.- Since {x} and{w} are statistically independents $r_{yy}(m) = r_{xx}(m) + \sigma_w^2.\delta(m)$ Using the taps involved in the proposed estimate,

$$r_{yy}(0)=r_{xx}(0)+\sigma_w^2 \quad ; \quad r_{yy}(1)=r_{xx}(1) \text{ and } \hat{a}(1)=-\frac{r_{yy}(1)}{r_{yy}(0)}=\frac{-\frac{r_{xx}(1)}{r_{xx}(0)}}{1+\frac{\sigma_w^2}{r_{xx}(0)}}=\frac{a(1).SNR}{1+SNR}$$

c.- $s_{yy}(w)=s_{xx}(w)+\sigma_w^2$ (Just use FT in the acf formula of section (b).

The model for {y} is AR plus a constant, in consequence, the resulting model for the obsevartion r.p. is ARMA (1,1).

d.-

$$x(n)=y(n)-w(n)$$

$$because \quad x(n)=-a(1).x(n-1)+v(n)$$

$$y(n)-w(n)=-a(1).(y(n-1)-w(n-1))+v(n)$$

$$or \quad y(n)=-a(1).y(n-1)+[w(n)+a(1).w(n-1)]+v(n)$$

clearly y(n-2) does not depends neither w(n) not w(n-1) resulting

$$r_{yy}(2)=-a(1).r_{yy}(1)----> a(1)=-\frac{r_{yy}(2)}{r_{yy}(1)}$$

$$r_{yy}(1)=r_{xx}(1)=-a(1).r_{xx}(0)$$

also

$$and \quad \sigma_w^2=r_{yy}(0)-r_{xx}(0)=r_{yy}(0)+\frac{r_{yy}(1)}{a(1)}=r_{yy}(0)-\frac{r_{yy}^2(1)}{r_{yy}(2)}$$

II.13.23.

a.- Just assigning the following vectors:

$\underline{1}=[1 \quad 1 \quad \ldots \quad 1]^H$ y $\underline{N}=[-N+1 \quad -N+2 \quad \ldots \quad N-2 \quad N-1]^H$

the rest is straight forward.

b.- Since $\underline{\underline{\Phi}}=[\underline{1} \quad \underline{N}]$ thus

$$\underline{\underline{\Phi}}^H.\underline{\underline{\Phi}}=\begin{bmatrix}\underline{1}^H\\ \underline{N}^H\end{bmatrix}.[\underline{1} \quad \underline{N}]=\begin{bmatrix}\underline{1}^H.\underline{1} & \underline{1}^H.\underline{N}\\ \underline{N}^H.\underline{1} & \underline{N}^H.\underline{N}\end{bmatrix}=\begin{bmatrix}2N-1 & 0\\ 0 & S\end{bmatrix}$$

from which the inverse can be found.

c.- The likelihood is $Ln(\Lambda)=cte.-\frac{|\underline{X}-\underline{\underline{\Phi}}.\underline{a}|^2}{\sigma^2}$ since the noise is white. Taking derivative with respect $\underline{a}^H$ and setting to zero:

$\underline{\underline{\Phi}}^H.(\underline{X}-\underline{\underline{\Phi}}.\underline{a})=\underline{0}$ o bien $\underline{a}^{ML}=(\underline{\underline{\Phi}}^H.\underline{\underline{\Phi}})^{-1}.\underline{\underline{\Phi}}^H.\underline{X}$

The expected value is: $E(\underline{a}^{ML})=(\underline{\underline{\Phi}}^H.\underline{\underline{\Phi}})^{-1}.\underline{\underline{\Phi}}^H.E(\underline{X})=(\underline{\underline{\Phi}}^H.\underline{\underline{\Phi}})^{-1}.\underline{\underline{\Phi}}^H.[\underline{\underline{\Phi}}.\underline{a}]=\underline{a}$ Which proves that the estimate is unbiased.

d.- The form of the estimate is:

$$\begin{bmatrix}A^{ML}\\ B^{ML}\end{bmatrix}=\begin{bmatrix}1/(2N-1) & 0\\ 0 & 1/S\end{bmatrix}.\begin{bmatrix}\underline{1}^H\\ \underline{N}^H\end{bmatrix}.\underline{X}=\begin{bmatrix}\underline{1}^H/(2N-1)\\ \underline{N}^H/S\end{bmatrix}.\underline{X}$$

Thus the estimates are:

$$A^{ML}=\frac{\underline{1}^H.\underline{X}}{(2N-1)} \quad y \quad B^{ML}=\frac{\underline{N}^H.\underline{X}}{S}$$

e.- The variance will be:

$$E\left[\left(\underline{a}^{ML}-\underline{a}\right).\left(\underline{a}^{ML}-\underline{a}\right)^{H}\right]=E\left[\left\{\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}.\underline{\underline{\Phi}}^{H}.\underline{X}-\underline{a}\right\}.\left\{\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}.\underline{\underline{\Phi}}^{H}.\underline{X}-\underline{a}\right\}^{H}\right]=$$

$$\left|\begin{array}{l}\text{using}\\ \underline{a}=\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}.\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}.\underline{a}\end{array}\right|=E\left[\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}\underline{\underline{\Phi}}^{H}\left\{\underline{X}-\underline{\underline{\Phi}}\underline{a}\right\}.\left\{.......\right\}^{H}\right]=\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}\underline{\underline{\Phi}}^{H}\underline{\underline{\Phi}}\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}.\sigma^{2}=$$

$$=\left(\underline{\underline{\Phi}}^{H}.\underline{\underline{\Phi}}\right)^{-1}.\sigma^{2}=\sigma^{2}.\begin{bmatrix}1/(2N-1) & 0\\ 0 & 1/S\end{bmatrix}$$

Where both terms tend to zero when the length tends to infinity.

III.10.1.-

$$S_{xx}(z) = \frac{\sigma_0^2}{|A(z)|^2} + \sigma^2 = \frac{\sigma_0^2 + \sigma^2 |A(z)|^2}{|A(z)|^2}$$

$$S_{xx}(z).A(z) = \frac{\sigma_0^2 + \sigma^2 |A(z)|^2}{A(z^{-1})}$$

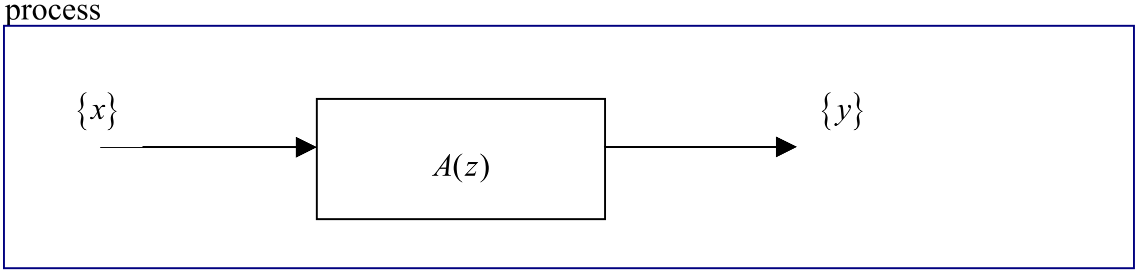
since the right hand term is cero for those samples above Q, the inverse transform of the length hand term verifies:

$$r_x(m) * a(m) = 0 \quad \forall m \geq Q+1$$

$$r_x(m) + \sum_{q=1}^{Q} a(q).r_x(m-q) = 0 \quad m = Q+1, \infty$$

From these extended Y-W equations, given the acf of the random process, the coefficients of the denominator of the model can be obtained.

Filtering the original data with the coefficients obtained before, the output is a pure MA process



Since the spectral density is $S_{yy}(w) = \sigma_0^2 + \sigma^2.|A|^2$, the two parameters can be derived from the first two terms of the acf of the new data record as:

$$r_y(0) = \sigma_0^2 + \sigma^2.\sum_{q=0}^{Q} a^2(q)$$

$$r_y(1) = \sigma^2.\sum_{q=0}^{Q-1} a(q).a(q+1)$$

from which σ_0^2 and σ^2 can be obtained

III.10.2.-

From a bank filter approach to spectral estimation, the filter associated to given value of the Periodogram is $\underline{A} = \frac{S}{Q}$ this filter results from the following minimization problem:

$$\underline{A}^H.\underline{S} = 1 \quad \textit{steers the desired frequency } (0dB.\,response)$$

$$\underline{A}^H.\underline{A}\Big|_{MIN} \quad \textit{with the lowest response to white noise}$$

Moving the same arguments to the case when interference is present at $\underline{S}_i$ results in the following problem:

$\underline{A}^H.[\underline{S} \quad \underline{S}_i] = [1 \quad 0] \quad or \quad \underline{A}^H.\underline{\underline{S}} = \underline{1}^H$

$with \quad \underline{A}^H.\underline{A}\Big|_{MIN}$

the minimization is as follows:

$\Im = \underline{A}^H.\underline{A} - \left(\underline{A}^H.\underline{\underline{S}} - \underline{1}^H\right).\underline{\lambda}\,; \quad \frac{\partial \Im}{\partial \underline{A}^H} = \underline{0}$ taking this solution to the constraint equation

$\underline{A} - \underline{\underline{S}}.\underline{\lambda} = \underline{0} \quad or \quad \underline{A} = \underline{\underline{S}}.\underline{\lambda}$

$\underline{\lambda}^H.\left(\underline{\underline{S}}^H.\underline{\underline{S}}\right) = \underline{1}^H \quad or \quad \underline{\lambda} = \left(\underline{\underline{S}}^H.\underline{\underline{S}}\right)^{-1}.\underline{1}$

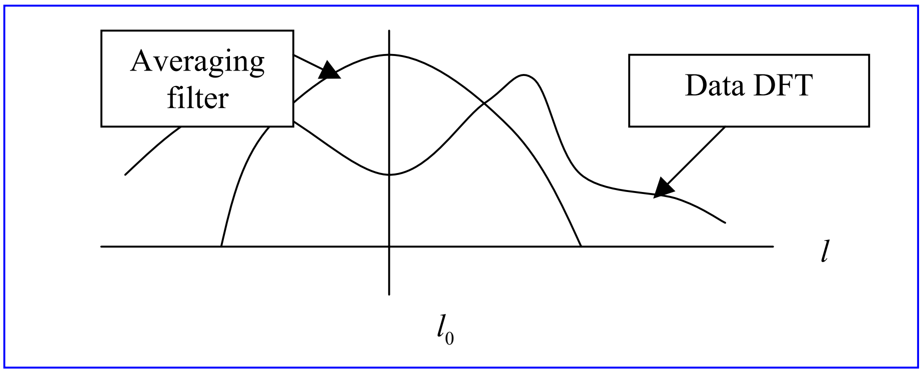
and since the spectral estimate is $\frac{\underline{A}^H.\underline{\underline{R}}.\underline{A}}{\underline{A}^H.\underline{A}}$

$\underline{A} = \underline{\underline{S}}.\left(\underline{\underline{S}}^H.\underline{\underline{S}}\right)^{-1}.\underline{1}$

then $\hat{S}_{xx}(w) = \dfrac{\underline{1}^H.\left(\underline{\underline{S}}^H.\underline{\underline{S}}\right)^{-1}.\underline{\underline{S}}^H.\underline{\underline{R}}.\underline{\underline{S}}.\left(\underline{\underline{S}}^H.\underline{\underline{S}}\right)^{-1}.\underline{1}}{\underline{1}^H.\left(\underline{\underline{S}}^H.\underline{\underline{S}}\right)^{-1}.\underline{1}} \quad \forall w \neq w_i$

III.10.3.-

For every frequency, denoted with index l_0, the estimate is and average of the surrounding DFT samples.



In consequence, the estimate is given by the convolution of the averaging function w(n), moved to frequency l_0 with the original sequence

$$\sum_{q=-Q}^{Q} w(q).x(n-q).\exp(-j2\pi l_0 n/N)$$

The procedure is equivalent to modulate the original data, in such a way that index l_0 moves to the cero frequency, and then a low pas filter W(l) is applied

MAJOR DRAWBACKS

-Averaging in the frequency domain does not take into account that the phase of the DFT may produce undesired cancellations. A better procedure is to average directly the Periodogram samples. This implies that the equivalence at the time domain will be the convolution of the sample autocorrelation with the inverse Fourier transform.
- The equivalence at the time domain is not strict since a finite support is used in the frequency domain. This implies that the time support of the time window may exceed the duration of the original record.

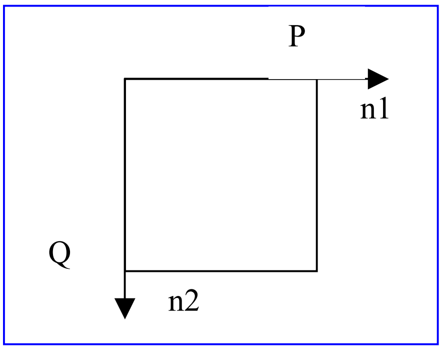
III.10.4.-

See the solution in A. Papoulis "The Fourier Integral and its applications" a pioneer work (first book) describing the role of Fourier transform in electrical engineering and communications.

III.10.5.-

The problem of 2-D spectral estimation is fully described in chapter V of these course notes. This section of chapter V is easy to read from the background obtained from this chapter. The exercise is solved for completeness here.

The filter is $a(n1,n2)$



Being (w,ψ) the frequencies corresponding to the 2D Fourier transform of indexes n1 and n2 respectively, the frequency response of the filter is:

$$A(w,\psi)=\sum_{n1=0}^{P}\sum_{n2=0}^{Q} a(n1,n2).\exp\left(-j\left[w.n1+\psi.n2\right]\right)$$

or $A(w,\psi)=\underline{a}^H.\underline{S}$ where

$$\underline{a}^H=\left[a(0,0),a(1,0),....,a(P,0),a(0,1),....,a(P,1),a(0,2),.....,a(P,Q)\right]$$

$$\underline{S}^H=\left[1,\exp(jw),..,\exp(jPw),\exp(j\xi),....,\exp(j\psi+jPw),\exp(j2\xi),..,\exp(jQ\xi+jPw)\right]$$

the filter output is

$$y(n,m)=\sum_{n1,n2} x(n-n1,m-n2).a(n1,n2)=\underline{a}^H.\underline{x}$$

$$with\quad \underline{x}^H=\left[x(n,m),x(n-1,m),..,x(n-P,m),x(n,m-1),..,x(n-P,m-1),....,x(n-P,m-Q)\right]$$

The filter design

$$\underline{a}^H.\underline{S}=1 \qquad \text{with} \quad \underline{\underline{C}}=E\left[\underline{x}_{n,m}.\underline{x}_{n,m}^H\right] \quad \textit{or its estimate} \quad \hat{\underline{\underline{C}}}=\frac{1}{N.M}\sum_{n,m}\underline{x}_{n,m}.\underline{x}_{n,m}^H$$

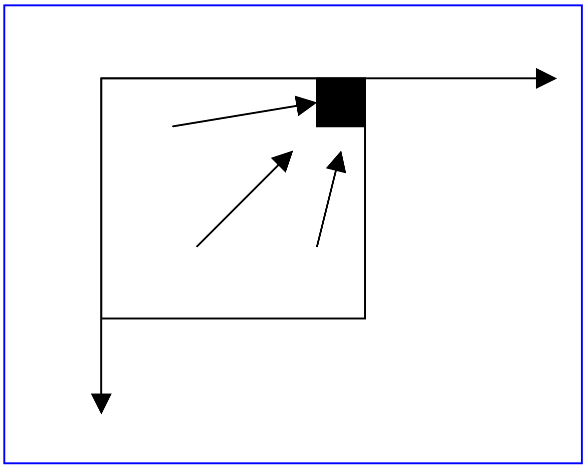
$$\underline{a}^H.\underline{\underline{C}}.\underline{a}\Big|_{MIN}$$

The power and the density estimates are:

$$P(w,\psi)=\underline{a}^H.\underline{\underline{C}}.\underline{a}=\frac{1}{\underline{a}^H.\underline{\underline{C}}^{-1}.\underline{a}} \quad \textit{and} \quad S_{xx}(w,\psi)=\frac{\underline{a}^H.\underline{\underline{C}}.\underline{a}}{\underline{a}^H.\underline{a}}=\frac{\underline{a}^H.\underline{\underline{C}}^{-1}.\underline{a}}{\underline{a}^H.\underline{\underline{C}}^{-2}.\underline{a}}$$

III.10.6.-

Given the 2D predictor $a(n1,n2)$, the first problem is to decide which sample is going to be predicted. At least four of them are possible. Note that the concept of causality, usual in 1D is lost.



Assuming that this corner is selected, the predictor coefficient for this lag will be forced to be one.

$\underline{a}^H.\underline{1}=1$ where $\underline{1}^H=\left[0,...0,1(\textit{position } P,1),0,..,0\right]$ Now to minimize the prediction error we need to minimize $\underline{a}^H.\underline{\underline{C}}.\underline{a}\Big|_{MIN}$

The solution is $\underline{a}=\dfrac{\underline{\underline{C}}^{-1}.\underline{1}}{\underline{1}^H.\underline{\underline{C}}^{-1}.\underline{1}}$ the prediction error is $\xi=\dfrac{1}{\underline{1}^H.\underline{\underline{C}}^{-1}.\underline{1}}$ and the spectral density estimate, i.e. the prediction error divided by the frequency response of the linear predictor is $\hat{S}_{xx}^{LP}(w,\psi)=\dfrac{\underline{1}^H.\underline{\underline{C}}^{-1}.\underline{1}}{\left|\underline{1}^H.\underline{\underline{C}}^{-1}.\underline{S}\right|^2}$

It is NOT possible to say that this estimate is a maximum entropy estimate as it was in 1D, since the acf support used for the filter computation exceeds the size of the predictor. To see this in a clearer manner, let us formulate the MEM estimate as it was in 1D

$$\iint Ln\left(S_{xx}^{MEM}(w,\psi)\right)dw.d\psi\Big|_{MAX}$$

subject to

$$\iint S_{xx}^{MEM}(w,\psi).\exp\left[j\left(w.n1+\psi.n2\right)\right].dw.d\psi = c\left(|n1|,|n2|\right) \quad \forall n1,n2 \le P,Q$$

assuming P and Q equal to 1. the number of values of matrix $\underline{\underline{C}}$ used to derive the MEM estimate will be 9 (i.e. (-1,-1),(-1,0),(-1,1),(0,-1),(0,0),(0,1),(1,-1),(1,0) and (1,1)).

For a linear predictor of 9 coefficients the matrix involved requires 37 acf values. Clearly the support or basis information is different.

MEM in 2D has to be computed iteratively (see references hereafter).

III.10.7.-

Being $\underline{a}_0,\underline{a}_1,....,\underline{a}_{Q-1}$ the set of vectors containing the optimum linear predictors for a given random process. Due to the different lengths (in increasing order), it is easy to check that $\underline{\underline{A}}=\left[\underline{a}_0,\underline{a}_1,...,\underline{a}_{Q-1}\right]$ is upper triangular. At the same time, due to the fact that the forward and the backward predictors for a stationary random process are equal, the mentioned matrix diagonalizes the acf matrix of the process.

$\underline{\underline{A}}^H.\underline{\underline{R}}.\underline{\underline{A}} = diag\left(\xi_0,\xi_1,..,\xi_{Q-1}\right)=\underline{\underline{\xi}}$ where the elements of the diagonal are the prediction error powers for the successive orders contained in matrix $\underline{\underline{A}}$. In summary:

$\underline{\underline{A}}^H.\underline{\underline{R}}.\underline{\underline{A}}=\underline{\underline{\xi}}$ and, in consequence $\underline{\underline{R}}=\underline{\underline{A}}^H.\underline{\underline{\xi}}.\underline{\underline{A}}$ and $\underline{\underline{R}}^{-1}=\underline{\underline{A}}^H.\underline{\underline{\xi}}^{-1}.\underline{\underline{A}}$.

Since matrix of the prediction errors is diagonal, this last expression can be further developed showing its explicit dependence on the linear predictors as:

$\underline{\underline{R}}^{-1}=\underline{\underline{A}}^H.\underline{\underline{\xi}}^{-1}.\underline{\underline{A}}=\sum_{q=0}^{Q-1}\frac{\underline{a}_q.\underline{a}_q^H}{\sigma_q^2}$ Multiplying both sides by $\underline{S}^H$. and by $.\underline{S}$ and taking the inverse we obtain

$$S_{xx(Q-1)}^{MLM}(w)=\frac{1}{\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}}=\frac{1}{\sum_{q=0}^{Q-1}\frac{1}{\sigma_q^2}\left|\underline{a}_q^H.\underline{S}\right|^2}=\frac{1}{\sum_{q=0}^{Q-1}\frac{1}{S_{xx(q)}^{MEM}(w)}} \qquad \text{or}$$

$\frac{1}{S_{xx(Q-1)}^{MLM}(w)}=\sum_{q=0}^{Q-1}\frac{1}{S_{xx(q)}^{MEM}(w)}$ In other words, The MLM estimate, being the "parallel union" of successive MEM estimates will have always poorer resolution than the corresponding MEM estimate for the same order or equal size of predictor MEM or filter MLM. THIS IS NOT LONGER TRUE FOR NMLM WHICH HAS SUPERIOR PERFORMANCE IN TERMS OF LOW SIDELOBE AND RESOLUTION THAN

MEM. Only when the process under analysis is actually and AR, the order of the predictor matches perfectly the model order and the number of data samples are above ten times the order MEM will be superior to NMLM.

III.10.8.-

$$\underline{\underline{R}} = \alpha.\underline{S}_0^H.\underline{S}_0 + \sigma^2.\underline{\underline{I}}$$

$\underline{\underline{R}}.\underline{S}_0 = \alpha.\left|\underline{S}_0\right|^2.\underline{S}_0 + \sigma^2.\underline{S}_0 = \left(\sigma^2 + \alpha.\left|\underline{S}_0\right|^2\right).\underline{S}_0$ thus $\dfrac{\underline{S}_0}{\left|\underline{S}_0\right|}$ is the eigenvector and

$\left(\sigma^2 + \alpha.\left|\underline{S}_0\right|^2\right)$ is the eigenvalue

III.10.9.-

$$\begin{bmatrix} r(m) & r(m-1) & \dots & r(m-Q+1) \\ r(m+1) & r(m) & \dots & r(m-Q) \\ \dots & \dots & \dots & \dots \\ r(m+P-1) & r(m+P-2) & \dots & r(m+P-Q) \end{bmatrix} . \begin{bmatrix} 1 \\ a(1) \\ \dots \\ a(Q-1) \end{bmatrix} = \underline{0} \quad \forall P \geq Q$$

This system of equations $\underline{\underline{R}}_e.\underline{a} = \underline{0}$ is over determined and has not a solution, unless the estimate is perfect and the order matches the model order. In general, there is no a solution for it.

One way out is to match the systems of equations in the MSE sense, i.e. find vector $\underline{a}$ such that $\left|\underline{\underline{R}}_e.\underline{a} - \underline{0}\right|^2$ is minimized, of course with the constraint that the first coefficient of the unknown vector is equal to one.

$\left.\underline{a}^H.\underline{\underline{R}}_e^H.\underline{\underline{R}}_e.\underline{a}\right|_{MIN}$, $\underline{a}^H.\underline{1} = 1$ the solution to this problem is $\underline{a} = \dfrac{\left(\underline{\underline{R}}_e^H.\underline{\underline{R}}_e\right)^{-1}.\underline{1}}{\underline{1}^H.\left(\underline{\underline{R}}_e^H.\underline{\underline{R}}_e\right)^{-1}.\underline{1}}$

III.10.10.-

See next exercise.

III.10.11.-

Since we are looking for two frequencies, two equations are enough to find them. These two equations reflect than the signal, without noise, are perfectly predictable since the signal is the solution of a deterministic differential equation. To set these two equations we use all the data available and, in order to do this, we have to select the forward and backward equations for the two border samples x(0) and x(N-1). This is correct since pure sinusoids present the same forward and backward evolution.

$$\begin{bmatrix} x(0) & x(1) & \ldots & x(N-3) & x(N-2) \\ x(N-1) & x(N-2) & \ldots & x(2) & x(1) \end{bmatrix} . \begin{bmatrix} a(0) \\ a(1) \\ \ldots \\ a(N-2) \end{bmatrix} = \begin{bmatrix} x(N-1) \\ x(0) \end{bmatrix}$$

in vector form $\underline{\underline{X}}.\underline{a} = \underline{x}$. Now, since the system is over determined, we take into account the presence of white noise by imposing that the solution has to minimize its response to the noise. This is equivalent to take the solution of the above system with the minimum norm among all the possible solutions.

$$\begin{matrix} \underline{\underline{X}}.\underline{a} = \underline{x} \\ \underline{a}^H.\underline{a}\big|_{MIN} \end{matrix}$$

the solution is $\underline{a} = \underline{\underline{X}}^H.\left(\underline{\underline{X}}.\underline{\underline{X}}^H\right)^{-1}.\underline{x}$ and the spectral estimate is derived based in the fact that the roots of the polynomial formed by the vector lie in one circle inside the unit circle, but two of them must lie in the frequencies corresponding with the location of the two frequencies contained in the data record. In summary, the two major peaks of $\hat{S}_{xx}(w) = \frac{1}{\left|\underline{a}^H.\underline{S}\right|^2}$ will coincide, or close depending on the signal to noise ratio, to the actual frequencies.

Since $\left(\underline{\underline{X}}^H.\underline{\underline{X}}\right)$ is 2x2, the product by vector x will be 2x1

$$\left(\underline{\underline{X}}^H.\underline{\underline{X}}\right)^{-1}.\underline{x} = \begin{bmatrix} \beta 1 \\ \beta 2 \end{bmatrix}$$

and

$$\underline{a}^H.\underline{S} = \begin{bmatrix} \beta 1 & \beta 2 \end{bmatrix}.\underline{\underline{X}}.\underline{S} = \beta 1.\left(DFT\{signal\ forward\} + \beta 2.DFT\{DFTsignal\ backward\}\right)$$

Then, the resulting procedure is basically the combination of two period grams weighted in such a manner that they null out at the actual frequency locations.

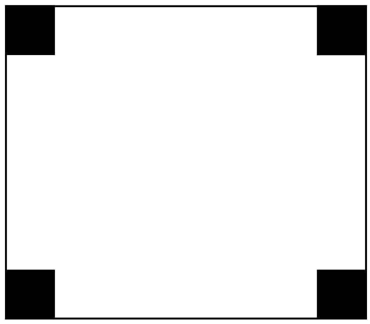
The similarity wit MUSIC or, much close to the pioneer work of Pisarenko, can be easily viewed when, instead of minimum norm from and over determined system of equations, we force unity norm with minimum prediction error (when vector $\underline{a}$ has been extended previously with anew coefficient a(N-1) in order to leave cero in the second term)

$$\begin{matrix} \underline{a}^H.\underline{a} = 1 \\ \left|\underline{\underline{X}}.\underline{a}\right|^2_{MIN} \end{matrix}$$

the solution to this system is the minimum eigenvector associated with matrix $\left(\underline{\underline{X}}^H.\underline{\underline{X}}\right)$. In fact, any of the N-2 eigenvectors with minimum eigenvalue will be adequate. (the reader may extent this idea to use all the noise eigenvalues to derive Music).

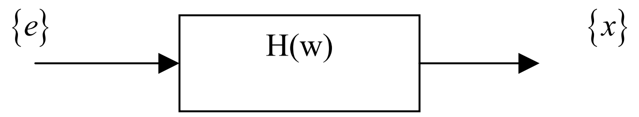
III.10.12.-

For the 2D case, there are four possible equations that allow the use of the overall data when formulating the exact prediction. These samples are the four corners of the original 2D data. Note that we search for 4 frequencies, two pairs (w,Φ).



Simples to be predicted from the rest of the data.
Four equations to define the four unknowns. From all the possible solution, again we select the minimum norm one.

III.10.13.-



a) $\hat{S}_x(w) = |H|^2 .\hat{S}_e(w)$

b) $E\left[\hat{S}_x(w)\right] = |H|^2 .E\left[\hat{S}_e(w)\right] = \left|\begin{matrix} Unbiased\ \ for \\ white\ noise \end{matrix}\right| = |H|^2 .N_0$

c) $\text{var}^2\left(\hat{S}_x(w)\right) = E\left[\left(\hat{S}_x - E\left(\hat{S}_x\right)\right)^2\right] = E\left(\hat{S}_x^2\right) - E^2\left(\hat{S}_x\right) = E\left(\hat{S}_x^2\right) - |H|^4 .N_0^2$ and

$E\left(\hat{S}_x^2\right) = |H|^4 .E\left(\hat{S}_e^2\right)$ altogether

$$\text{var}^2\left(\hat{S}_x(w)\right) = \left(|H|^2 N_0\right).\left[\frac{E\left(\hat{S}_e^2\right)}{N_0^2} - 1\right] = S_x^2(w).\left[\frac{E\left(\hat{S}_e^2\right)}{N_0^2} - 1\right]$$

because $E\left[\hat{S}_e^2\right] = \text{var}^2(\hat{S}_e) + E^2\left(\hat{S}_e\right) = \text{var}^2(\hat{S}_e) + N_0^2$ the desired result follows

$$\text{var}^2\left(\hat{S}_x(w)\right) = \hat{S}_x^2(w).\frac{\text{var}^2\left(\hat{S}_e^2\right)}{N_0^2} = S_x^2(w).\left[\frac{Sin^2\left(Q.w\right)}{Q^2.Sin^2\left(w\right)} + 1\right]$$

III.10.14.-

a) See chapter content.

b) $Q.P_x(w) = \underline{S}^H.\underline{\underline{R}}.\underline{S} = \alpha_d.\left|\underline{S}^H.\underline{S}_d\right|^2 + \underline{S}^H.\underline{\underline{R}}_o.\underline{S} = \alpha_d.Q^2 + \underline{S}^H.\underline{\underline{R}}_o.\underline{S}$ and

$$P_x = Q.\alpha_d + \frac{\underline{S}^H.\underline{\underline{R}}_o.\underline{S}}{Q}$$

c) $\left(\underline{S}^H.\underline{\underline{R}}.\underline{S}\right).\left(\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}\right) \geq \left|\underline{u}^H.\underline{v}\right|^2 = \left|\underline{S}^H.\underline{S}\right|^2$ *being* $\underline{u} = \underline{\underline{R}}^{1/2}.\underline{S}$ $\underline{v} = \underline{\underline{R}}^{-1/2}.\underline{S}$

d) Equal when $\underline{u} \propto \underline{v}$ or $\underline{\underline{R}}^{1/2}.\underline{S} = \beta.\underline{\underline{R}}^{-1/2}.\underline{S}$ *or* $\underline{\underline{R}}.\underline{S} = \beta.\underline{S}$. This last relationship is true when either the acf matrix is the identity, within a constant or the frequency $\underline{S}$ is orthogonal to the actual frequency $\underline{S}_d$

e) We like to prove that $\frac{\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}}{\underline{S}^H.\underline{\underline{R}}^{-2}.\underline{S}} \leq \frac{1}{\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}}$ because $S_{NMLM}(w) \leq S_{MLM}(w)$ implies greater resolution of the first with respect the second, whenever at the actual frequency they got the same value.

To prove this, note that $\left|\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}\right|^2 \leq |\underline{u}|^2.|\underline{v}|^2 = Q^2.\left(\underline{S}^H.\underline{\underline{R}}^{-2}.\underline{S}\right)$ *where* $\underline{u} = \underline{S}$ *and* $\underline{v} = \underline{\underline{R}}^{-1}.\underline{S}$.

Thus $S_{NMLM}(w) \leq Q^2.S_{MLM}(w)$ being equal when both vectors are proportional. This occurs at the actual frequency location since the actual steering vector is an eigenvector of the acf matrix.

----

f)
$$\Lambda\left(\alpha, \underline{\underline{R}}_0, \underline{S}\right) = \left(\underline{X}_n - \alpha.\underline{S}\right)^H.\underline{\underline{R}}_0^{-1}.\left(\underline{X}_n - \alpha.\underline{S}\right)$$
$$\alpha_{ML} = \left\{\frac{\partial\Lambda}{\partial\alpha^H} = 0\right\} = \frac{\underline{S}^H.\underline{\underline{R}}_0^{-1}.\underline{X}_n}{\underline{S}^H.\underline{\underline{R}}_0^{-1}.\underline{S}}$$

----

g)
$$\text{var}^2_{ML} = \frac{1}{\underline{S}^H.\underline{\underline{R}}_0^{-1}.\underline{S}}$$
$$\text{var}^2_{Capon} = \frac{\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{\underline{R}}_0.\underline{\underline{R}}^{-1}.\underline{S}}{\left(\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}\right)^2}$$
since
$$\left(\underline{S}^H.\underline{\underline{R}}_0^{-1}.\underline{S}\right).\left(\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{\underline{R}}_0.\underline{\underline{R}}^{-1}.\underline{S}\right) \geq \left(\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}\right)^2$$
where $\underline{u} = \underline{\underline{R}}_0^{1/2}.\underline{S}$ *and* $\underline{v} = \underline{\underline{R}}_0^{1/2}.\underline{\underline{R}}.\underline{S}$

then $\text{var}^2_{Capon} = \frac{\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{\underline{R}}_0.\underline{\underline{R}}^{-1}.\underline{S}}{\left(\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}\right)^2} \geq \frac{\underline{S}^H.\underline{\underline{R}}^{-1}.\underline{S}}{\left(\underline{S}^H.\underline{\underline{R}}_0^{-1}.\underline{S}\right)^2}$ and $\text{var}^2_{Capon} \geq \text{var}^2_{ML}$

$\underline{u} = \beta.\underline{v}$; $\underline{\underline{R}}^{1/2}.\underline{S} = \beta.\underline{\underline{R}}_0^{1/2}.\underline{\underline{R}}^{-1}.\underline{S}$; $\underline{\underline{R}}.\underline{\underline{R}}_0^{-1}\underline{S} = \beta.\underline{S}$

only when the noise is white

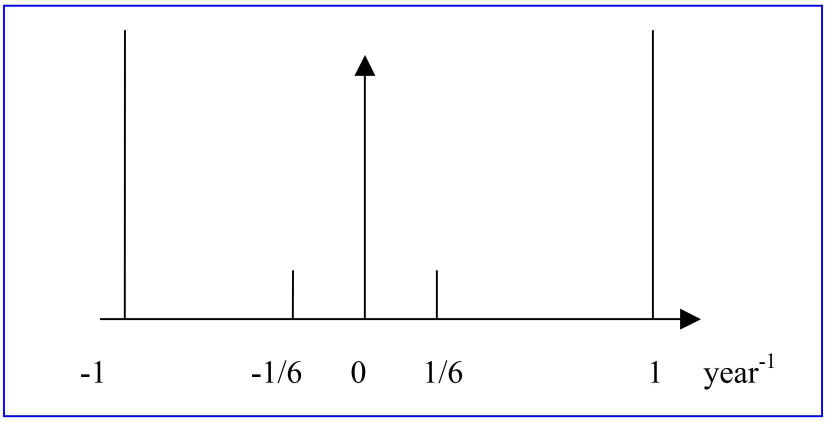
$\left(\alpha.\underline{S}_d.\underline{S}_d^H + \underline{I}\right).\underline{S} = \beta.\underline{S}$

and when the steering $\underline{S}$ *coincides with the actual frequency* $\underline{S}_d$

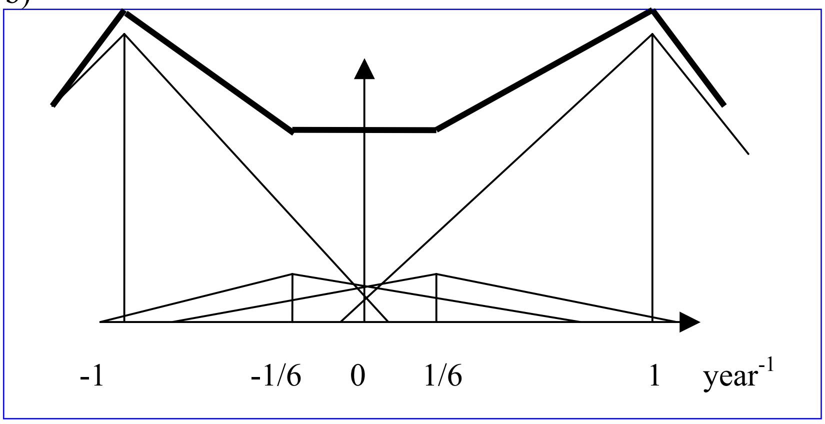
both var*iances coincide*

III.10.15.-

a)

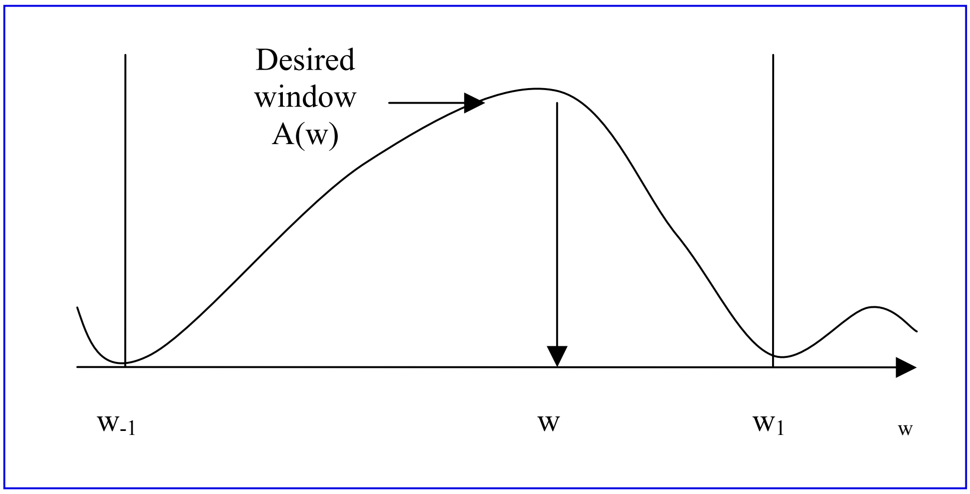


b)



Being so small the line a 1/6 is severely masked out due to the window leakage promoted by the strong line at 1.

Assuming that we are looking for the spectral content at frequency w_0, the best choice for the data window is to have cero response an f_l=1, i.e. the window will have a cero at w_0-w_1. Also a cero is necessary to remove leakage from line at -1, i.e. at w_0-w_{-1}.



d) Assuming that only one line has to be removed, the constraint is $A(w_1)=0$. The use of a single frequency of nulling in the design is for the shake of presentation and, without loss of generality, the design included hereafter can be extended to several nulling frequencies. Going back to the window design, we have two constraints: The value at w=0 has to be one (no bias) and cero to null w_1, in consequence:

$$\underline{a}^H\left[\underline{1} \quad \underline{S}_\Delta\right]=\left[1 \quad 0\right] \quad or \quad \underline{a}^H.\underline{\underline{\Psi}}=\underline{1}^H$$

where a is the vector containing the window taps

The design is completed with minimum bandwidth (white noise equivalent bandwidth, i.e. the norm of vector a minimum, which is equivalent to minimum d.c. response of the lag-window corresponding to the data window under design.

$$\underline{a}^H.\underline{a}\Big|_{MIN}$$

and the solution is: $\underline{a}=\underline{\underline{\Psi}}.\left(\underline{\underline{\Psi}}^H.\underline{\underline{\Psi}}\right)^{-1}.\underline{1}$

The norm of the window is:

$$\underline{a}^H.\underline{a}=\left[1 \quad 0\right].\left(\underline{\underline{\Psi}}^H.\underline{\underline{\Psi}}\right)^{-1}.\begin{bmatrix}1\\0\end{bmatrix}=\frac{1}{Q^2-\alpha^2}.\left(Q-\alpha^*\right).\begin{bmatrix}1\\0\end{bmatrix}=\frac{Q}{Q^2-\alpha^2}$$

$$also \quad \underline{a}^H.\begin{bmatrix}1\\..\\1\end{bmatrix}=1$$

e) The length of the segment, given the desired resolution, is 32 months, then

$$10dB.\Rightarrow 10 \; averages \Rightarrow 10 \; records \Rightarrow 320 \; months$$

e) It is easy to check that, for this length of the data record and maintaining the segment resolution, it does not change very much the results obtained for a rectangular or triangular window.

III.10.16.- and **III.10.17.-**

The solution is easily derived from the content of this chapter.

III.10.18.-

$$x_0\begin{cases}E\left(x_0\right)=x\\ \text{var}^2\left(x_0\right)=\sigma_0^2 \text{ (min.)}\end{cases} \qquad x_1=\rho.x_0$$

$$\xi=E\left[\left(x_1-x\right)^2\right]=E\left(x_1^2\right)+x^2-2.x.E\left(x_1\right)=\rho^2.E\left(x_0^2\right)+x^2-2.\rho.x^2$$

$$\frac{\partial\xi}{\partial\rho}=2.\rho.E\left(x_0^2\right)-2.x^2 \quad ; \quad \rho=\frac{x^2}{E\left(x_0^2\right)} \quad ; \quad \sigma_0^2=E\left(x_0^2\right)-x^2$$

in consequence: $\rho = \dfrac{x^2}{\sigma_0^2 + x^2}$

In the case of Periodogram, $\sigma_0^2 = x^2$ and $\rho = \dfrac{1}{2}$

This exercise shows how the MSE is a trade-off between bias and variance. In any case, the relevance of a scale factor is removed when a logarithmic scale is used to plot the estimates.

III.10.19.-

a) $\underline{h}_{pr} = \underline{1}$ *all ones vector*

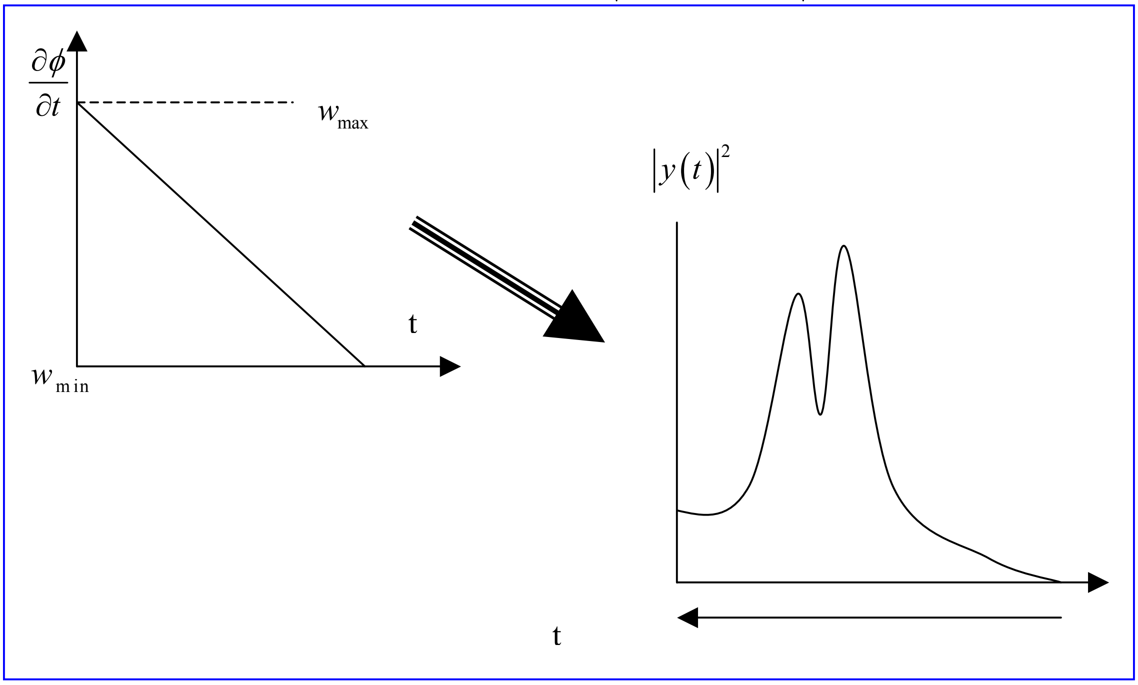
b) The filter, at the frequency domain, is not invariant and it depends on the data.

$$y(t) = x(t).\exp\left(j.\phi(t)\right) * h(t) = \int h(\tau).x(t-\tau).\exp\left(j.\phi(t-\tau)\right).d\tau =$$

c) $\cong \int h(\tau).\exp\left(-j.w_0.\tau\right).x(t-\tau).\exp\left(j.\phi(t)\right).d\tau = \left| \begin{matrix} with \\ w_0 = \dfrac{\partial\phi}{\partial t} \end{matrix} \right| =$

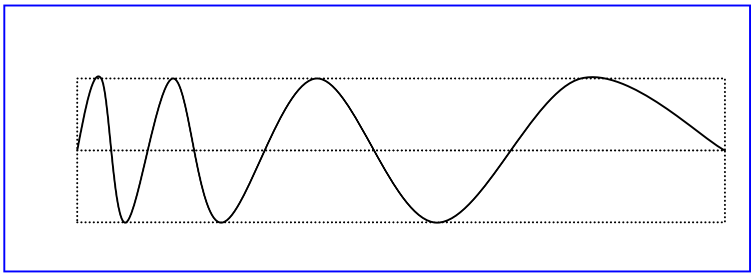
$$= \exp\left(j.\phi(t)\right).\int_T h(\tau).x(t-\tau).\exp\left(-j.w_0.\tau\right).d\tau$$

This last integral is just the DFT of signal x(.) windowed by h(.), along the duration of the impulse response T). In summary $\left|y(t)\right|^2 = \left|DFT(x)_{w=w_0(t)}\right|^2$



d) The filter length (close to 1/T Hz.)

e) The instantaneous frequency decreases with time



f) It is similar to the Periodogram or modulus square DFT.

III.10.21

MEM set correlation constrains in such a way that the estimate the same autocorrelation values that the data record.

The 2Q+1 correlation constrains are:

$$\frac{1}{2\pi}.\int_{-\pi}^{\pi} s_x(w).\exp(jqw).dw = r(q); q = -Q,Q$$

The objective is the maximum flatness of the estimate that, at the same time, satisfies the correlation constrains. Maximum flatness is equivalent to maximum entropy. Thus, the objective is:

$$\left.\frac{1}{2\pi}.\int_{-\pi}^{\pi} Ln(s_x(w)).dw\right]_{MAX}$$

2A.- When there is only one correlation constrain the formulation of the estimate is:

$$\left.\frac{1}{2\pi}.\int_{-\pi}^{\pi} Ln(s_x(w)).dw\right]_{MAX}$$

$$\frac{1}{2\pi}.\int_{-\pi}^{\pi} s_x(w).dw = r(0)$$

The Lagrangian would be (Only a single multiplier).

$\Psi(s_x(w)) = \int_{-\pi}^{\pi}\left[Ln(s_x(w) - \lambda_0.s_x(w)\right].dw$ deriving with respect the estimate is equivalent to set to zero the derivative of the integrand. Setting to zero the derivative the estimate results.

$\frac{1}{s_x(w)} - \lambda_0 = 0$ i.e. the solution is: $s_x(w) = \frac{1}{\lambda_0}; \quad \forall w$. In consequence the maximum entropy estimate with power constrain (Just the zero lag of the autocorrelation) is white noise with the same power.

3A.- Solving the general problem, the integrand of the Lagrangian is:

$Ln(s_x(w)) - \sum_{q=-Q}^{Q} \lambda_q . s_x(w).\exp(jqw)$ and its derivative set to zero results in:

$$\frac{1}{s_x(w)} - \sum_{q=-Q}^{Q} \lambda_q .\exp(jqw) = 0 \text{ or } s_x(w) = \frac{1}{\sum_{q=-Q}^{Q} \lambda_q .\exp(jqw)}.$$

The denominator polynomial is even and, as a consequence, it can be decomposed in two polynomials containing the inside and outside roots of the unit circle respectively.

$s_x(w) = \frac{1}{\sum_{q=-Q}^{Q} \lambda_q .\exp(jqw)} = \frac{1}{B(\exp(-jw)).B(\exp(jw))}$ and, normalizing the first coefficient to one $B(\exp(jw)) = \sum_{p=0}^{Q} b(p).\exp(jpw) = b(0).\left[1 + \sum_{p=1}^{Q} \left(\frac{b(p)}{b(0)}\right).\exp(jpw)\right]$ or

$B(\exp(jw)) = \left(\frac{1}{\sigma}\right).A(\exp(jw))$ In summary, the maximum entropy with correlation constrains can be written as:

$s_x(w) = \frac{\sigma^2}{|A(\exp(jw))|^2}$ which coincides with the AR model estimate of the same order.

4A.-

The Yules-Walker equations, allowing the computation of the coefficients of an AR model of order Q are as follows:

$\underline{\underline{R}}.\underline{A} = \sigma^2.\begin{bmatrix} 1 \\ 0 \\ .. \end{bmatrix} = \sigma^2.\underline{1}$ being $\underline{1}^T = [1 \quad 0 \quad .. \quad 0]$, $\underline{\underline{R}}$ is the autocorrelation matrix, $\underline{A}$ the vector containing the denominator coefficients with the first coefficient equal to one.

$\underline{A} = [1 \quad a(1) \quad ... \quad a(Q)]^T$

The design of a filter with minimum power at its output when the input is the signal {x} implies to minimize $\underline{A}^H.\underline{\underline{R}}.\underline{A}\big|_{MINIMO}$ with the constrain of the first coefficient equal to one that can be set, in a vector form, as $\underline{A}^H.\underline{1} = 1$. The solution to this problem is:

$\underline{A} = \frac{\underline{\underline{R}}^{-1}.\underline{1}}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}}$ or $\underline{\underline{R}}.\underline{A} = \left(\frac{1}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}}\right).\underline{1}$ where it is evident that these equations are identical to the Yules-Walker equations. In summary, the solution coincides for both designs.

Clearly, the noise power at the input of the AR model can be computed as: $\sigma^2 = \frac{1}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}}$

5A.- When the signal under analysis is pure AR then the white noise at the input of the model is Gaussian and verifies the AR equation as:

$w(n) = x(n) + \sum_{q=1}^{Q} a(q).x(n-q) = \underline{A}^H.\underline{X}_n$ The likelihood will be:

$\Pr\left({w(n)}/{\underline{A}} \right) = cte.\exp\left\{ -\underline{A}^H.\underline{\underline{R}}.\underline{A} \right\}$ Clearly the maximization of the likelihood implies the minimization of the exponent yet preserving the first coefficient equal to one. Thus, the previous design coincides with the maximum likelihood procedure to estimate the AR coefficients.

IV.9.1.-

$$\left|\underline{\underline{X}}_n.\underline{a}-\underline{d}_n\right|^2_{MIN} \quad \underline{a}^H.\underline{a}=1$$

solving this problem, the solution is: $\underline{a}=\left(\underline{\underline{X}}_n^H.\underline{\underline{X}}_n+\lambda.\underline{\underline{I}}\right)^{-1}.\left(\underline{\underline{X}}_n^H.\underline{d}_n\right)$

Note that the Lagrange multiplier, always greater than cero, acts as white noise is added to the original data. This procedure is also known as the diagonal loading method to make more robust to impairments the resulting Wiener filter.

IV.9.2.-

$$\underline{X}_n=\underline{d}+\underline{w}_n \quad \underline{\underline{R}}=\underline{d}.\underline{d}^H+\sigma^2.\underline{\underline{I}} \quad \underline{\underline{R}}^{-1}=\frac{1}{\sigma^2}.\left(\underline{\underline{I}}-\frac{\underline{d}.\underline{d}^H}{\sigma^2+\left|\underline{d}\right|^2}\right) \quad \underline{P}=P_d.\underline{d}$$

then $\underline{\underline{R}}^{-1}.\underline{P}=\frac{P_d}{\sigma^2}.\underline{d}.\left(1-\frac{\left|\underline{d}\right|^2}{\sigma^2+\left|\underline{d}\right|^2}\right)=\underline{d}.\frac{P_d}{\sigma^2+\left|\underline{d}\right|^2}$ which reveals that the solution coincides, within a constant that does not modify the output SNR, with the vector containing the deterministic signal component.

IV.9.3.-

$H(l)=\frac{\sum Y(l).X^*(l)}{\sum X(l).X^*(l)}$ this expression reveals that the computation in the frequency domain of the Wiener filter from the DFTs of segments, corresponding to the input and output signals, faces a problem of spectral estimation. In fact, see chapter II when describing the spectral coherence, the actual expression of the optimum filter is given by the quotient of the cross-spectral density between the output and the input divided by the spectral density of the input. $H(w)=\frac{S_{xy}(w)}{S_{xx}(w)}$. Windows, number of segments, resolution, etc. are the problems that this way of getting the Wiener filter have to overpass. On the other hand, the processing and design is done only with FFTs. It is more convenient than traditional time-domain design when the length of the filter is long (above 64 samples).

IV.9.4.-

$$\hat{x}(n+2)=-a(1).x(n-1)-a(2).x(n-2)$$

since the prediction error has to be orthogonal to the data,

$$\begin{bmatrix} r(3) \\ r(4) \end{bmatrix}+\begin{pmatrix} r(0) & r(1) \\ r(1) & r(0) \end{pmatrix}\begin{bmatrix} a(1) \\ a(2) \end{bmatrix}=\underline{0} \quad \textit{and the prediction error} \quad \xi=r(0)+a(1).r(3)+a(2).r(4)$$

or in compact form (Y-W equations for this case)

$$\begin{pmatrix} r(0) & r(3) & r(4) \\ r(3) & r(0) & r(1) \\ r(4) & r(1) & r(0) \end{pmatrix} \begin{pmatrix} 1 \\ a(1) \\ a(2) \end{pmatrix} = \begin{pmatrix} \xi \\ 0 \\ 0 \end{pmatrix}$$

IV.9.5.-

The proper choice is to select those lags when the estimate acf shows the greatest absolute values. The reason is that the prediction is favored with the fact that the samples use to make the prediction have high correlation (positive or negative) with the sample to be predicted.

IV.9.6.-

$$\begin{bmatrix} 1 & 0.5 & a \\ 0.5 & 1 & 0.5 \\ a & 0.5 & 1 \end{bmatrix} = \underline{\underline{R}}$$ Using Levinson for this matrix:

$$k_1 = a_1^1 = -0.5 \quad \sigma_1^2 = 1.(1-0.5^2) = 0.75$$

....

$$\Delta_1 = a + 0.5.a_1^1 = a - 0.25 \quad k_2 = -\frac{a-0.25}{0.75} = a_2^2 \quad \sigma_2^2 = 0.75.\left(1-k_2^2\right)$$

Clearly the MEM extrapolation for r(2) is that the corresponding Parcor is cero, i.e. the value of a that coincides with the MEM extrapolation of r(0) and r(1) is 0.25.

The range of values for a (see IV.35) is $r_{MEM}(2) \pm \sigma_1^2 = \begin{cases} 1 \\ -0.5 \end{cases}$

IV.9.7.-

$$k_1 = 0.5 \Rightarrow r(1) = -k_1.r(0) = -0.5.r(0) \Rightarrow \sigma_1^2 = (1-0.25).r(0)$$
$$k_2 = 0.2 \Rightarrow \Delta_2 = -0.2.\sigma_1^2 = 0.15.r(0) \Rightarrow r(2) + 0.5.r(1) = 0.15.r(0) \Rightarrow r(2) = 0$$

IV.9.8.-

$$k_Q = \frac{2.\sum F.B}{\sum F^2 + \sum B^2} \leq \frac{\sum F.B}{\sqrt{\left(\sum F^2\right).\left(\sum B^2\right)}} \leq 1$$

We used that the arithmetic mean is less than or equal than the geometric mean and, later on $\sum |u|^2 . \sum |v|^2 \geq \left(\sum u.v\right)^2$

IV.9.9.-

$$e_b^Q(n) = \left(\underline{a}_b^Q\right)^H . \begin{bmatrix} x(n-Q) \\ x(n-Q+1) \\ \ldots \\ x(n) \end{bmatrix} \quad and$$

$$e_b^{Q-1}(n) = \left(\underline{a}_b^{Q-1}\right)^H . \begin{bmatrix} x(n-Q+1) \\ x(n-Q+2) \\ \ldots \\ x(n) \end{bmatrix} = e_b^Q(n) = \left(\begin{bmatrix} 0 \ \underline{a}_b^Q \end{bmatrix}\right)^H . \begin{bmatrix} x(n-Q) \\ x(n-Q+1) \\ \ldots \\ x(n) \end{bmatrix}$$

then

$$E\left[e_b^Q(n).e_b^{Q-1}(n)\right] = \left(\underline{a}_b^Q\right)^H .E\left[\begin{pmatrix} x(n-Q) \\ x(n-Q+1) \\ \ldots \\ x(n) \end{pmatrix}.\begin{bmatrix} x(n-Q) & x(n-Q+1) & \ldots & x(n) \end{bmatrix}\right]\begin{bmatrix} 0 \\ a_b^{Q-1} \end{bmatrix}$$

or $E\left[e_b^Q(n).e_b^{Q-1}(n)\right] = \left(\underline{a}_b^Q\right)^H .\underline{\underline{R}}.\begin{bmatrix} 0 \\ a_b^{Q-1} \end{bmatrix}$.

Now using the expression of the Y-W equations for the length Q predictor

$\underline{\underline{R}}.\underline{a}_b^Q = \begin{bmatrix} \sigma_{b,Q}^2 \\ 0 \\ .. \\ 0 \end{bmatrix}$ we have $E\left[e_b^Q(n).e_b^{Q-1}(n)\right] = \begin{bmatrix} \sigma_{b,Q}^2 & 0 & .. & 0 \end{bmatrix}.\underline{\underline{R}}^{-1}.\underline{\underline{R}}.\begin{bmatrix} 0 \\ \underline{a}_b^{Q-1} \end{bmatrix} = 0$

IV.9.10.-

$$\underline{X}_n = \alpha.e^{jw_d.n}.\underline{S}_d + \underline{w}_n$$

$$\underline{h} = \frac{\underline{\underline{R}}^{-1}.\underline{1}}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}} \qquad \xi_{MIN} = \frac{1}{\underline{1}^H.\underline{\underline{R}}^{-1}.\underline{1}}$$

$$\underline{\underline{R}} = \alpha^2.\underline{S}_d.\underline{S}_d^H + \underline{\underline{R}}_0 \qquad \underline{\underline{R}}^{-1} = \underline{\underline{R}}_0^{-1} - \frac{\alpha^2}{1+\rho}.\underline{\underline{R}}_0^{-1}.\underline{S}_d.\underline{S}_d^H.\underline{\underline{R}}_0^{-1} \quad with \quad \rho = \alpha^2.\underline{S}_d.\underline{\underline{R}}_0^{-1}.\underline{S}_d^H$$

$$\underline{\underline{R}}^{-1}.\underline{1} = \left(\underline{\underline{I}} - \frac{\alpha^2}{1+\rho}.\underline{\underline{R}}_0^{-1}.\underline{S}_d.\underline{S}_d^H\right).\underline{\underline{R}}_0^{-1}.\underline{1}$$

$$\underline{1}^H.\underline{\underline{R}}^{-1}\underline{1} = \left(\underline{1}^H.\underline{\underline{R}}_0^{-1}\underline{1}\right)\left(1 - \frac{\alpha^2}{1+\alpha^2.\underline{S}_d.\underline{\underline{R}}_0^{-1}.\underline{S}_d^H}.\frac{\left|\underline{1}^H.\underline{\underline{R}}_0^{-1}.\underline{S}_d\right|^2}{\left(\underline{1}^H.\underline{\underline{R}}_0^{-1}\underline{1}\right)}\right)$$

Now, using the definition of the spectral estimates of MLM and MEM for the noise covariance matrix,

$$\xi_{MIN} = \left(\underline{1}^H . \underline{\underline{R}}_0^{-1} \underline{1}\right)^{-1} \left(1 - \frac{\alpha^2}{1 + \dfrac{\alpha^2}{S_0^{MLM}(w_d)}} . \frac{1}{S_0^{MEM}(w_d)}\right)^{-1} \quad \text{or}$$

$$\xi_{MIN} = \left(\underline{1}^H . \underline{\underline{R}}_0^{-1} \underline{1}\right)^{-1} \left(\frac{1 + \alpha^2 \left(\dfrac{1}{S_0^{MLM}(w_d)} \right)}{1 + \alpha^2 . \left(\dfrac{1}{S_0^{MLM}(w_d)} - \dfrac{1}{S_0^{MEM}(w_d)} \right)} . \right)$$

Note that depending on the spectral density of the noise at the frequency of interest the noise reduction will change. When the MEM estimate of the noise is very high then $\xi_{MIN} = \left(\underline{1}^H . \underline{\underline{R}}_0^{-1} \underline{1}\right)^{-1}$. Meanwhile the density is very low, i.e. a band pass filter may take out the line, then $\xi_{MIN} \Rightarrow 0$.

IV.9.11.-

Since $\underline{x} = \underline{\underline{U}}.\underline{b}$ *and* $\underline{y} = \underline{\underline{V}}.\underline{a}$ where matrixes $\underline{\underline{U}}$ and $\underline{\underline{V}}$ contain the samples of the input and the output respectively and properly arranged.

Now

$$\underline{e} = \underline{x} - \underline{y}$$

$$E\left[\underline{e}.\underline{e}^H\right] = E\left[\left(\underline{x} - \underline{y}\right).\left(\underline{x} - \underline{y}\right)^H\right] =$$

$$= E\left[\underline{b}.\underline{\underline{U}}.\underline{\underline{U}}^H \underline{b}^H\right] + E\left[\underline{a}.\underline{\underline{V}}.\underline{\underline{V}}^H \underline{a}^H\right] - E\left[\underline{b}.\underline{\underline{U}}.\underline{\underline{V}}^H \underline{a}^H\right] - E\left[\underline{a}.\underline{\underline{V}}.\underline{\underline{U}}^H \underline{b}^H\right] =$$

$$= \underline{b}^H . \underline{\underline{R}}_{uu} . \underline{b} + \underline{a}^H . \underline{\underline{R}}_{vv} . \underline{a} - \underline{b}^H . \underline{\underline{R}}_{uv} . \underline{a} - \underline{a}^H . \underline{\underline{R}}_{vu} . \underline{b}$$

The minimization of the error entails the following gradients equal to cero:

$$\xi = \underline{b}^H . \underline{\underline{R}}_{uu} . \underline{b} + \underline{a}^H . \underline{\underline{R}}_{vv} . \underline{a} - \underline{b}^H . \underline{\underline{R}}_{uv} . \underline{a} - \underline{a}^H . \underline{\underline{R}}_{vu} . \underline{b}$$

....

$$\frac{\partial \xi}{\partial \underline{b}^H} = \underline{0} = \underline{\underline{R}}_{uu} . \underline{b} - \underline{\underline{R}}_{uv} . \underline{a} \quad \Rightarrow \quad \underline{\underline{R}}_{uu} . \underline{b} = \underline{\underline{R}}_{uv} . \underline{a}$$

$$\frac{\partial \xi}{\partial \underline{a}^H} = \underline{0} = \underline{\underline{R}}_{vv} . \underline{a} - \underline{\underline{R}}_{vu} . \underline{b} \quad \Rightarrow \quad \underline{\underline{R}}_{vv} . \underline{a} = \underline{\underline{R}}_{vu} . \underline{b}$$

These two equations correspond with the Wiener solutions of one given the another one.

Combining both: $\underline{\underline{R}}_{vv} . \underline{a} = \underline{\underline{R}}_{vu} . \underline{\underline{R}}_{uu}^{-1} . \underline{\underline{R}}_{uv} . \underline{a}$

Solving this equation may be is not possible since it implies the existence of an eigenvalue equal to one. When solved for all the eigenvalues,

$\lambda.\underline{a} = \left(\underline{\underline{R}}_{vv} - \underline{\underline{R}}_{vu}.\underline{\underline{R}}_{uu}^{-1}.\underline{\underline{R}}_{uv}\right).\underline{a}$

and then the error is not longer cero but equals to:

$\underline{b} = \underline{\underline{R}}_{uu}^{-1}.\underline{\underline{R}}_{uv}.\underline{a}$

$$\xi_{MIN} = \underline{a}^{H}.\left(\underline{\underline{R}}_{vv} - \underline{\underline{R}}_{vu}.\underline{\underline{R}}_{uu}^{-1}.\underline{\underline{R}}_{vu}\right).\underline{a} = \lambda$$

which implies that the optimum is the eigenvector associated to the maximum eigenvalue. Note that this technique does not guarantee that the denominator will be a minimum phase polynomial.

IV.9.12, IV.9.13 and IV.9.14

They can be solved directly from the chapter content and previous exercises

IV.9.16.

$$\Pr\left(\underline{X}_n / i(n)\right) = cte.\exp\left\{-\left[\left(\underline{X}_n - \underline{h}.i(n)\right)^H \underline{\underline{R}}_0^{-1}.\left(\underline{X}_n - \underline{h}.i(n)\right)\right]\right\}$$

$$\frac{d(.)}{di^*(n)} = 0 = \underline{h}^H.\underline{\underline{R}}_0^{-1}.\left(\underline{X}_n - \underline{h}.i(n)\right) \qquad i^{ML}(n) = \frac{\underline{h}^H.\underline{\underline{R}}_0^{-1}.\underline{X}_n}{\underline{h}^H.\underline{\underline{R}}_0^{-1}.\underline{h}}$$

2B

$\xi = \left|\underline{X}_n - \underline{h}.i(n)\right|^2$ which is the same than the exponent of the likelihood when $\underline{\underline{R}}_0$ is a constant by the identity matrix. In other words: $i^{MSE}(n) = \dfrac{\underline{h}^H.\underline{X}_n}{\underline{h}^H.\underline{h}}$

2C

$\Psi = \left[\underline{h}_e^H.\underline{X}_n - i(n)\right]^H.\left[\underline{h}_e^H.\underline{X}_n - i(n)\right] \qquad \dfrac{d\Psi}{d\underline{h}_e^H} = \underline{X}_n.\left[\underline{X}_n^H\underline{h}_e - i^*(n)\right]$ and minimizing the expected value of Ψ we arrive to: $E\left(\underline{X}_n.\underline{X}_n^H\right)\underline{h}_e = E\left(\underline{X}_n.i^*(n)\right)$ which is the corresponding Wiener filter for the given problem. $\underline{h}_e = \underline{\underline{R}}^{-1}.\underline{h}$

2D

The Wiener filter is $\underline{h}_e = \underline{\underline{R}}^{-1}.\underline{h}$, also $\underline{\underline{R}} = \underline{h}.\underline{h}^H + \underline{\underline{R}}_0$ and, using the Inverse Lemma we obtain:

$\underline{h}_e = \left(\underline{h}.\underline{h}^H + \underline{\underline{R}}_0\right)^{-1}.\underline{h} = \left(\underline{\underline{R}}_0^{-1}.\underline{h} - \dfrac{\underline{\underline{R}}_0^{-1}.\underline{h}.\underline{h}^H.\underline{\underline{R}}_0^{-1}.\underline{h}}{1 + \underline{h}^H.\underline{\underline{R}}_0^{-1}.\underline{h}}\right) = \underline{\underline{R}}_0^{-1}.\underline{h}.\left(1 - \dfrac{\rho}{1+\rho}\right)$ thus, both differ in a constant.

IV.9.17.

a.- The output signal of the system is $\begin{aligned} x(n) &= d(n) + a.d(n-1) \\ x(n-1) &= d(n-1) + a.d(n-2) \end{aligned}$ thus, the vector formulation for the two samples will be:

$$\begin{bmatrix} x(n) \\ x(n-1) \end{bmatrix} = \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & a \end{bmatrix} . \begin{bmatrix} d(n) \\ d(n-1) \\ d(n-2) \end{bmatrix} \quad \text{and } \underline{X}_n = \underline{\underline{H}}.\underline{d}_n \text{ e } \underline{Y}_n = \underline{X}_n + \underline{w}_n = \underline{\underline{H}}.\underline{d}_n + \underline{w}_n$$

since d(n) is uncorrelated with unity power ($E\left(\underline{d}_n.\underline{d}_n^H\right) = \underline{\underline{I}}_3$) , and the noise is uncorrelated and independent of y(n), we have $\underline{\underline{H}}.\underline{\underline{H}}^H + \sigma^2.\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

b.- Given a reference at the output of the first linear system, since the reference at the output of the first linear system has a delay of one simple (it passes by a FIR filter of two coefficients) it has no sense to ask that the system recovers by anticipation. Also it can be seen that the 3 samples of d(n) that participate on vector $\underline{X}_n$, the simple that appears more times than the rest is d(n-1), in consequence it is intuitive that this lag will be the most easy to handle.

c.-

$$\underline{\underline{H}}.\underline{\underline{H}}^H + \sigma^2.\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & a \end{bmatrix} . \begin{bmatrix} 1 & 0 \\ a & 1 \\ 0 & a \end{bmatrix} + \sigma^2.\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+a^2+\sigma^2 & a \\ a & 1+a^2+\sigma^2 \end{bmatrix}$$

then $\underline{\underline{R}}^{-1} = \dfrac{1}{\left(1+a^2+\sigma^2\right)^2 - a^2}\begin{bmatrix} 1+a^2+\sigma^2 & -a \\ -a & 1+a^2+\sigma^2 \end{bmatrix}$ Also, the P-vector of the Wiener filter will be:

$\underline{P} = E\left(\underline{Y}_n.d(n-1)\right) = \underline{\underline{H}}.\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ 1 \end{bmatrix}$ Then, multiplying by the autocorrelation inverse the Wiener filter results.

The minimum error power is given by $\xi_{\min} = P_d - \underline{P}^H.\underline{\underline{R}}^{-1}.\underline{P}$ and, being the reference's power equal to one and using the expression of the Wiener filter,

$$\xi_{\min} = 1 - \underline{P}^H.\underline{A} = 1 - \frac{a^2.\left(a^2+\sigma^2\right) + 1 + \sigma^2}{\left[\left(1+a^2+\sigma^2\right)^2 - a^2\right]}$$

d.- The denominator of the step size is the trace of the data autocorrelation matrix and the numerator the miss-adjusment, in consequence:

$\mu = \dfrac{0.1}{2.\left(1+a^2+\sigma^2\right)}$ At the same time since the number of iterations for convergence is

$nc = -\dfrac{Ln(10)}{Ln\left(1-\mu.\lambda_{\min}\right)}$ small values of the miss-adjusment implies that the number of iteration for convergence is bounded by:

$nc \cong Ln(10).\dfrac{\lambda_{\max}}{\lambda_{\min}} < Ln(10).\dfrac{Trace\left(\underline{\underline{R}}\right)}{\lambda_{\min}} = Ln(10).\dfrac{2.(1+a^2+\sigma^2)}{\left(1+\sigma^2\right)}$ Note that the number is the closest integer to the above expression.

e.- Since z(n)=y(n)+b.d(n-1)=d(n)+a.d(n-1)+w(n)+b.d(n-1), or z(n)=(d(n)+w(n))+ (a+b).d(n-1) then the error would be:

e(n)=z(n)-d(n)=w(n)+(a+b).d(n-1)

Clearly, being w(n) y d(n) uncorrelated, the solution in order to minimize the power of the error is obtained when the second coefficient will be zero, i.e. b=-a.

In can be seen that this second solution is a system that perfectly inverts with an IIR the FIR channel. This solution requires that the value of a has to be strictly less than one otherwise the system will be non stable, precluding that the detector may obtain the desired signal at its output. In fact, the major problem of the filter stays on that requires that the detector provides a perfectly regenerated reference. This implies that any time z(n) has to be as close as possible to d(n), i.e. the signal to noise ratio has to be high (or moderate high). Under these two circumstances, a lower than one and good SNR the second system performs better than the first. Nevertheless, in general the IIR solution is usually disregarded in practical systems.

V.8.1.-

$$\mu = \frac{M}{Trace(\underline{\underline{R}})} = \frac{0.1}{M.P_x}$$

$$P_x = \beta.P_x + (1-\beta).\frac{\underline{X}_n^H.\underline{X}_n}{M} \quad with \quad \beta \geq 0.99 \quad equivalent\ to\ more\ than\ 100\ samples$$

c) $n_c = \dfrac{2.3}{Ln(1-\mu.\lambda_{\min})} = \dfrac{2.3}{Ln\left(1 - \dfrac{0.1.\sigma^2}{Trace(\underline{\underline{R}})}\right)}$

d)

Since $M = trace(\underline{\underline{R}}.\underline{\underline{\Sigma}})$ and $\underline{\underline{\Sigma}} = diagonal\left(\dfrac{A_{\max}^2}{3.2^b}\right)$ then $M_{quant.} = \dfrac{A_{\max}^2.Trace(\underline{\underline{R}})}{3.2^b}$

V.8.2.-

a) The eigenvalue spread is given by the ex-centric are the plots. Approximately the length of the major with respect the minor axis in the plot is 0.8.

b) The spreading around the optimum seems to be 5 times larger in the second case than in the first. This is the same for the convergence rate. In consequence μ in the second case looks 5 times greater than in the first case.

c) The same reasoning that it was done in (b).

V.8.3.-

a)

$\underline{a}^{Q+1} = \begin{bmatrix} \underline{a}^Q \\ 0 \end{bmatrix} + k_Q.\begin{bmatrix} 0 \\ \underline{a}_r^Q \end{bmatrix}$ where "r" indicates reverse order: Start with $\underline{a}_0 = [1]$

b)

$k_q(n) = k_q(n) - \mu.\left(\gamma.e_b^{Q-1}(n-1).e_f^Q(n) + (1-\gamma).e_f^{Q-1}(n).e_b^Q(n)\right)$ being the objective

$$\xi^Q = \gamma.E\left[\left(e_f^Q(n)\right)^2\right] + (1-\gamma).E\left[\left(e_b^Q(n)\right)^2\right]$$

c) The data used in this gradient lattice is $data = \begin{bmatrix} e_b^{Q-1}(n) \\ e_f^{Q-1}(n) \end{bmatrix}$ and its trace is $F^{Q-1} + B^{Q-1}$. Then the steep size will be: $\mu = \dfrac{0.1}{F^{Q-1} + B^{Q-1}}$

d) It is obvious that the steep size will grow since the denominator at each section are the prediction errors for successive orders that, in general will strictly decrease. Only they remain the same when the adequate order of an AR process is over-passed.

d) Already done in the previous sections. The trade-off parameter was γ .

V.8.4.-

a) Check the chapter content to prove this formula

$$\xi = \xi_{MIM} + \left(\underline{A}_n - \underline{A}_{opt}\right)^H . \underline{\underline{R}} . \left(\underline{A}_n - \underline{A}_{opt}\right)$$

b) Defining $\tilde{\underline{A}}_n = \left(\underline{A}_n - \underline{A}_{opt}\right)$ and $\underline{\underline{\Sigma}} = E\left[\left(\underline{A}_n - \underline{A}_{opt}\right)\left(\underline{A}_n - \underline{A}_{opt}\right)^H\right]$ because the commutative property of the trace

$$E[\xi] = \xi_{MIM} + E\left[\left(\underline{A}_n - \underline{A}_{opt}\right)^H . \underline{\underline{R}} . \left(\underline{A}_n - \underline{A}_{opt}\right)\right] =$$

$$= \xi_{MIM} + Trace\left(E\left[\left(\underline{A}_n - \underline{A}_{opt}\right)^H . \underline{\underline{R}} . \left(\underline{A}_n - \underline{A}_{opt}\right)\right]\right) =$$

$$= \xi_{MIM} + Trace\left[\underline{\underline{R}} . E\left[\left(\underline{A}_n - \underline{A}_{opt}\right)\left(\underline{A}_n - \underline{A}_{opt}\right)^H\right]\right] =$$

$$= \xi_{MIM} + Trace\left[\underline{\underline{R}} . E\left(\tilde{\underline{A}} . \tilde{\underline{A}}^H\right)\right] = \xi_{MIM} + Trace\left[\underline{\underline{R}} . \underline{\underline{\Sigma}}\right]$$

c)

$$\underline{A}_{n+1} = \underline{A}_n + \mu . \underline{X}_n . (d(n) - y(n))$$

Subtracting the optimum in both terms

$$\tilde{\underline{A}}_{n+1} = \tilde{\underline{A}}_n + \mu . \underline{X}_n . (d(n) - y(n))$$

Computing the covariance

$\Sigma_{n+1} = \Sigma_n + \mu^2 . E\left[\varepsilon^2(n)\right] E\left[\underline{X}_n . \underline{X}_n^H\right] - 2.\mu . E\left[\underline{X}_n . \varepsilon(n) . \tilde{\underline{A}}_n^H\right]$ where we have assumed that the error is orthogonal to the data since we are close to convergence (note that we are dealing with miss adjustment)

Because, being close of convergence, $\varepsilon(n) \approx \underline{X}_n^H . \tilde{\underline{A}}_n + w(n)$ then $E\left[\underline{X}_n . \varepsilon(n) . \tilde{\underline{A}}_n^H\right] \approx \underline{\underline{R}} . \underline{\underline{\Sigma}}_n$ and $E\left[\varepsilon^2(n)\right] \approx \xi_{\min}$.

In summary: $\Sigma_{n+1} = \Sigma_n + \mu^2 . \xi_{\min} . \underline{\underline{R}} - 2.\mu . \underline{\underline{R}} . \underline{\underline{\Sigma}}_n$ and forcing that $\underline{\underline{\Sigma}}_{n+1} = \underline{\underline{\Sigma}}_n = \Sigma \quad \forall n \geq n_{convergence}$

$$\Sigma = \Sigma + \mu^2 . \xi_{\min} . \underline{\underline{R}} - 2.\mu . \underline{\underline{R}} . \underline{\underline{\Sigma}} \quad \Rightarrow \quad \mu^2 . \xi_{\min} . \underline{\underline{R}} = 2.\mu . \underline{\underline{R}} . \underline{\underline{\Sigma}}$$

and $\underline{\underline{\Sigma}} = \frac{\mu}{2} . \xi_{\min} . \underline{\underline{I}}$

d)

From the definition of the missadjusment $M = \frac{E[\xi_n] - \xi_{\min}}{\xi_{\min}} = Trace\left(\underline{\underline{R}} . \underline{\underline{\Sigma}}\right) = \frac{\mu}{2} . Trace\left(\underline{\underline{R}}\right)$

e)

Done in the previous exercise

f)

$$SNR_{opt} = \frac{P_d}{\xi_{\min}} \quad and \quad SNR = \frac{P_d}{E[\xi_n]}$$

Since *then*

$$SNR = \frac{SNR_{opt}}{\left(E[\xi_n] / \xi_{\min} \right)} = \frac{SNR_{opt}}{M+1}$$

V.8.5.-

a.-

$$\underline{P} = E\left\{ \underline{y}_n . x(n-1) \right\} \quad and \quad \underline{\underline{R}} = E\left\{ \underline{y}_n . \underline{y}_n^H \right\}$$

$$\underline{h} = \underline{\underline{R}}^{-1} . \underline{P}$$

$$\left(y(n) \quad .... \quad y(n-M+1) \right) = \underline{\underline{C}}^H . \underline{\underline{X}}_n \quad ; \quad \underline{\underline{X}}_n = \begin{bmatrix} x(n) & x(n-1) & ... \\ x(n-1) & x(n-2) & ... \end{bmatrix}$$

$1xM$ $\quad$ $1x2 \quad 2xM$

or

$$\begin{bmatrix} y(n) \\ .. \\ y(n-m+1) \end{bmatrix} = \begin{bmatrix} c(1) & c(2) & 0 & .. & 0 & 0 \\ 0 & c(1) & c(2) & .. & 0 & 0 \\ .. & .. & .. & .. & c(1) & c(2) \end{bmatrix} . \begin{bmatrix} x(n) \\ x(n-1) \\ .. \\ x(n-M) \end{bmatrix}$$

$$\underline{y}_n = \underline{\underline{C}} . \underline{X}_n$$

Thus, the filter design equations are:

$$\underline{P} = \underline{\underline{C}} . E\left(\underline{X}_n . x(n-1) \right) = \underline{\underline{C}} . \begin{bmatrix} 0 \\ 1 \\ 0 \\ .. \\ 0 \end{bmatrix} = \begin{bmatrix} c(2) \\ c(1) \\ 0 \\ .. \end{bmatrix}$$ Vector of M components.

$$\underline{\underline{R}} = \underline{\underline{C}} . \underline{\underline{R}}_{xx} . \underline{\underline{C}}^H + N_0 . \underline{\underline{I}} = \underline{\underline{C}} . \underline{\underline{C}}^H + N_0 . \underline{\underline{I}}$$

b.- For M=2 we have:

$$\underline{\underline{R}} = \begin{bmatrix} c(1)^2 + c(2)^2 & c(1).c(2) \\ c(1).c(2) & c(1)^2 + c(2)^2 \end{bmatrix} + N_0 . \underline{\underline{I}} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} + N_0 . \underline{\underline{I}}$$

$$\underline{\underline{R}}^{-1} = \frac{1}{(1+N_0)^2 - 0.25} . \begin{bmatrix} 1+N_0 & -0.5 \\ -0.5 & 1+N_0 \end{bmatrix}$$

$$\underline{h} = \frac{1/\sqrt{2}}{(1+N_0)^2 - 0.25} . \begin{bmatrix} 0.5+N_0 \\ 0.5+N_0 \end{bmatrix}$$

c.- $\xi = P_x - \underline{P}^H.\underline{\underline{R}}^{-1}.\underline{P} = P_x - \underline{P}^H.\underline{h} = 1 - \frac{1+2.N_o}{1.5+4.N_0+2.N_0^2} = \frac{0.5+2.N_o+2.N_0^2}{1.5+4.N_0+2.N_0^2}$

d.- $\underline{h}_n = \underline{h}_{n-1} + \mu.\varepsilon(n).\underline{y}_n$

e.- $\mu \le \frac{1}{\lambda_{\max}} \le \frac{2}{Tr\left(\underline{\underline{R}}\right)} = \frac{2}{2.N_0 + 2.\left(c(1)^2 + c(2)^2\right)} = \frac{1}{N_0+1}$

f.- $n_{con} = k_0.\frac{1}{\mu.\lambda_{\min}} \approx \frac{N_0+1}{N_0}.k_0 = k_0.\left(1+\frac{1}{N_0}\right)$

g.-

$$M = \alpha\,\% \quad when \quad \mu = \frac{2.\alpha}{Tr\left(\underline{\underline{R}}\right)}$$

$$if \quad M = 0.1 \quad then \quad \alpha = 0.1$$

for 0.01 criteria of convergence

VI.8.1.-

a.- Since $\underline{\underline{A}}^H.\underline{\underline{A}} = 4.\underline{\underline{I}}$ then $\underline{\underline{B}} = 0.5.\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}$

b.- $\underline{\underline{C}}_x = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ the eigenvalues and eigenvectors are

$$\lambda_1 = 1+\rho \quad \underline{e}_1 = \frac{1}{\sqrt{2}}.\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 1-\rho \quad \underline{e}_2 = \frac{1}{\sqrt{2}}.\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

thus $\underline{\underline{\Phi}} = \frac{1}{\sqrt{2}}.\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

c.-

$$\underline{z} = \underline{\underline{B}}.\underline{x} \qquad \underline{z}^Q = \underline{z} + \underline{\varepsilon} = \underline{\underline{B}}.\underline{x} + \underline{\varepsilon} \qquad \underline{\hat{x}} = \underline{\underline{B}}^H.\underline{z}^Q = \underline{x} + \underline{\underline{B}}^H.\underline{\varepsilon}$$

$$\underline{\varsigma} = \underline{x} - \underline{\hat{x}} = \underline{\underline{B}}^H.\underline{\varepsilon} \qquad \underline{\varsigma}^H.\underline{\varsigma} = \underline{\varepsilon}^H.\underline{\underline{B}}.\underline{\underline{B}}^H.\underline{\varepsilon} = Tr\left(\underline{\underline{B}}^H.\underline{\varepsilon}.\underline{\varepsilon}^H.\underline{\underline{B}}\right)$$

taking the expected value of the last expression we obtain the error power E

$$MSE = Tr\left(\underline{\underline{B}}^H.E\left\{\underline{\varepsilon}.\underline{\varepsilon}^H\right\}.\underline{\underline{B}}\right) = Tr\left(\underline{\underline{B}}^H.\underline{\underline{E}}.\underline{\underline{B}}\right) = Tr\left(\underline{\underline{E}}.\underline{\underline{B}}^H.\underline{\underline{B}}\right) = Tr\left(\underline{\underline{E}}\right)$$

Now, $\underline{\underline{E}} = diag(\underline{\underline{R}}_z).\begin{bmatrix} 1/2^{2.k1} & 0 \\ 0 & 1/2^{2.k2} \end{bmatrix}$ since the quantization error is the power of the signal to be quantized, reflected in the diagonal of the acf matrix, divided by two raised to two times the number of bits used.

$$being \quad \underline{\underline{R}}_z = \underline{\underline{B}}^H.\underline{\underline{R}}_x.\underline{\underline{B}}$$

Case of the proposed transform:

$$0.25.\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}.\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix} = \begin{bmatrix} \frac{4+2.\sqrt{3}.\rho}{4}.2^{-2.k1} & xx \\ xx & \frac{4-2.\sqrt{3}.\rho}{4}.2^{-2.k2} \end{bmatrix}$$

and the error is: $MSE(PT) = \frac{4+2.\sqrt{3}.\rho}{4}.2^{-2.k1} + \frac{4-2.\sqrt{3}.\rho}{4}.2^{-2.k2}$

Case of the KL:

$$\underline{\underline{\Phi}}^H.\underline{\underline{R}}.\underline{\underline{\Phi}} = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}$$

$$MSE(KL) = (1+\rho).2^{-2.k1} + (1-\rho).2^{-2.k2}$$

When k2=0

$$MSE^{KL} = (1+\rho).2^{-2.k}$$

$$MSE^{proposed} = \left(\frac{4+2.\sqrt{3}.\rho}{4}\right).2^{-2.k} = \left(1+\frac{\sqrt{3}}{2}.\rho\right).2^{-2k}$$

$d.-$

$$MSE^{direct} = 2.2^{-2.k}$$

e.- When k2=0 then:

$$2.k1^{KL} = \log_2\left(\frac{1+\rho}{D}\right)$$

$$2.k1^{proposed} = \log_2\left(\frac{4+2.\sqrt{3}.\rho}{4.D}\right)$$

$e.-$

$$2.k^{direct} = \log_2\left(\frac{2}{D}\right)$$

f.- Obvious.

VI8.3.-

a)The given expression $\underline{\underline{X}}_n = \sum_{r=1}^{Q} \varphi_n(r).\underline{u}_{r,n}.\underline{v}_{r,n}^H$, can be written as $\underline{\underline{X}}_n = \underline{\underline{U}}_n.diag\underline{\underline{\varphi}}_n.\underline{\underline{V}}_n^H$ which indicates that matrixes $\underline{\underline{U}}_n$ y $\underline{\underline{V}}_n$ diagonalize directly the given matrix. For this reason for every sub-image it is necessary to transmit both matrices, which implies a severe waste of channel capacity.

b) In this case $\underline{\underline{X}}_n = \sum_{r=1}^{Q}\sum_{s=1}^{Q} \phi_n(r,s).\underline{u}_r.\underline{u}_s^H$ or $\underline{\underline{X}}_n = \underline{\underline{U}}.\underline{\underline{\phi}}_n.\underline{\underline{U}}^H$ which does not diagonalize the original matrix, it is just a mere transformation. At the same time, this transform does not depend on every sub-image under processing and can be computed as an average over an ensemble of sub-images.

In any case, it is important to set additional criteria in order to select a transform in such a way that the transmission of $\underline{\underline{\phi}}_n$ reports more advantages than the direct transmission of the original image. Otherwise the transform would be un-useful.

c) Since the determinant of a definite positive matrix $\underline{\underline{R}}_x$, which trace is fixed since the elements of the main diagonal are the energy of every component, is always lower than the product of its main diagonal components, then B is maximum when the off-diagonal of the transformed matrix are zero. In other words, the components are uncorrelated.

d) Since the power is $P_y = Trace\left(\underline{\underline{R}}_y = E\left[\underline{Y}_n.\underline{Y}_n^H\right]\right) = Trace\left(\underline{\underline{U}}.\underline{\underline{R}}_x.\underline{\underline{U}}^H\right) =$ The trace has the circular property, i.e. altering the order does not change the trace=$Traza\left(\underline{\underline{R}}_x.\underline{\underline{U}}^H.\underline{\underline{U}}\right)$.

In order that this last expression be equal to $P_x = Traza\left(\underline{\underline{R}}\right)$ is clear that the transform matrix must be orthonormal. $\underline{\underline{U}}^H.\underline{\underline{U}} = \underline{\underline{I}}$

e) The orthonormality constrain together with the un-correlation of the components implies that the transform matrix has to be set equal to the eigenvectors of the original correlation matrix.

$\underline{\underline{R}}.\underline{\underline{U}} = \underline{\underline{U}}.diag\left[\lambda\right]$ and $\underline{\underline{R}}_y = diag\left[\lambda\right]$

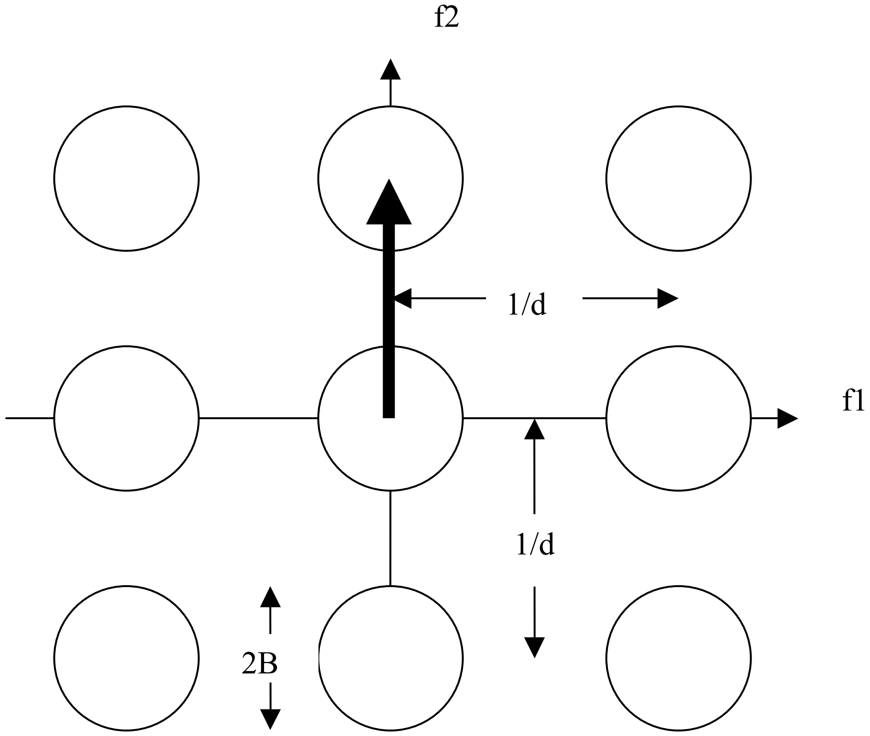
Clearly B is preserved (it does not increases) since the determinant of the resulting or transformed correlation matrix is just the product of the eigenvalues which is the same that the determinant of the original matrix.

f) The DFT is also an orthonormal transform. The difference is that the autocorrelation matrix of the transformed data is not diagonal. In other words, assuming that the transform is the set of outputs of a filter bank, the DFT does not use orthonormal filters they have aliasing precluding the un-correlation of their outputs. Only when the size of the signal tends to infinity then the filters become ideal and their outputs uncorrelated. Nevertheless, for reasonable sample size the DFT may show good decorrelation among its components.

Any non-linear processing tends to spread the energy in the frequency domain. Since the DCT implies an additional extrapolation making even the original signal this provides better result that the zero-padding associated with the DFT. The energy leakage due to non-linearities is always from frequency regions of similar energy levels. In addition, DCT implies only real operations.

V.18.4.

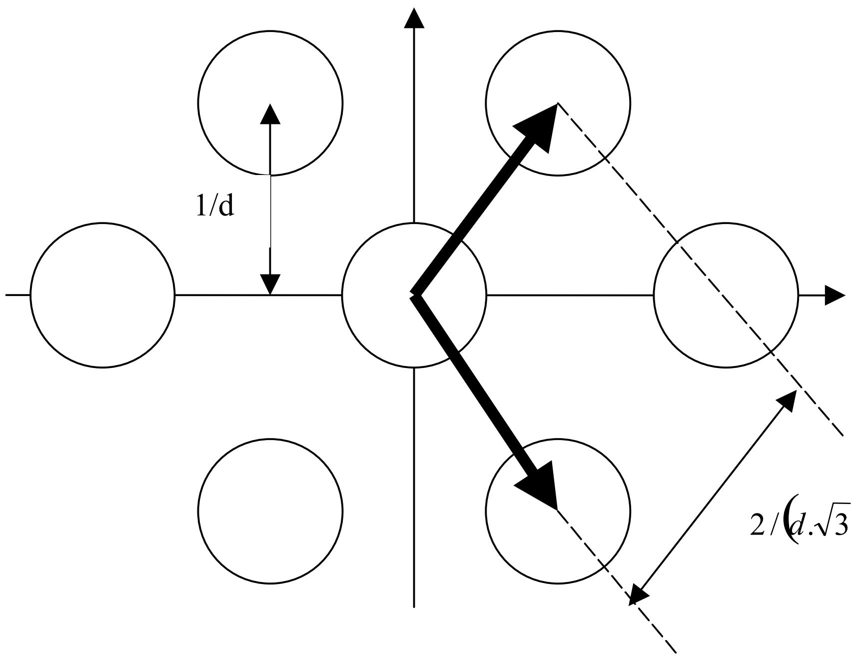
a.- UIT a rectangular sampling matrix $\underline{\underline{U}}_R = d.\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, the repetition matrix in the frequency plane would be $\underline{\underline{V}}_R^T.\underline{\underline{U}} = 2\pi\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\underline{\underline{V}}_R = \frac{2\pi}{d}.\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. With this matrix the spectrum plane is:



To prevent aliasing it is necessary that $d \leq 2B$

b.- Doing the same that in the previous section,

$\underline{\underline{V}}_H^T \underline{\underline{U}}_H = \underline{\underline{V}}_H^T d.\begin{bmatrix} \sqrt{3}/2 & \sqrt{3}/2 \\ 1/2 & -1/2 \end{bmatrix} = 2\pi.\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\underline{\underline{V}}_H^T = \frac{2\pi}{d}.\frac{2}{\sqrt{3}}.\begin{bmatrix} 1/2 & \sqrt{3}/2 \\ 1/2 & -\sqrt{3}/2 \end{bmatrix}$ in consequence: $\underline{\underline{V}}_H^T = \frac{2\pi}{d}.\frac{2}{\sqrt{3}}.\begin{bmatrix} 1/2 & 1/2 \\ \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$. The spectrum plane will be:



Now the closest spectrum are located at a distance equal to $\frac{1}{d.\sqrt{3}}$ and the constrain for no-aliasing is $d \le (2B).\sqrt{3}$

c.- The hexagonal sampling is better since it requires a longer distance in sampling (i.e. the sampling frequency yet preserving the anti-alisaing condition is lower for hexagonal than for rectangular sampling)

d.- The inverse Fourier transform for the discrete signal is:

$$f(\underline{n}) = cte.\int_{\underline{w}} F_d(\underline{w}).\exp(j\underline{n}^T.\underline{\underline{U}}_R.\underline{w}).d\underline{w}$$

Sampling in the $\underline{w}$ plane with $2\pi\underline{\underline{\Phi}}.\underline{m}$ we obtain:

$$f_d(\underline{n}) = cte.\sum_{\underline{m}} F_d(2\pi\underline{\underline{\Phi}}.\underline{m}).\exp(2\pi j\underline{n}^T.\underline{\underline{U}}_R^T.\underline{\underline{\Phi}}\underline{m}).d\underline{w}$$

This is a periodic function in the spatial domain and it has to show repetition when passing from $\underline{n}$ to $\underline{n}+\underline{\underline{N}}.\underline{l}$. Clearly, to avoid aliasing, matrix $\underline{\underline{N}}$ must be diagonal (rectangular case) with entries above M where M.d>2D. Thus

$$\underline{\underline{N}} = \begin{bmatrix} M & 0 \\ 0 & M \end{bmatrix}$$

Finally, the existing relationship between the sampling matriz and the sampling matrix at the frequency domain is derived from setting factor $\underline{\underline{N}}^T.\underline{\underline{U}}_R.\underline{\underline{\Phi}}$ equal to the identity matrix. In consequence:

$$2\pi\underline{\underline{\Phi}} = \underline{\underline{V}}.\underline{\underline{N}}^{-T}$$

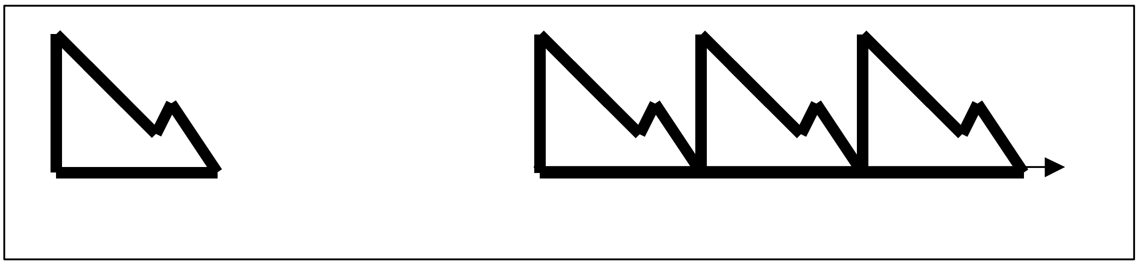
And, because $f_d(\underline{n})$ is the basic function repeated without aliasing, we can write the following expression:

$$f(\underline{n}) = cte.\sum_{\underline{m}} F_d(\underline{m}).\exp(2\pi j\underline{n}^T.\underline{\underline{N}}^{-T}.\underline{m}).d\underline{w}$$

e.- The reasons for using the DCT instead of the DFT in image processing are all of them based on the real and positive character of the signal at the spatial domain. The major differences among both transforms are:
1.- Always use real numbers (no complex quantities).
2.- Twice resolution of the DCT with respect the DFT of the same size of data signal.

As a consequence of this second property, the repetition is performed over the even part of the original signal, this implies than any non-linear processing, like quantification included in all the standards for image coding, the leakage of energy motivated by the non-linear processing in the borders of the data segment does not occurs on different border (left or right in the figure. This property is essential in reducing distortion on coding and compression procedures.

Periodicity on DFT.



Periodicity on DCT for the same record length