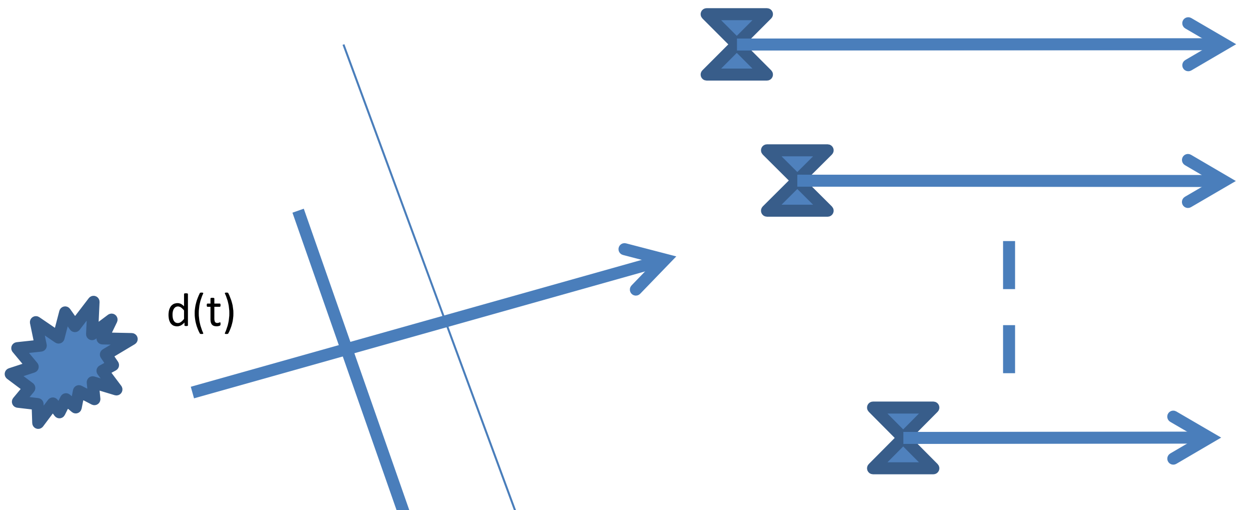
ESTIMACION Y FILTRADO

Procesado de Señal
Miguel A. Lagunas

Noción de estimation o filtrado



$$\underline{X}_t = d(t).\underline{H} + \underline{w}_n$$

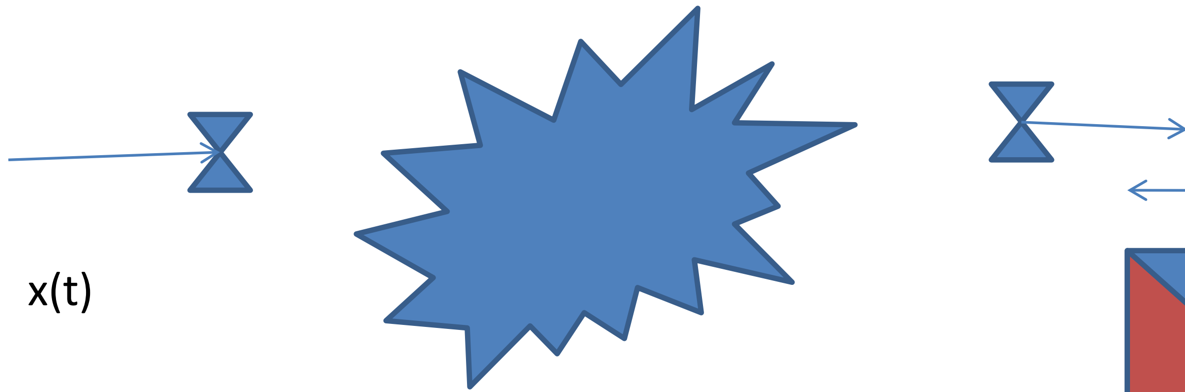
Dado Xt se pretende conocer d(t), H y la potencia de ruido.

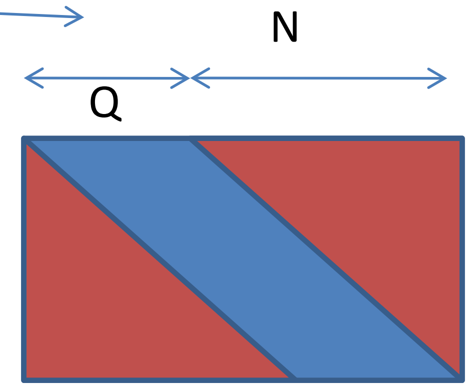
SE CONOCE:

- El modelo
- El ruido es Gaussiano de media nula y covarianza Ro

$$E\left(\underline{w}_n\right) = \underline{0}$$

$$E\left(\underline{w}_n \underline{w}_n^H\right) = \underline{\underline{R}}_0$$



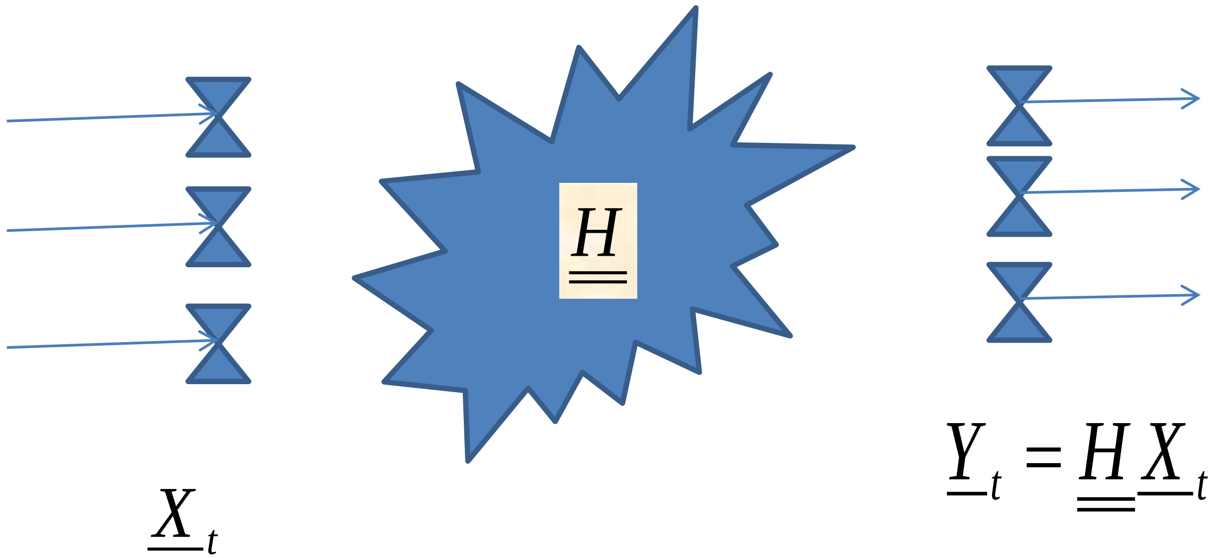


$$y(t)=\sum_{q=0}^{Q-1}h(q)x(t-qT)+w(t)$$

$$\begin{bmatrix} y(t) \\ .. \\ y(t-NT) \end{bmatrix} = \begin{bmatrix} \underline{h}^T & 0 & .. & 0 \\ 0 & \underline{h}^T & .. & 0 \\ .. & .. & .. & .. \\ 0 & 0 & 0 & \underline{h}^T \end{bmatrix} \begin{bmatrix} x(t) \\ x(t-T) \\ .. \\ x(t-(N+Q-1)T \end{bmatrix} + \begin{bmatrix} w(t) \\ .. \\ w(t-NT) \end{bmatrix}$$

De nuevo el objetivo es estimar x(t) dado h, el modelo y que el ruido es gaussiano de media nula y covarianza conocidda.

FILTRADO DE Yt PARA OBTENER (ESTIMAR) Xt



$$\underline{Y}_t = \underline{\underline{H}}\,\underline{X}_t + \underline{w}_t$$

Note que el problema anterior es un caso particular de este problema que consiste en estimar la señal x(t) cuando se conoce y(t), el modelo y el ruido (gaussiano,media y covarianza)

Otra situacion es: Dado x(t), y(t) el modelo y caracterizacion del ruido, estimar h.

IDENTIFICACION DE SISTEMAS

$$y(t)=\sum_{q=0}^{Q-1}h(q)x(t-qT)+w(t)$$

$$\begin{bmatrix} y(t) \\ .. \\ y(t-NT) \end{bmatrix} = \begin{bmatrix} \underline{X}_t^T \\ \ldots \\ \underline{X}_{t-NT}^T \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ .. \\ h(Q-1) \end{bmatrix} + \begin{bmatrix} w(t) \\ .. \\ w(t-NT) \end{bmatrix}$$

$$\underline{Y}_t = \underline{\underline{X}}_t \underline{h} + \underline{w}_t$$

ESTIMAR: ¿QUE CRITERIO USAR?

INTUICION:
El mejor estimador de una cantidad es su valor mas probable

INFORMACION: Cuando ocurre lo menos probable es la maxima informacion

Es decir, estimar bien es seleccionar el valor que mas ocurre, el que menos informacion da, es como tomar la “moda” o valor medio.

Asi pues, conocida la distribucion de probabilidad del vector a estimar, el mejor estimador sera:

$$\underline{\hat{a}} = \max_{\underline{a}} \Pr(\underline{a})$$

Informacion y entropia o informacion media:

$$I(\underline{a}) = -Ln(\Pr(\underline{a}))$$

$$H = \int I(\underline{a}) \Pr(\underline{a}) d\underline{a} = -\int Ln(\Pr(\underline{a})) \Pr(\underline{a}) d\underline{a}$$

Resumen

- El mejor estimador es el valor mas probable.
- El valor mas probable es el de minima informacion.
- Entropia es la informacion asociada a un evento con una determinada pdf.
- La entropia maxima es cuando todos los valores dan la misma informacion (UNIFORME). Este es el caso de mas indecision a la hora de estimar el valor que tomara la variable.

ESTIMACION COHERENTE CON LOS DATOS

Cuando existe un vector de datos, la intuicion de nuevo nos dice que la estimacion mas correcta seria el valor mas probable, observados los datos.
EL ESTIMADOR OPTIMO, conocido como MAP (Maximo A Posteriori) seria:

$$\hat{\underline{a}}_{MAP} = \max_{\underline{a}} \Pr\left(\underline{a} \big/ \underline{X}_t \right)$$

Para conocer un poco mas la estructura de este estimador, usaremos el teorema de Bayes

$$\Pr\left(\underline{a} \big/ \underline{X}_t \right) \propto \Pr\left(\underline{X}_t \big/ \underline{a} \right) \Pr(\underline{a})$$

-Notese que los datos disminuyen la informacion aun mas (Mejoran la estimacion)
-La funcion que se maximiza es el producto (suma en logaritmos) de la VEROSIMILITUD de los datos con respecto al vector a estimar por (mas) el conocimiento a PRIORI del vector a estimar.

EL ESTIMADOR DE MEDIA CONDICIONAL

Cuando la probabilidad es unimodal y simetrica (Este es el caso de distribucion Gaussiana), el maximo de la probabilidad coincide con la MEDIA. De este modo el estimador MAP coincidira con el estimador MED

$$\hat{\underline{a}}_{MAP} = \max_{\underline{a}} \Pr\left({\underline{a}}/{\underline{X}_t} \right) =$$

$$= \hat{\underline{a}}_{MED} = E\left[\Pr\left({\underline{a}}/{\underline{X}_t} \right) \right] = \int \underline{a} \Pr\left({\underline{a}}/{\underline{X}_t} \right) d\underline{a}$$

EL ESTIMADOR MSE O DE MINIMO ERROR CUADRATICO MEDIO

Un nuevo estimador, aun sin saber si coincide con el optimo MAP, seria del valor que mejor se "ajusta" a la distribucion, en terminos de error cuadratico. Su definicion seria:

Derivando con respecto al vector c conjugado

$$\nabla_{\underline{c}^H} \int (\underline{c}-\underline{a})^H (\underline{c}-\underline{a}) \Pr\left(\underline{a} / \underline{X}_t \right) d\underline{a} = \int (\underline{c}-\underline{a}) \Pr\left(\underline{a} / \underline{X}_t \right) d\underline{a} = \underline{0}$$

$$\int \underline{c} \Pr\left(\underline{a} / \underline{X}_t \right) d\underline{a} = \underline{c} \int \Pr\left(\underline{a} / \underline{X}_t \right) d\underline{a} = \underline{c} = \int \underline{a} \Pr\left(\underline{a} / \underline{X}_t \right) d\underline{a}$$

EL ESTIMADOR MSE COINCIDE CON EL DE MEDIA CONDICIONAL. AMBOS SON OPTIMOS SI LA PDF ES UNIMODAL Y SIMETRICA

$$\underline{\hat{a}}_{MSE} = \int \underline{a} \Pr\left(\underline{a} / \underline{X}_t \right) d\underline{a} = \underline{\hat{a}}_{MED}$$

Note que la denominada MATRIZ del MSE viene dada por:

$$\underline{\underline{E}} = \int \left(\underline{\hat{a}} - \underline{a}\right)\left(\underline{\hat{a}} - \underline{a}\right)^H \Pr\left(\underline{a} / \underline{X}_t \right) d\,\underline{a}$$

Como el operador traza es circular tr(AB)=tr(BA).
Y la traza de us escalar es el mismo tr(a)=a

-EL ESTIMADOR MSE MINIMIZA LA TRAZA DE LA MATRIZ DEL MSE.
- LA MATRIZ DEL MSE PASA A SER LA MINIMA QUE ES LA COVARIANZA DEL VECTOR A ESTIMAR.

EL ESTIMADOR DE MAXIMA VEROSIMILITUD O ML

Algunas veces es muy dificil conocer la probabilidad MAP $\Pr\left(\underline{a} / \underline{X}_t \right)$

$$\Lambda(\underline{a}) = Ln\left(\Pr\left(\underline{X}_t / \underline{a} \right) \right)$$

$$\log - likelihood$$

En su lugar, es facil de formular la probabilidad de los datos dado el parametro a estimar o VEROSIMILITUD

El estimador resultante de maximizar la verosimilitud se le denomina estimador ML o de maxima verosimilitud.
Este estimador coincide con el optimo cuando el conocimiento a priori es nulo (o lo que es lo mismo, uniforme).

$$\underline{\hat{a}}_{ML} = \max_{\underline{a}} \Lambda(\underline{a}) = \max_{\underline{a}} Ln\left(\Pr\left(\underline{X}_t / \underline{a} \right) \right)$$

Ejemplo: La Media de un Proceso

$$\underline{X}_t = a\underline{1} + \underline{w}_t$$

$$\underline{1}^T = \begin{bmatrix} 1 & .. & 1 \end{bmatrix}$$

$$\underline{X}_t = \begin{bmatrix} x(t) \\ x(t-T) \\ .. \\ x(t-(Q-1)T) \end{bmatrix}$$

Dado que:

$$E\left(\underline{X}_t\right) = a\underline{1}$$

$$Cov\left(\underline{X}_t\right) = E\left(\underline{w}_t \underline{w}_t^H\right) = \underline{\underline{R}}_0$$

La probabilidad de los datos dado el modelo sera:

$$\Pr\left(\underline{X}_t / a, \underline{1}, \underline{\underline{R}}_0\right) \propto \frac{1}{\det\left(\underline{\underline{R}}_0\right)} \exp\left[-\left(\underline{X}_t - a\underline{1}\right)^H \underline{\underline{R}}_0^{-1} \left(\underline{X}_t - a\underline{1}\right)\right]$$

Para maximizar, notese que solo el exponente depende del parametro que se desea estimar.
Al derivar el exponente con respecto al parametro conjugado, se obtiene:

$$\underline{1}^H \underline{\underline{R}}_0^{-1}\left(\underline{X}_t - a\underline{1}\right) = \underline{0}$$

De lo que se obtiene el estimador ML de la media de un proceso

$$\boxed{\underline{\hat{a}}_{ML} = \frac{\underline{1}^H \underline{\underline{R}}_0^{-1} \underline{X}_t}{\underline{1}^H \underline{\underline{R}}_0^{-1} \underline{1}}}$$

Notese que difiere de lo que seria la media de los datos

$$\underline{\hat{a}}_{am} = \frac{\underline{1}^H \underline{X}_t}{\underline{1}^H \underline{1}} = \frac{1}{Q}\sum_{q=0}^{Q-1} x(t-qT)$$

Solo coinciden cuando el ruido es blanco o incorrelado y su matriz de correlacion es una constante por la identidad

Ejemplo: Modem sincronizacion, una portadora en ruido

$$\underline{X}_t = b\underline{S} + \underline{w}_t = a\exp\left(jw_0 + \theta\right)\begin{bmatrix} 1 \\ \exp(-jw_0) \\ .. \\ \exp\left(-j(Q-1)w_0\right) \end{bmatrix}$$

$$\underline{S}^H = \left[1 \quad \exp(jw_0)\ldots\ldots \quad \exp(j(Q-1)w_0\right]$$

$$\underline{X}_t = \begin{bmatrix} x(t) \\ x(t-T) \\ .. \\ x(t-(Q-1)T) \end{bmatrix}$$

$$\Pr\left(\underline{X}_t / b, \underline{S}, \underline{\underline{R}}_0\right) \propto \frac{1}{\det\left(\underline{\underline{R}}_0\right)} \exp\left[-\left(\underline{X}_t - b\underline{S}\right)^H \underline{\underline{R}}_0^{-1}\left(\underline{X}_t - b\underline{S}\right)\right]$$

Derivando con respecto a b conjugado se obtiene: $\underline{S}^H \underline{\underline{R}}_0^{-1}\left(\underline{X}_t - b\underline{S}\right) = \underline{0}$

Y, el estimador de la amplitud compleja

$$b_{ML} = \frac{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{X}_t}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}}$$

Si el ruido es blanco $\underline{\underline{R}}_0 = \sigma^2 \underline{\underline{I}}$

El estimador de la amplitud se reduce a:

$$\boxed{b_{ML} = \frac{\underline{S}^H \underline{X}_t}{\underline{S}^H \underline{S}} = \frac{\underline{S}^H \underline{X}_t}{Q}}$$

La VEROSIMILITUD que dependia de magnitud b, frecuencia via el vector S y covarianza, al sustituir el estimador de la magnitud compleja solo depende de la frecuencia y de la potencia de ruido

$$\Pr\left(\underline{X}_t / \underline{S}, \underline{\underline{R}}_0\right) \propto \frac{1}{\left(\sigma^2\right)^Q} \exp\left[-\left(\underline{X}_t - \underline{S}\frac{\underline{S}^H \underline{X}_t}{Q}\right)^H \frac{1}{\sigma^2}\left(\underline{X}_t - \underline{S}\frac{\underline{S}^H \underline{X}_t}{Q}\right)\right]$$

Operando con el exponente

$$\left(\underline{X}_t - \underline{S}\,\frac{\underline{S}^H\,\underline{X}_t}{Q}\right)^H \frac{1}{\sigma^2}\left(\underline{X}_t - \underline{S}\,\frac{\underline{S}^H\,\underline{X}_t}{Q}\right) =$$

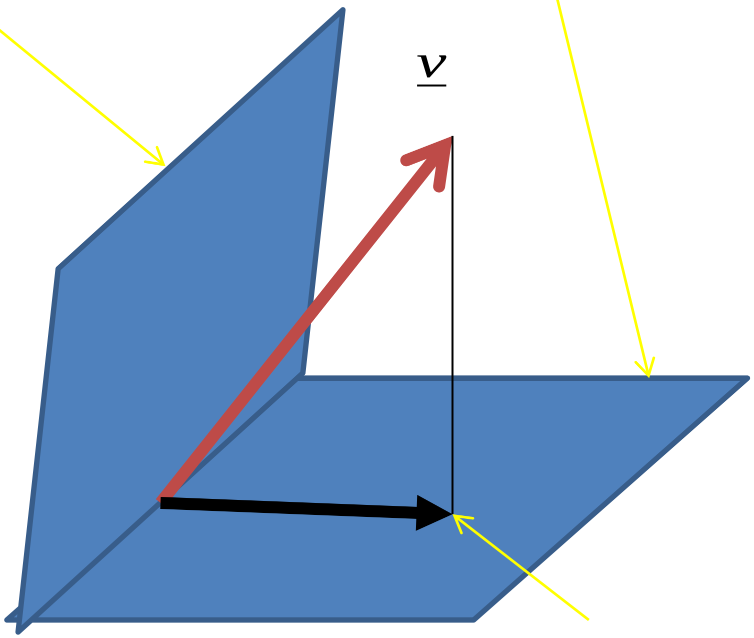
$$= \underline{X}_t^H\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q}\right)\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q}\right)\underline{X}_t\,\frac{1}{\sigma^2} = \underline{X}_t^H\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q}\right)\underline{X}_t\,\frac{1}{\sigma^2}$$

P es un operador de proyeccion.

$$\underline{\underline{P}} = \left(\underline{\underline{I}} - \underline{\underline{S}} \left(\underline{\underline{S}}^H \underline{\underline{S}} \right)^{-1} \underline{\underline{S}}^H \right)$$

Este operador proyecta en el espacio ortogonal al definido por la matriz S

$$\underline{\underline{P}}_{=} = \left(\underline{\underline{S}}^H \right)$$



Asi pues P proyecta cualquier vector en el espacio ortogonal al vector S. Es decir, es la proyeccion de los datos al resto de frecuencias excluyendo la que tiene el vector S

$$\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \right) \underline{X}_t$$

Como la traza de un escalar es el mismo y la traza es circular

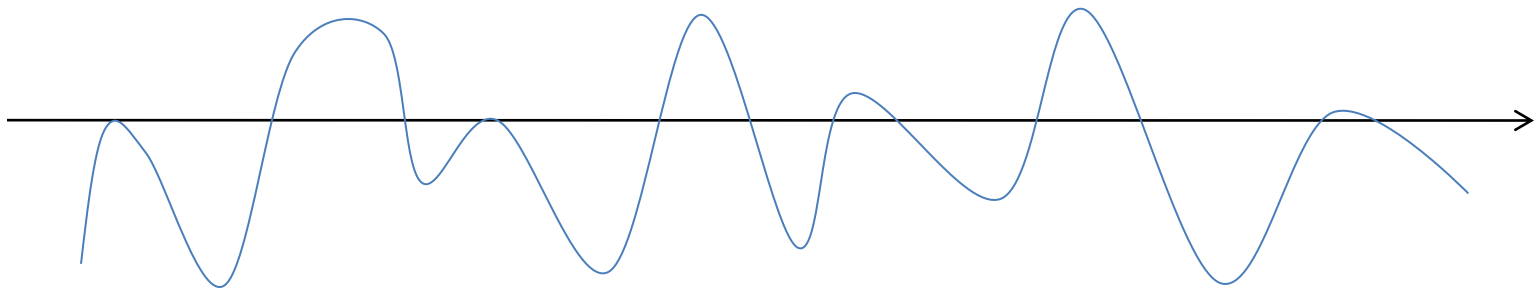
$$\frac{1}{\sigma^2} tr\left(\underline{X}_t^H \left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \right) \underline{X}_t \right) = \frac{1}{\sigma^2} tr\left(\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \right) \underline{X}_t \underline{X}_t^H \right)$$

$$tr\left(\underline{\underline{A}}\,\underline{\underline{B}}\right) = tr\left(\underline{\underline{B}}\,\underline{\underline{A}}\right)$$

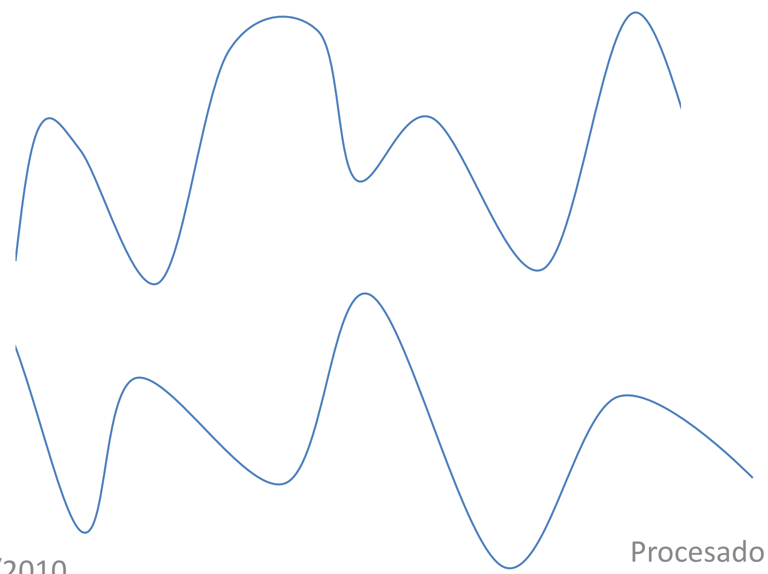
$$tr\left(\underline{a}^H \underline{b}\right) = escalar = \underline{a}^H \underline{b} = tr\left(\underline{b}\,\underline{a}^H\right)$$

ERGODICIDAD y SEGMENTACION

Una realizacion del proceso de N muestras



Aunque,… tal vez podrian ser dos realizaciones de N/2 muestras



Esta es la idea mas importante y practica de ergodicidad: Para tener mas realizaciones se puede segmentar cada realizacion disponible

Como la traza de un escalar es el mismo y la traza es circular

$$\frac{1}{\sigma^2} tr\left(\underline{X}_t^H \left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \right) \underline{X}_t \right) = \frac{1}{\sigma^2} tr\left(\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \right) \underline{X}_t \, \underline{X}_t^H \right)$$

La amplitud y la fase (CAG y sincronismo de fase, al cambiar rapidamente (radio) no se benefician de tomar segementos largos pues se necesita un estimador de ellas cada pocos simbolos. Sin embargo, el ruido del front-end y la frecuencia si que se benefician de tener varias realizaciones (segmentos) pues permanecen mas. Recuerde que no es que no se puedan hacer con un solo segemento sino que se estiman mejor cuantos mas se dispone. (NOCION DE ESTACIONARIDAD)

Si se dispone de mas segmentos de Q muestras independientes, digamos M segmentos, al ser independientes la verosimilitud sera el producto (al fin y al cabo es una probabilidad)

$$\prod_1^M \Pr\left(\underline{X}_t / \underline{S}, \underline{\underline{R}}_0 \right) \propto \left(\sigma^2 \right)^{-QM} \exp\left[-\left(\underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \right) \sum_1^M \underline{X}_t \, \underline{X}_t^H \right]$$

Tomando logaritmo neperiano (es monotona creciente y no cambia la posicion de extremos de la funcion) y denominando con la matriz P al operador de proyeccion, se obtiene:

$$Ln\left(\prod_{1}^{M} \Pr\left(\underline{X}_t / \underline{S}, \underline{\underline{R}}_0\right)\right) \propto -QMLn\left(\sigma^2\right) - \frac{1}{\sigma^2} tr\left(M \underline{\underline{P}}\,\underline{\underline{R}}\right)$$

$$siendo \quad \underline{\underline{P}} = \underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \quad y \quad \underline{\underline{R}} = \frac{1}{M} \sum_{1}^{M} \underline{X}_t \underline{X}_t^H$$

Ahora se puede derivar con respecto a la potencia de ruido para asi obtener su estimador ML

$$-QM \frac{1}{\sigma^2} + \frac{1}{\left(\sigma^2\right)^2} tr\left(M \underline{\underline{P}}\,\underline{\underline{R}}\right) = 0$$

De donde se obtiene el estimador ML de la potencia de ruido

$$\boxed{\sigma_{ML}^2 = tr\left(\underline{\underline{P}}\,\underline{\underline{R}}\right) / Q}$$

¿EN QUE INFLUYE M, EL NUMERO DE SEGMENTOS?

Al insertar la potencia de ruido estimada, la verosimilitud tan solo dependera de la frecuencia.

$$Ln\left(\prod_{1}^{M}\Pr\left(\underline{X}_t/\underline{S}\right)\right)\propto -QMLn\left(\sigma^2_{ML}\right)+Q$$

Como deseamos el maximo segun S (para estimar la frecuencia), implica que el S maximo ha de maximizar la potencia de ruido estimada. Como la potencia de ruido estimada es:

$$MIN_{\underline{S}}\,\sigma^2_{ML}=MIN_{\underline{S}}\,tr\left(\underline{\underline{P}}\,\underline{\underline{R}}\right)/Q=MIN_{\underline{S}}\,\frac{1}{Q}tr\left(\underline{\underline{R}}-\frac{\underline{S}\,\underline{S}^H\underline{\underline{R}}}{Q}\right)=$$

$$=MAX_{\underline{S}}\,tr\left(\frac{\underline{S}\,\underline{S}^H\underline{\underline{R}}}{Q}\right)=MAX_{\underline{S}}\,\frac{\underline{S}^H\underline{\underline{R}}\,\underline{S}}{Q}=MAX_{\underline{S}}\,\frac{1}{QM}\sum_{1}^{M}\left|\underline{S}^H\underline{X}_t\right|^2$$

$$\underline{S}_{ML}=MAX_{\underline{S}}\,\frac{1}{QM}\sum_{1}^{M}\left|\underline{S}^H\underline{X}_t\right|^2$$

En definitiva. La estimacion de frecuencia es el maximo del promedio del modulo de la DFT de cada segmento al cuadrado

$$\boxed{\omega_{0ML}=\max_w\frac{1}{QM}\sum_{1}^{M}\left|DFT\,\underline{X}_t\right|^2}$$

RESUMEN

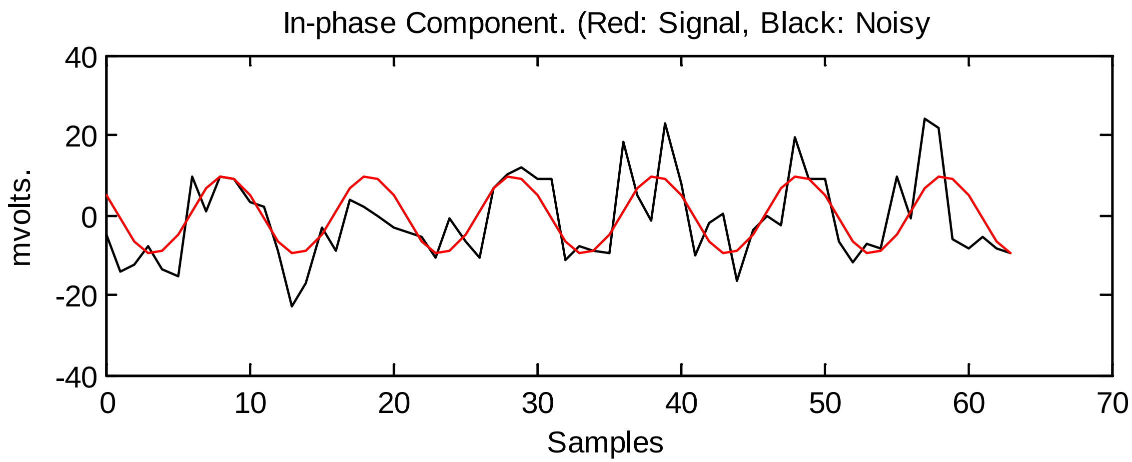
$$\omega_{0ML} = \max_w \frac{1}{QM} \sum_1^M \left| DFT\, \underline{X}_t \right|^2$$

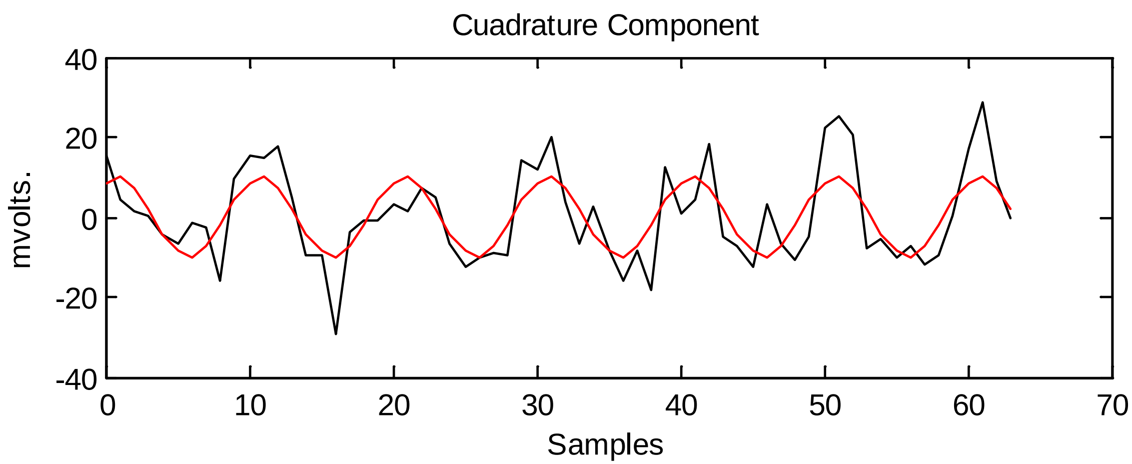
$$\underline{\underline{P}} = \underline{\underline{I}} - \frac{\underline{S}\,\underline{S}^H}{Q} \quad y \quad \underline{\underline{R}} = \frac{1}{M} \sum_1^M \underline{X}_t\, \underline{X}_t^H$$

$$\sigma_{ML}^2 = tr\left(\underline{\underline{P}}\,\underline{\underline{R}}\right) / Q$$

$$b_{ML} = \frac{\underline{S}^H\, \underline{X}_t}{\underline{S}^H\, \underline{S}} = \frac{\underline{S}^H\, \underline{X}_t}{Q}$$

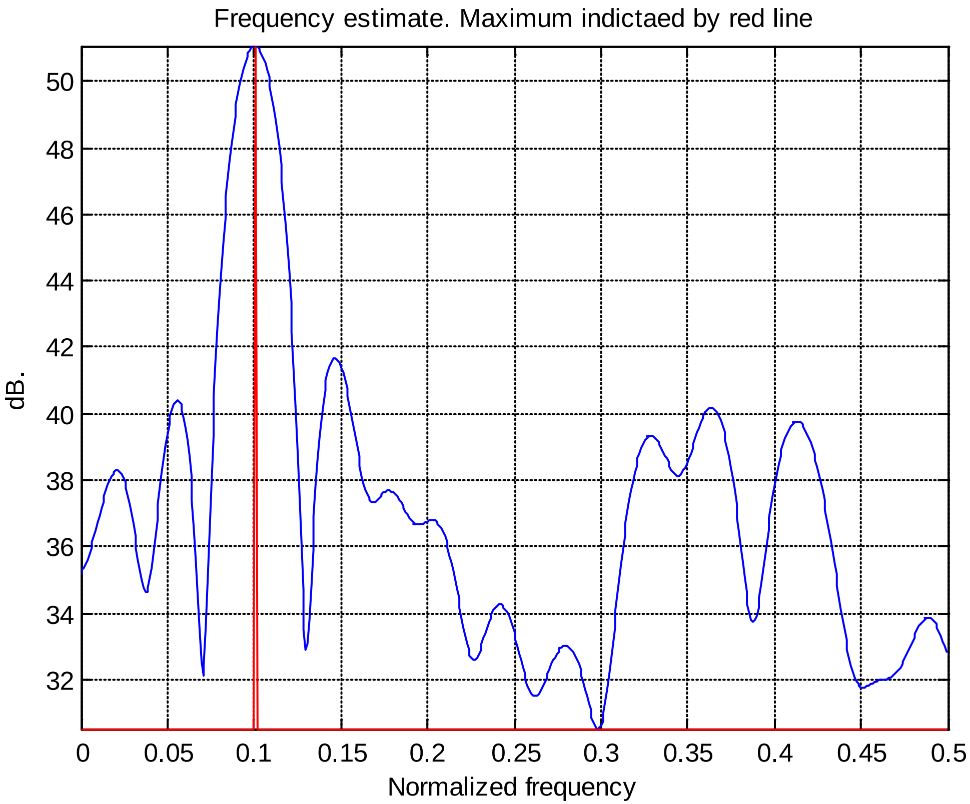
Simulacion



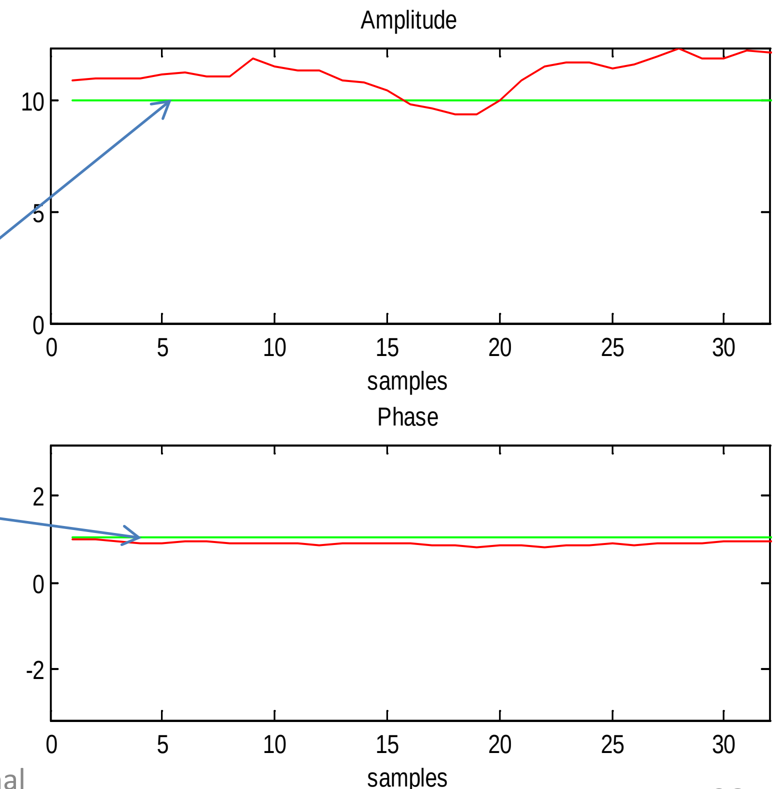


Current amplitude	10
Current phase	1.0472
Current frequency	0.1

SNR=0dB
64 Samples
Segments of 16
4 Segments.



Estimated frequency 0.10059
Estimated noise level 20.3307



Valores correctos de amplitud y fase en verde.
En rojo el valor estimado en cada segmento.

EJEMPLO: Identificacion de Canal

$$y(t)=\sum_{q=0}^{Q-1} h(q)x(t-qT)+w(t)$$

$$\begin{bmatrix} y(t) \\ .. \\ y(t-NT) \end{bmatrix} = \begin{bmatrix} \underline{X}_t^T \\ \dots \\ \underline{X}_{t-NT}^T \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ .. \\ h(Q-1) \end{bmatrix} + \begin{bmatrix} w(t) \\ .. \\ w(t-NT) \end{bmatrix}$$

$$\underline{Y}_t = \underline{\underline{X}}_t \underline{h} + \underline{w}_t$$

La verosimilitud sera:

$$\Pr\left(\underline{Y}_t / h, \underline{\underline{R}}_0, \underline{\underline{X}}_t\right) \propto \frac{1}{\det\left(\underline{\underline{R}}_0\right)} \exp\left[-\left(\underline{Y}_t - \underline{\underline{X}}_t \underline{h}\right)^H \underline{\underline{R}}_0^{-1} \left(\underline{Y}_t - \underline{\underline{X}}_t \underline{h}\right)\right]$$

El gradiente con respecto al vector canal conjugado es:

$$\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \left(\underline{Y}_t - \underline{\underline{X}}_t \underline{h}\right) = \underline{0}$$

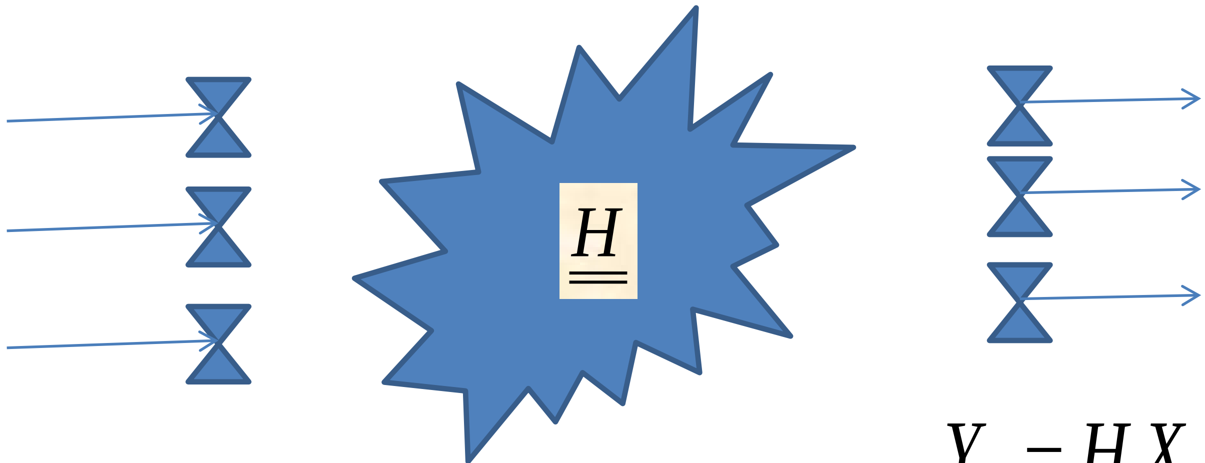
Y el estimador del canal

$$\underline{h}_{ML} = \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1} \underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{Y}_t$$

Si el ruido es gaussiano de potencia No, el estimador sera:

$$\boxed{\underline{h}_{ML} = \left(\underline{\underline{X}}_t^H \underline{\underline{X}}_t\right)^{-1} \underline{\underline{X}}_t^H \underline{Y}_t}$$

EJEMPLO: Filtrado (Estimacion MAP)



$$\underline{X}_t$$

$$\underline{Y}_t = \underline{\underline{H}}\,\underline{X}_t + \underline{w}_t$$

Para calcular la probabilidad de los datos Y condicionado a la entrada a estimar X, se usa un vector Z compuesto con media y covarianza segun se indica:

$$\underline{Z} = \begin{pmatrix} \underline{X}_t \\ \underline{Y}_t \end{pmatrix} \qquad E\left(\underline{Z}\right) = \begin{pmatrix} \underline{m}_x \\ \underline{m}_y \end{pmatrix} \qquad \underline{\underline{C}}_z = E\left(\left(\underline{Z} - \underline{m}_z\right)\left(\underline{Z} - \underline{m}_z\right)^H\right) = \begin{pmatrix} \underline{\underline{R}}_{xx} & \underline{\underline{R}}_{yx} \\ \underline{\underline{R}}_{xy} & \underline{\underline{R}}_{yy} \end{pmatrix}$$

La distribucion de Z sera la conjunta de X e Y

$$\Pr\left(\underline{X}_t, \underline{Y}_t\right) = \Pr\left(\underline{Z}\right) \propto \frac{1}{\det\left(\underline{\underline{C}}_z\right)} \exp\left[-\left(\underline{Z} - \underline{m}_z\right)^H \underline{\underline{C}}_z^{-1} \left(\underline{Z} - \underline{m}_z\right)\right]$$

Para la inversa de la covarianza de Z usaremos la formula de Schur

$$\underline{\underline{C}}_z^{-1} = \begin{pmatrix} \underline{\underline{C}}_{xx} & \underline{\underline{C}}_{yx} \\ \underline{\underline{C}}_{xy} & \underline{\underline{C}}_{yy} \end{pmatrix}^{-1} =$$

$$\begin{pmatrix} \underline{\underline{I}} & \underline{\underline{0}} \\ -\underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{yx} & \underline{\underline{I}} \end{pmatrix} \begin{pmatrix} \left(\underline{\underline{C}}_{xx} - \underline{\underline{C}}_{xy} \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{yx}\right)^{-1} & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{C}}_{yy}^{-1} \end{pmatrix} \begin{pmatrix} \underline{\underline{I}} & -\underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \\ \underline{\underline{0}} & \underline{\underline{I}} \end{pmatrix}$$

Ahora, definiendo:

$$\begin{pmatrix} \underline{x1} \\ \underline{x2} \end{pmatrix} = \begin{pmatrix} \underline{\underline{I}} & -\underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \\ \underline{\underline{0}} & \underline{\underline{I}} \end{pmatrix} \begin{pmatrix} \underline{X}_t - \underline{m}_x \\ \underline{Y}_t - \underline{m}_y \end{pmatrix}$$

Note que Y es identica a x2

Se tendra la expresion de la probabilidad MAP que permite estimar X a partir de Y y el modelo

$$\Pr\left(\underline{X}_t \big/ \underline{Y}_t \right) = \frac{\Pr\left(\underline{X}_t , \underline{Y}_t\right)}{\Pr\left(\underline{Y}_t\right)}$$

$$\propto \frac{\dfrac{1}{\det\left(\underline{\underline{C}}_{x1x2}\right)} \exp\left[-\begin{pmatrix} \underline{x1} \\ \underline{x2} \end{pmatrix}^H \begin{pmatrix} \left(\underline{\underline{C}}_{xx} - \underline{\underline{C}}_{xy}\underline{\underline{C}}_{yy}^{-1}\underline{\underline{C}}_{yx}\right)^{-1} & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{C}}_{yy}^{-1} \end{pmatrix} \begin{pmatrix} \underline{x1} \\ \underline{x2} \end{pmatrix} \right]}{\dfrac{1}{\det\left(\underline{\underline{C}}_{x2}\right)} \exp\left[-\left(\underline{x2}\right)^H \underline{\underline{C}}_{yy}^{-1} \left(\underline{x2}\right) \right]}$$

Usando de nuevo la formula de Schur para el determinante. Ademas, notando que el producto del numerador se desacopla en dos independientes, siendo uno de ellos identico al denominador queda que:

$$\Pr\left(\underline{X}_t \big/ \underline{Y}_t \right) \propto \frac{1}{\det\left(\underline{\underline{C}}_{x1}\right)} \exp\left[-\left(\underline{x1}\right)^H \left(\underline{\underline{C}}_{xx} - \underline{\underline{C}}_{xy}\underline{\underline{C}}_{yy}^{-1}\underline{\underline{C}}_{yx}\right)^{-1} \left(\underline{x1}\right) \right]$$

Recuperando la definicion de x1

$$\begin{pmatrix} \underline{x1} \\ \underline{x2} \end{pmatrix} = \begin{pmatrix} \underline{\underline{I}} & -\underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \\ \underline{\underline{0}} & \underline{\underline{I}} \end{pmatrix} \begin{pmatrix} \underline{X}_t - \underline{m}_x \\ \underline{Y}_t - \underline{m}_y \end{pmatrix} \Rightarrow \underline{x1} = \underline{X}_t - \underline{m}_x - \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \left(\underline{Y}_t - \underline{m}_y \right)$$

Entonces

$$\Pr\left(\underline{X}_t / \underline{Y}_t \right) \propto \exp\left[-\left(\underline{X}_t - \underline{m}_x - \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \left(\underline{Y}_t - \underline{m}_y \right) \right)^H \left(\underline{\underline{C}}_{xx} - \underline{\underline{C}}_{xy} \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{yx} \right)^{-1} \left(\underline{X}_t - \underline{m}_x - \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \left(\underline{Y}_t - \underline{m}_y \right) \right) \right]$$

Para obtener el estimador MAP es necesario encontrar la X que maximiza esta expresion. Dado que la probabilidad es Gaussiana, como ya se menciono, el maximo o MAP es tambien el estimador de la media condicional.

En definitiva se obtiene el estimador MAP y la covarianza o error que presentara la estimacion.

$$\underline{\hat{X}}_{t,MAP} = \underline{m}_x + \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \left(\underline{Y}_t - \underline{m}_y \right)$$

$$\underline{\underline{C}}_{\hat{x}\hat{x}} = \underline{\underline{C}}_{xx} - \underline{\underline{C}}_{xy} \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{yx}$$

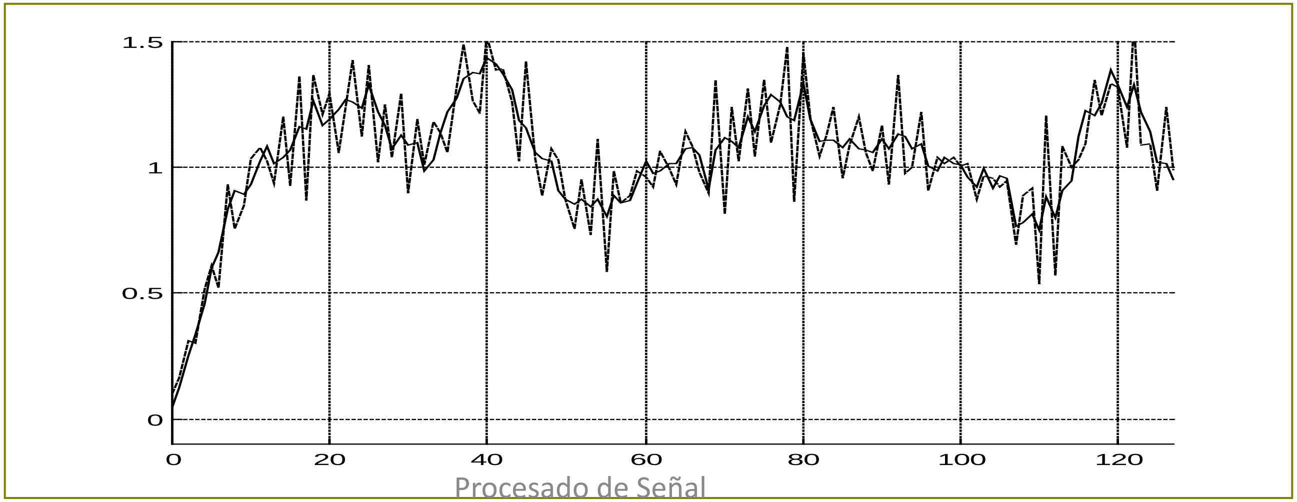
Si se quita la media a los datos y esta no interesa o es cero del vector a estimar, la expresion del estimador MAP se simplifica a:

$$\underline{\hat{X}}_{t,MAP} = \underline{\underline{C}}_{yy}^{-1} \underline{\underline{C}}_{xy} \underline{Y}_t$$

LA CALIDAD DE UN ESTIMADOR

Dado que un estimador de, digamos por ejemplo la media, depende de los datos X y estos son de un proceso estocastico, el estimador en si es una variable aleatoria que fluctua con los datos.

En la figura que sigue se puede ver el estimador de la media, alo largo de varios segmentos de señal del proceso. La media exacta es 10. No obstante, la estimacion fluctua aleatoriamente, mas o menos, alrededor del valor exacto. Que es calidad de un estimador?



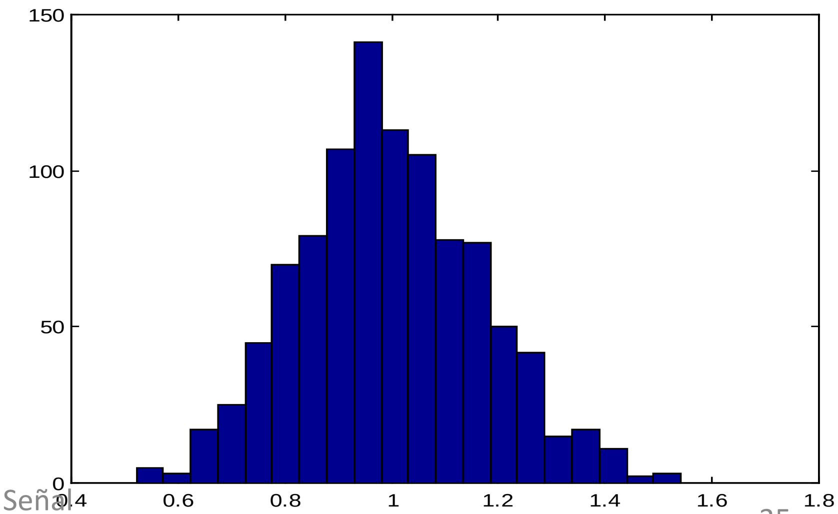
Sea el siguiente estimador de un vector $\underline{\alpha} = f\left(\underline{X}_t\right)$

Asumiendo que el valor exacto del vector que buscamos es el vector a, lo que deseariamos del estimador perfecto es que la distribucion del vector alfa (del estimador) fuese una delta centrada en el vector a (SITUACION IDEAL)

$$\Pr\left(\underline{\alpha}\right) = \partial\left(\underline{\alpha} - \underline{a}\right)$$

Obviamente, lo que ocurre en la practica es que el histograma (estimador de la pdf) al variar el segmento utilizado no es una delta.

La figura presenta el histograma de un estimador de la media cuyo valor exacto es 1.



Sesgo y Varianza de un estimador

Sabiendo que la pdf del estimador no sera, desafortunadamente, una delta, deseariamos que la media coincidiese con el valor exacto. Si NO es asi, diremos que el estimador esta sesgado.

$$\boxed{SESGO \equiv \underline{b} = E\left(\underline{\alpha}\right) - \underline{a}}$$

Es decir el centro de gravedad de la pdf del estimador ha de ser lo que se desea estimar. De otro modo, el estimador sera sesgado.

Mas importante que la nocion de sesgo es la de variabilidad (fiabilidad) del estimador. Es decir, un buen estimador tendra un momento de inercia bajo o, dicho de otro modo una varianza pequeña.

$$\boxed{VARIANZA^2 \equiv \underline{\underline{C}}_{\alpha} = E\left[\left(\underline{\alpha} - E\left(\underline{\alpha}\right)\right)\left(\underline{\alpha} - E\left(\underline{\alpha}\right)\right)^H\right]}$$

Claramente la covarianza es una matriz cuando lo que se estima es un vector.

Es interesante ver la relacion que vector de sesgo y matriz de covarianza tienen con la matriz de error cuadratico medio entre el estimador y el valor exacto

$$MSE \equiv \underline{\underline{C}}_{MSE} = E\left[\left(\underline{\alpha} - \underline{a}\right)\left(\underline{\alpha} - \underline{a}\right)^H\right] =$$

$$E\left[\left(\left(\underline{\alpha} - E\left(\alpha\right)\right) - \left(\underline{a} - E\left(\underline{\alpha}\right)\right)\right)\left(........\right)^H\right]$$

Al desarrollar los cuatro productos: El primero por el primero es la matriz de covarianza del estimador; el segundo por el segundo, es un vector deterministico sobre el que no actua el valor esperado y resulta ser el vector sesgo por su traspuesto; finalmente, los dos productos cruzados con signo menos son productos del primer vector (Aleatorio/deterministico) por el segundo (Deterministico/Aleatorio) siendo iguales al sesgo por su traspuesto. En definitiva el resultado es:

$$MSE \equiv \underline{\underline{C}}_{MSE} = \underline{\underline{C}}_{\alpha} + \underline{b}\underline{b}^H = COVAR + Sesgo.Sesgo^H$$

La cota de Cramer-Rao

Se puede desmostrar que la matriz de covarianza de un estimador insesgado es siempre superior a una matriz CR

$$E\left[\left(\underline{\hat{\theta}}-\theta\right)\left(\underline{\hat{\theta}}-\theta\right)^{H}\right]=\underline{\underline{C}}_{\theta}\succ\underline{\underline{CR}}$$

Notese que una matriz es mayor que otra cuando todos sus autovalores son superiores , es decir, la diferencia es definida positiva (esto implica que tanto traza como determinante son mayores

siendo

$$\underline{\underline{CR}}=\left[E\left\{\left(\nabla_{\underline{\theta}^{H}}L\right)\left(\nabla_{\underline{\theta}^{H}}L\right)^{H}\right\}\right]^{-1}$$

Donde L es la log-likelihood de los datos, i.e. Pr(datos/teta)

Cualidades deseables de un estimador…

- Que sea insesgado
- Que su varianza sea minima (i.e. alcance la cota de C-R)
- Que la varianza tienda a cero cuando la longitud de los datos disponibles tienda a infinito (CONSISTENTE)
- Que sea lineal
- Complejidad baja
- ……….

Ejemplo: El estimador de la media (Sincronizacion)

$$\underline{X}_t = b\underline{S} + \underline{w}_t = a\exp\left(jw_0 + \theta\right)\begin{bmatrix} 1 \\ \exp(-jw_0) \\ .. \\ \exp\left(-j(Q-1)w_0\right) \end{bmatrix}$$

$$\underline{S}^H = \begin{bmatrix} 1 & \exp(jw_0)\ldots\ldots & \exp(j(Q-1)w_0 \end{bmatrix}$$

$$\underline{X}_t = \begin{bmatrix} x(t) \\ x(t-T) \\ .. \\ x(t-(Q-1)T) \end{bmatrix}$$

$$\boxed{b_{ML} = \frac{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{X}_t}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}}}$$

Calculo del valor esperado

$$E(b_{ML}) = E\left(\frac{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{X}_t}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}}\right) = \frac{\underline{S}^H \underline{\underline{R}}_0^{-1} E(\underline{X}_t)}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}} = \frac{\underline{S}^H \underline{\underline{R}}_0^{-1} b \underline{S}}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}} = b$$

INSESGADO

Calculo de la covarianza

$$C_b = E\left[(b_{ML} - b)(b_{ML} - b)^H\right] =$$

$$E\left[\left(\frac{\underline{S}^H \underline{\underline{R}}_0^{-1}}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}}(\underline{X}_t - b\underline{S})\right)(\ldots\ldots\ldots)^H\right] = \frac{\underline{S}^H \underline{\underline{R}}_0^{-1}}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}} E\left(\underline{w}_t \underline{w}_t^H\right) \frac{\underline{\underline{R}}_0^{-1} \underline{S}}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}}$$

$$\boxed{C_b = \frac{1}{\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}}}$$

Veamos si es consistente, para ello usaremos estas dos propiedades

$$\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S} = tr\left(\underline{S}^H \underline{\underline{R}}_0^{-1} \underline{S}\right) = tr\left(\underline{\underline{R}}_0^{-1} \underline{S}\,\underline{S}^H\right)$$

$$\lambda_{\min}\left(\underline{\underline{A}}\right) tr\left(\underline{\underline{B}}\right) \le tr\left(\underline{\underline{A}}\,\underline{\underline{B}}\right) \le \lambda_{\max}\left(\underline{\underline{A}}\right) tr\left(\underline{\underline{B}}\right)$$

Al usar estas dos propiedades, se obtiene que

$$\lambda_{\min}\left(\underline{\underline{R}}_0^{-1}\right)Q \leq tr\left(\underline{S}^H \underline{\underline{R}}_0^{-1}\underline{S}\right) = tr\left(\underline{\underline{R}}_0^{-1}\underline{S}\,\underline{S}^H\right) \leq \lambda_{\max}\left(\underline{\underline{R}}_0^{-1}\right)Q$$

$$o \quad bien$$

$$\frac{Q}{\lambda_{\max}\left(\underline{\underline{R}}_0\right)} \leq tr\left(\underline{S}^H \underline{\underline{R}}_0^{-1}\underline{S}\right) = tr\left(\underline{\underline{R}}_0^{-1}\underline{S}\,\underline{S}^H\right) \leq \frac{Q}{\lambda_{\max}\left(\underline{\underline{R}}_0\right)}$$

Asi pues, la covarianza del estimador esta acotada segun sigue:

$$\frac{\lambda_{\min}\left(\underline{\underline{R}}_0\right)}{Q} \leq \frac{1}{\underline{S}^H \underline{\underline{R}}_0^{-1}\underline{S}} \leq \frac{\lambda_{\max}\left(\underline{\underline{R}}_0\right)}{Q}$$

Esto permite asegurar que el estimador es consistente

$$\lim_{Q\to\infty}\frac{1}{\underline{S}^H \underline{\underline{R}}_0^{-1}\underline{S}} = 0$$

Ejemplo: Identificacion de canal

$$y(t) = \sum_{q=0}^{Q-1} h(q)x(t - qT) + w(t)$$

$$\begin{bmatrix} y(t) \\ .. \\ y(t-NT) \end{bmatrix} = \begin{bmatrix} \underline{X}_t^T \\ \ldots \\ \underline{X}_{t-NT}^T \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ .. \\ h(Q-1) \end{bmatrix} + \begin{bmatrix} w(t) \\ .. \\ w(t-NT) \end{bmatrix}$$

$$\underline{Y}_t = \underline{\underline{X}}_t \underline{h} + \underline{w}_t$$

$$\boxed{\underline{h}_{ML} = \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t \right)^{-1} \underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{Y}_t}$$

Claramente, el estimador es insesgado

$$E(\underline{h}_{ML}) = \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1} \underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} E(\underline{Y}_t) =$$
$$= \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1} \underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t \underline{h} = \underline{h}$$

La matriz de covarianza es:

$$\underline{\underline{C}}_h = \left\langle \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1} \underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \right\rangle E\left[\left(\underline{Y}_t - \underline{\underline{X}}_t \underline{h}\right)\left(\ldots\ldots\ldots\right)^H\right]\left\langle \ldots\ldots \right\rangle^H =$$

$$\boxed{\underline{\underline{C}}_h = \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1}}$$

Veamos como es la traza de la matriz de covarianza

$$tr\left(\underline{\underline{C}}_h\right) = tr\left[\left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1}\right]$$

Usando la propiedad siguiente:

$$\left[tr\left(\underline{\underline{A}}\,\underline{\underline{B}}\right)\right]^2 \leq tr\left(\underline{\underline{A}}\,\underline{\underline{A}}^H\right) tr\left(\underline{\underline{B}}\,\underline{\underline{B}}^H\right)$$
$$igual \quad cuando \quad \underline{\underline{A}} = \underline{\underline{B}}$$
$$en \quad particular$$
$$\left[tr\left(\underline{\underline{I}}\right)\right]^2 \leq tr\left(\underline{\underline{A}}\right) tr\left(\underline{\underline{A}}^{-1}\right)$$
$$igual \quad cuando \quad \underline{\underline{A}} = \underline{\underline{I}}$$

Asi pues:

$$tr\left(\underline{\underline{C}}_h\right)=tr\left[\left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1}\right]\geq \frac{Q^2}{tr\left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)}=$$

$$=\frac{Q^2}{tr\left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)}=\frac{Q^2}{tr\left(\underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t \underline{\underline{X}}_t^H\right)}\geq \frac{Q^2 \lambda_{\min}\left(\underline{\underline{R}}_0\right)}{tr\left(\underline{\underline{X}}_t \underline{\underline{X}}_t^H\right)}$$

Para conseguir el minimo, hemos de asegurarnos que las dos desigualdades se cumplen con igualdad.

-Para la primera desigualdad se requiere que:

- La segunda es obvia

$$\underline{\underline{X}}_t \underline{\underline{X}}_t^H = cte.\underline{\underline{R}}_0$$

Como la potencia de la señal de entrada x(n) es

$$P_x = \frac{tr\left(\underline{\underline{X}}_t \underline{\underline{X}}_t^H\right)}{NQ}$$

La mejor señal para identificar el sistema seria la que sigue, siendo el vector v ruido blanco de potencia unidad

$$\underline{\underline{X}}_t = \underline{\underline{R}}_0^{1/2}.\begin{bmatrix} \underline{v}_t^T \\ \cdots \\ \underline{v}_{t-NT}^T \end{bmatrix}\left(\frac{NP_x}{tr\left(\underline{\underline{R}}_0\right)}\right)^{1/2}$$

La covarianza, usando esta entrada optima para identificar el sistema seria:

$$\underline{\underline{C}}_h = \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1} = \left(\underline{\underline{V}}^H \underline{\underline{V}}\right)^{-1} \frac{tr\left(\underline{\underline{R}}_0\right)}{NQP_x} = \frac{tr\left(\underline{\underline{R}}_0\right)}{N^2 QP_x} \underline{\underline{I}}_Q$$

Note que la covarianza de los coeficientes del canal estimado es diagonal, es decir su estimacion no esta correlada (INDEPENDENCIA de errores)

Cuando el ruido es de potencia sigma al cuadrado, la varianza pasa a ser

$$\underline{\underline{C}}_h = \left(\underline{\underline{X}}_t^H \underline{\underline{R}}_0^{-1} \underline{\underline{X}}_t\right)^{-1} = \frac{\sigma^2}{P_x N} \underline{\underline{I}}_Q$$

En resumen: La covarianza es inversamente proporcional a la SNR a la entrada y disminuye con la longitud de los datos de entrada ES CONSISTENTE pues la covarianza tiende a cero cuando N tiende a infinito

Estimacion Matriz de Correlacion

La matriz que se desea estimar es la siguiente (NxN)

$$\underline{\underline{R}} = E\left\{\underline{X}_n(N)\underline{X}_n^H(N)\right\} = \begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(N-2) & r(N-1) \\ r(-1) & r(0) & r(1) & \cdots & r(N-3) & r(N-2) \\ r(-2) & r(1) & r(0) & \cdots & r(N-4) & r(N-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r(-N+2) & r(-N+3) & r(-N+4) & \cdots & r(0) & r(1) \\ r(-N+1) & r(-N+2) & r(-N+3) & \cdots & r(-1) & r(0) \end{bmatrix}$$

- Hermtiica
- Toeplitz
- DEFINIDA POSITIVA

En concreto, la densidad espectral en W./Hz se define como:

$$S(\omega) = \lim_{Q\to\infty} \frac{1}{\underline{S}^H \underline{S}} \underline{S}^H \underline{\underline{R}}\underline{S} = \lim_{Q\to\infty} \frac{1}{Q} \underline{S}^H \underline{\underline{R}}\underline{S} = r(0) + \lim_{Q\to\infty} \sum_{q=-Q}^{Q} r(q)\exp(jq\omega)$$

$$\underline{S}^H = \begin{bmatrix} 1 & \exp(-j\omega) & \cdots & \exp(-jQ\omega) \end{bmatrix}$$

La primera iniciativa seria sustituirla por el valor instantaneo, con ello el estimador seria:

$$\underline{\hat{R}} = \underline{X}_n(N)\underline{X}_n^H(N)$$

Es claro que es insesgado, es decir su valor esperado es la matriz de autocorrelacion exacta. Es definida positiva. El problema radica en su varianza.

Tomando un elemento cualquiera de la matriz estimada:

$$E\{r_{i,j}\} = E\{x(n-i)x^*(n-j)\} = r_x(j-i)$$

Para calcular la varianza, usaremos que en un proceso gaussiano, se verifica que:

$$E\{x_1x_2x_3x_4\} = E\{x_1x_2\}E\{x_3x_4\} + E\{x_1x_3\}E\{x_2x_4\} + E\{x_1x_4\}E\{x_2x_3\}$$

$$\operatorname{var}^2\{r_{i,j}\} = r_x^{\ 2}(j-i) + r_x^{\ 2}(0)$$

$$\operatorname{var}^2 \geq 2r_x^2(j-i)$$

!!! Este resultado equivale a decir que se estima x con error mas/menos 1.47x !!!

Es claro que se necesita reducir varianza y la mejor manera de hacerlo es promediando mas segmentos. El numero de promedios sera el maximo posible y se realiza con segmentos solapados con todas las muestras menos la primera y la ultima. El estimador de la matriz QxQ sera:

$$\hat{\underline{\underline{R}}} = \frac{1}{M}\sum_{m=0}^{M-1}\underline{X}_m(Q)\underline{X}_m^H(Q)$$

Este estimador carece de sesgo y la varianza de cada elemento sera el valor del elemento dividido por el numero de segmentos utilizados.

¿ Si tan solo se desea la funcion de autocorrelacion, como se estima?

Como la definicion de la funcion de autocorrelacion es la de ser la transformada inversa de la densidad espectral, basta usar el estimador de la matriz en la formula mostrada anteriormente e identificar. Se obtiene que:

$$\hat{S}(\omega) = \frac{1}{\underline{S}^H\underline{S}}\underline{S}^H\hat{\underline{\underline{R}}}\underline{S} = \frac{1}{Q}\underline{S}^H\hat{\underline{\underline{R}}}\underline{S} = \hat{r}(0) + \sum_{q=-Q}^{Q}\hat{r}(q)\exp(jq\omega)$$

$$siendo \quad \boxed{\hat{r}(q) = \frac{1}{N}\sum_{n=0}^{N-|q|-1}x^*(n)x(n+q) \quad para \quad q=-Q,\ldots,Q}$$

Ejemplo

N=6 muestras, Matriz a estimar de 3x3

Segmentos disponibles

$$\begin{bmatrix} x(0) \\ 0 \\ 0 \end{bmatrix} \ldots \begin{bmatrix} x(2) \\ x(1) \\ x(0) \end{bmatrix} \ldots. \begin{bmatrix} x(5) \\ x(4) \\ x(3) \end{bmatrix} \ldots \begin{bmatrix} 0 \\ 0 \\ x(5) \end{bmatrix}$$

Matrices a promediar

$$\begin{bmatrix} x(0)x(0) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \ldots + \begin{bmatrix} x(2)x(2) & x(2)x(1) & x(2)x(0) \\ x(1)x(2) & x(1)x(1) & x(1)x(0) \\ x(0)x(2) & x(0)x(1) & x(0)x(0) \end{bmatrix} + \ldots +$$

$$+ \begin{bmatrix} x(5)x(5) & x(5)x(4) & x(5)x(3) \\ x(4)x(5) & x(4)x(4) & x(4)x(3) \\ x(3)x(5) & x(3)x(4) & x(3)x(3) \end{bmatrix} + \ldots + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & x(5)x(5) \end{bmatrix}$$

Al sumar, la matriz es simetrica, Toeplitz y definida positiva

$$\begin{bmatrix} x(0)x(0)+x(1)x(1)+x(2)x(2)+x(3)x(3)+x(4)x(4)+x(5)x(5) & x(1)x(0)+x(2)x(1)+x(3)x(2)+x(4)x(3)+x(5)x(4) & x(2)x(0)+x(3)x(1)+x(4)x(2) \\ x(0)x(1)+x(1)x(2)+x(2)x(3)+x(3)x(4)+x(4)x(5) & x(0)x(0)+x(1)x(1)+x(2)x(2)+x(3)x(3)+x(4)x(4)+x(5)x(5) & x(1)x(0)+x(2)x(1)+x(3)x(2)+x(4)x(3)+x(5)x(4) \\ x(0)x(2)+x(1)x(3)+x(2)x(4)+x(3)x(5) & x(0)x(1)+x(1)x(2)+x(2)x(3)+x(3)x(4)+x(4)x(5) & x(0)x(0)+x(1)x(1)+x(2)x(2)+x(3)x(3)+x(4)x(4)+x(5)x(5) \end{bmatrix}$$

La funcion de correlacion seria:

$$r(0)=\frac{1}{6}\big(x(0)x(0)+x(1)x(1)+x(2)x(2)+x(3)x(3)+x(4)x(4)+x(5)x(5)\big)$$

$$r(1)=\frac{1}{6}\big(x(0)x(1)+x(1)x(2)+x(2)x(3)+x(3)x(4)+x(4)x(5)\big)$$

$$r(2)=\frac{1}{6}\big(x(0)x(2)+x(1)x(3)+x(2)x(4)+x(3)x(5)\big)$$

$$\boxed{sesgo \quad E\big(\hat{r}(q)\big)=\left(1-\frac{|q|}{Q}\right)r(q) \quad para \quad q=-Q,\ldots,Q}$$

IMPORTANTE

-Al incluir vectores con zeros, la matriz es Toeplitz pero pasa a tener sesgo

-Como el termino r(0) es la potencia se divide por el numero de sumandos de este que es el numero total de muestras.

Si se suprime el transitorio:

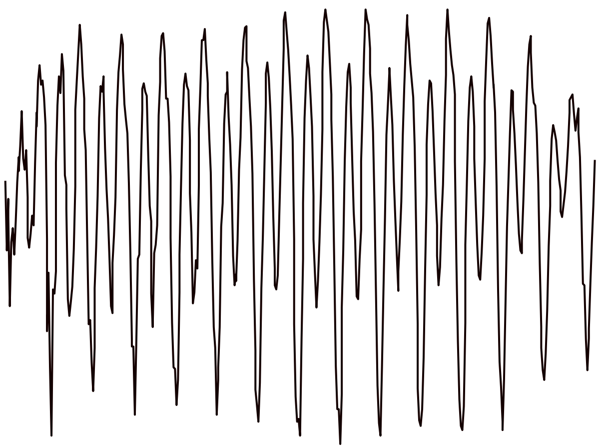
$$\begin{bmatrix} x(2)x(2)+x(3)x(3)+x(4)x(4)+x(5)x(5) & x(2)x(1)+x(3)x(2)+x(4)x(3)+x(5)x(4) & x(2)x(0)+x(3)x(1)+x(4)x(2) \\ x(1)x(2)+x(2)x(3)+x(3)x(4)+x(4)x(5) & x(0)x(0)+x(1)x(1)+x(2)x(2)+x(3)x(3)+x(4)x(4) & x(1)x(0)+x(2)x(1)+x(3)x(2)+x(4)x(3) \\ x(0)x(2)+x(1)x(3)+x(2)x(4)+x(3)x(5) & x(0)x(1)+x(1)x(2)+x(2)x(3)+x(3)x(4) & x(0)x(0)+x(1)x(1)+x(2)x(2)+x(3)x(3) \end{bmatrix}$$

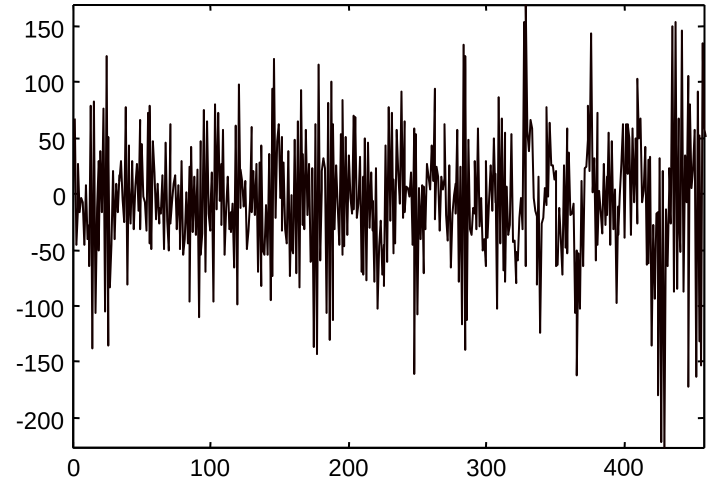
Desaparece el sesgo, No es Toeplitz, Es definida positiva
En terminos practicos, proporciona mejores resultados que la matriz anterior. (PRACTICO)

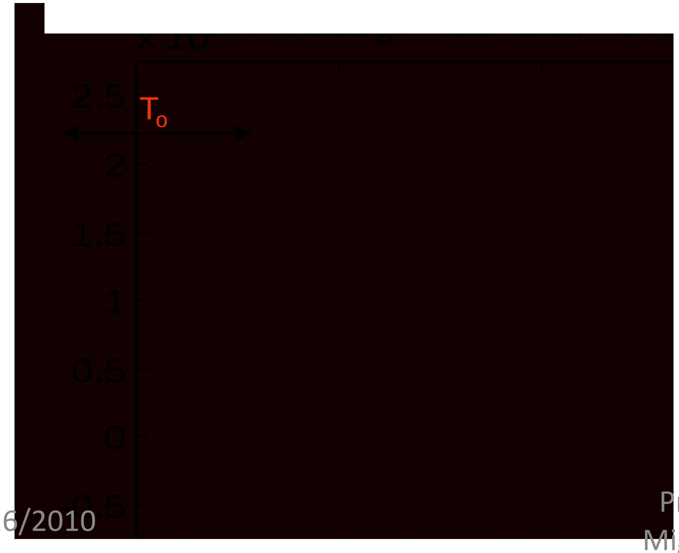
COMENTARIOS ADICIONALES

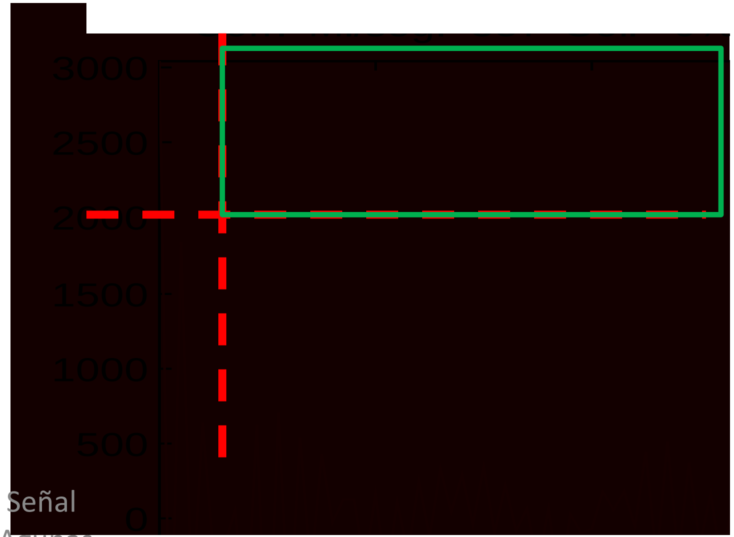
- Si se suman los vectores backward en el metodo de covarianza, la matriz se hace Toeplitz.
-El estimador derivado es de maxima verosimilitud
- Si el numero de segmentos es menor que el orden, la matriz estimada se forma a partir de los autovectores y autovalores distintos de cero. Del mismo modo se procede si se necesita usar la inversa, esta estara formada por la inversa de los autovalores distintos de cero y los correspondientes autovectores.

Ejemplo: Señal pseudo-periodica y señal aleatoria.



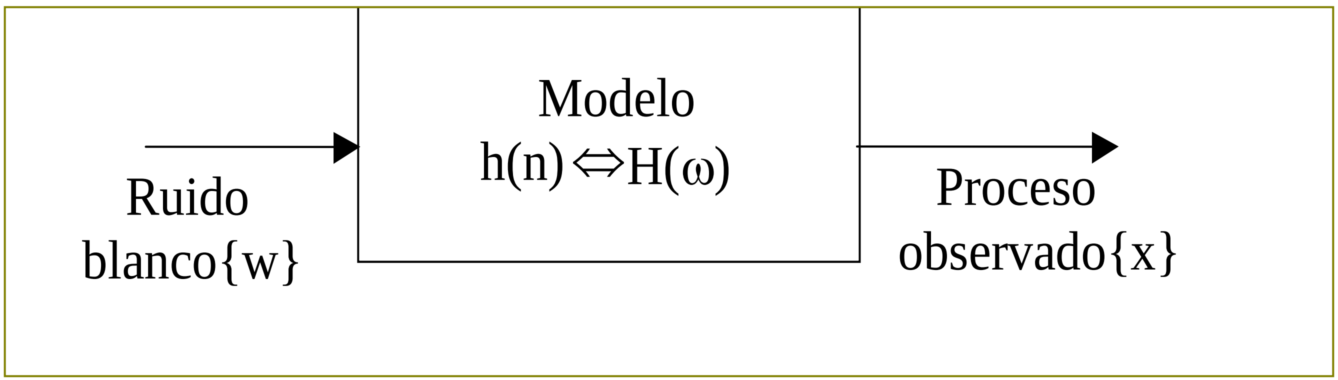






MODELOS RACIONALES

Toda señal perteneciente a un proceso gaussiano se puede considerar generada por ruido blanco de potencia sigma cuadrado aplicado a un sistema lineal.



Por esta razon, usando formulas habituales de sistemas lineales se obtienen las siguientes relaciones

$$x(n) = \sum_{m=0}^{\infty} h(m)w(n-m)$$

$$X(z) = H(z)W(z)$$

$$R_x(z) = \sum_{-\infty}^{\infty} r(n)z^{-n} = E\left(X(z)X(1/z)\right) = \sigma^2 H(z)H(1/z)$$

$$s_x(w) = R_x(z)\Big|_{z=\exp(jwT)} = \sigma^2 H\left(e^{jwT}\right)H\left(e^{-jwT}\right) = \sigma^2 \left|H\left(e^{jwT}\right)\right|^2$$

¿ CUAL ES NUESTRO INTERES POR SISTEMAS RACIONALES?

Si dados un conjunto de valores de la autocorrelacion del proceso rx(q), somos capaces de encontrar los coeficientes del modelo. Dispondremos de la densidad espectral del proceso (Resonancias, antiresonancias o abosorcion, ancho de banda, etc.), de su funcion de autocorrelacion completa, los parametros adquiriran sentido fisico en muchas aplicaciones.

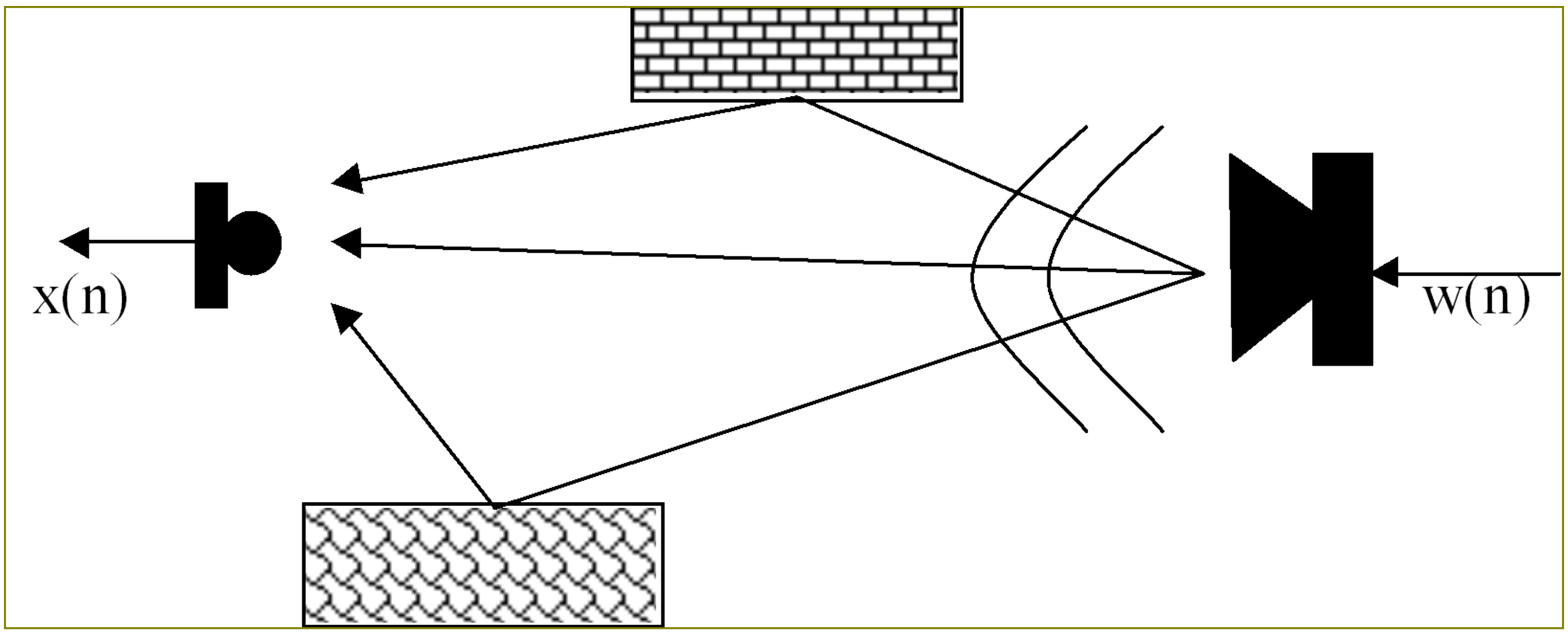
¿ QUE AMBITOS SE HAN BENEFICIADO MAS DEL USO DE MODELOS RACIONALES ?

- Acustica de recintos
- Propagacion Sonar (Geofisica y prospeccion en particular)
- Sistemas de prediccion
- Estudio de series temporales en economia

¿ ES POSIBLE ENCONTRAR EL MODELO DADOS UNOS CUANTOS VALORES DE LA CORRELACION?

- No. Salvo en determinadas circunstancias.

MODELOS MA(P) “Moving Average” orden P



Al propagar ondas longitudinales (presion) en espacio abiertos o no reverberantes, la señal recogida es una suma ponderada de la señal aplicada al proyector w(n).

$$x(n) = w(n) + \sum_{p=1}^{P} b(p)w(n-p) = b(n) * w(n)$$

$$X(\omega) = B(\omega)W(\omega) \quad siendo \ B(\omega) = 1 + \sum_{p=1}^{P} b(p)\exp(-jp\omega)$$

Claramente

$$R_x(z) = \sigma^2 B(z) B(1/z)$$

Asi pues, dado R(z), formado por P valores de la autocorrelacion de x(n) estimados,

$$R_x(z) = \sum_{n=-P}^{P} r_x(n) z^{-n}$$

Ahora se factoriza el polinomio, seleccionando la mitad de las raices para B(z), el coeficiente b(0) se saca factor comun y pasa a ser el ruido de la secuencia de entrada.

PROBLEMAS:

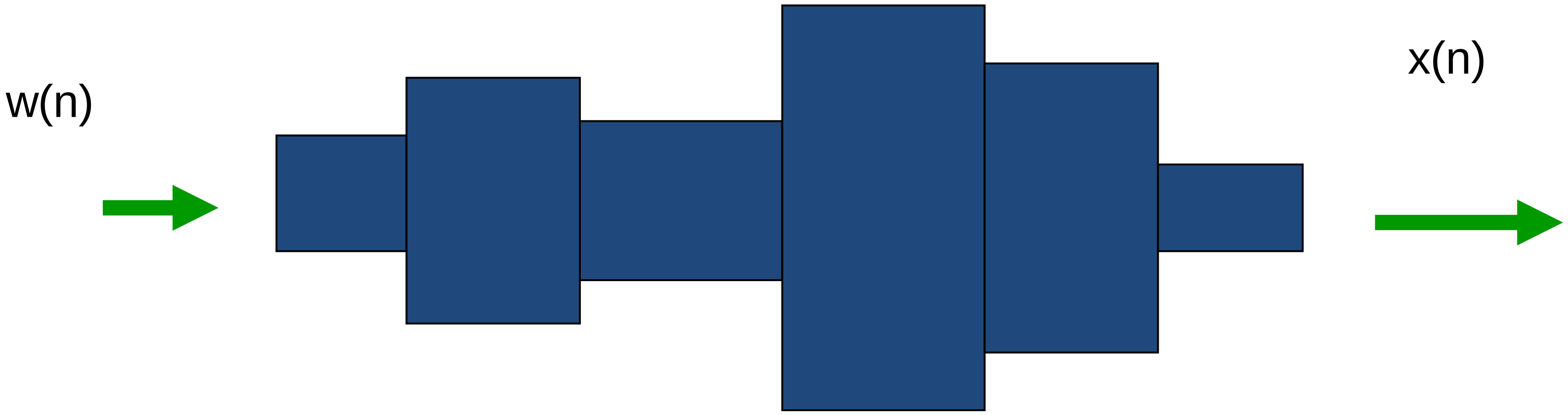
-Es necesario saber el orden P

- Los valores estimados de correlacion aumentan su sesgo cuando el lag se separa del origen (malos valores de autocorrelacion para n grande.

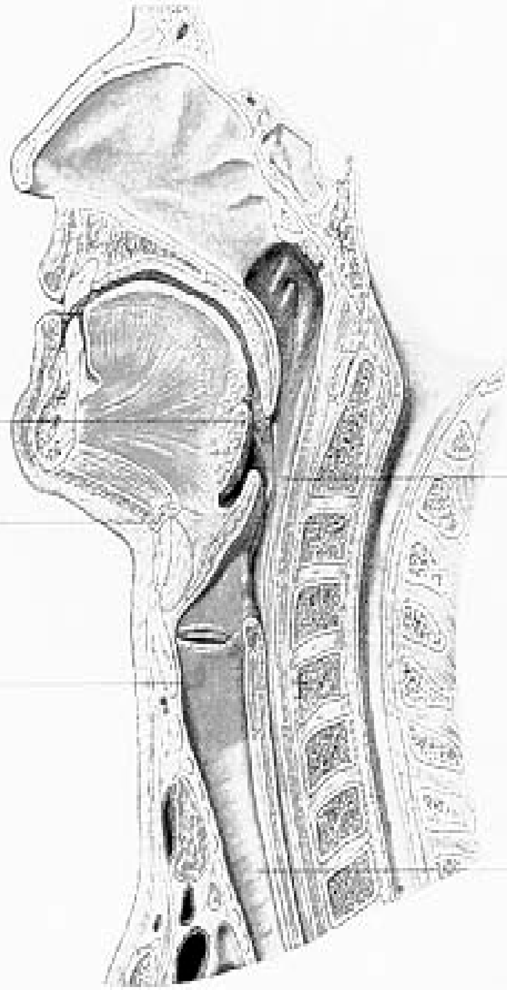
- Existe ambiguedad en las raices a seleccionar. Por ejemplo, todas dentro del circulo unidad, todas fuera o mezcla de ambas opciones. No obstante, para la densidad espectral del proceso dan el mismo resultado.

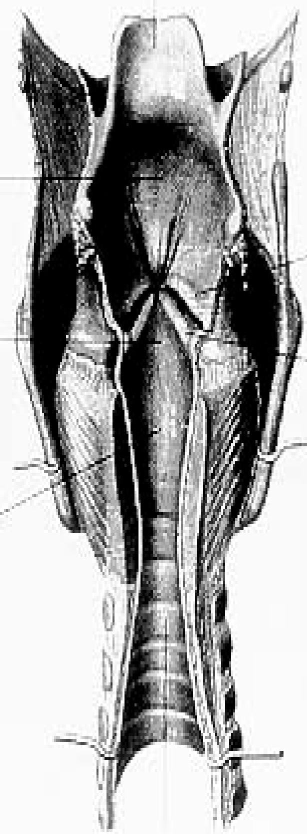
- Seleccionando todas las raices dentro del circulo unidad, el modelo es de retardo minimo y por esta razon es el preferido (¿Porque?).

MODELOS AR(Q) “Auto regressive” orden Q



Ondas de presion en recintos reverberantes. Ejemplos: Escape de vehiculos, Acustica de salas, Generacion de Voz, Prospeccion geofisica.


BASE DE LENGUA
HIOIDES
LARINGE
FARINGE
TRAQUEA


EPIGLOTIS
SUPRAGLOTIS
BANDA VENTRICULAR
GLOTIS
CUERDA VOCAL
SUBGLOTIS
TRAQUEA

Las ecuaciones que definen
un modelo AR(Q) seran:

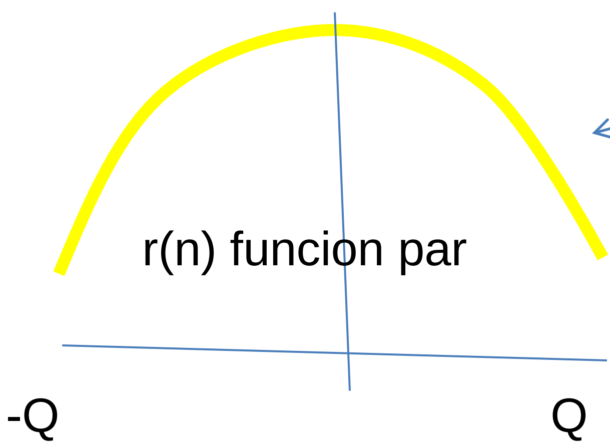
$$x(n) = w(n) - \sum_{q=1}^{Q} a(q)x(n-q)$$

$$S(\omega) = \frac{\sigma^2}{\left|A(\omega)\right|^2} = \left.\frac{\sigma^2}{A(z)A^*(1/z^*)}\right|_{z=\exp(j\omega T)}$$

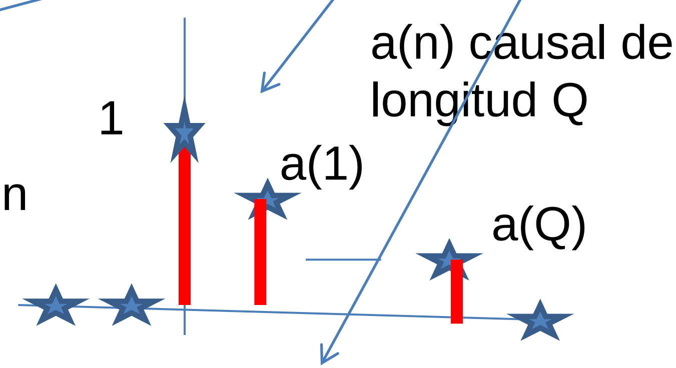
$$H(z) = \frac{1}{1+\sum_{q=1}^{Q} a(q)z^{-q}} = 1/A(z)$$

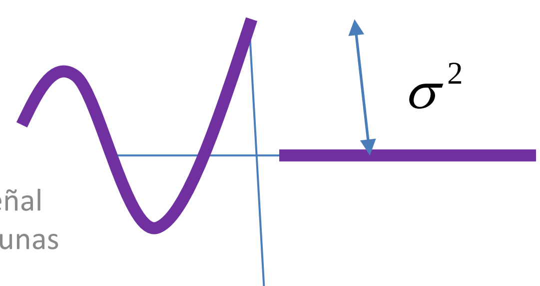
$$R(z)A(z) = \frac{\sigma^2}{A(1/z)}$$

Escribiendo la expresion de R(z) como se indica:

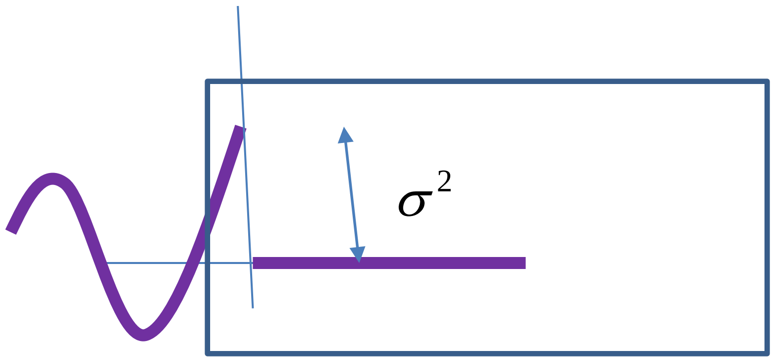


Convolucion





Igual a una funcion anticausal
con valor sigma cuadrado en el origen



La salida de esta convolucion representa infinitas ecuaciones de diseño. La primera en n=0 es la potencia del ruido de entrada, de n igual a uno hasta infinito la salida ha de ser cero

Escribiendo la convolucion en forma matricial, a(n) quieto y r(n) dado la vuelta y desplazandose se obtiene el siguiente conjunto de ecuaciones

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-Q) \\ r(1) & r(0) & \cdots & r(-Q+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(m) & r(m-1) & \cdots & r(m-Q) \\ r(m+1) & r(m) & \cdots & r(m-Q+1) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

Dado que la correlacion se estima mejor para lags pequeños, las ecuaciones de diseño del modelo AR se hacen con las primeras

Asi pues las ecuaciones de diseño seran:

$$\begin{bmatrix} r(1) & r(0) & .. & r(Q-2) & r(Q-1) \\ r(2) & r(1) & .. & r(Q-3) & r(Q-2) \\ .. & .. & .. & .. & .. \\ r(Q-1) & r(Q-2) & .. & r(0) & r(1) \\ r(Q) & r(Q-1) & .. & r(1) & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ .. \\ 0 \\ 0 \end{bmatrix}$$

Y la ecuacion que permite calcular la potencia de la señal de entrada es:

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

Otras cualidades del modelado AR

-Con Q+1 valores de la autocorrelacion se calculan los Q coeficientes del modelo y la potencia del ruido de entrada.
- La densidad espectral exacta del proceso AR viene dada por

$$s_x(w) = \frac{\sigma^2}{\left|A\left(e^{jwT}\right)\right|^2} = \frac{\sigma^2}{\left|1 + \sum_{q=1}^{Q} a(q) e^{jqwT}\right|^2}$$

- Conocidos los coeficientes, la autocorrelacion completa del proceso se puede calcular usando el resto de ecuaciones obtenidas

$$r(m) = -\sum_{q=1}^{Q} a(q) r(m-q) \quad m \geq Q+1$$

- Si el proceso no es de verdad AR(Q), la densidad espectral de arriba pasa a ser un buen estimador de la densidad del proceso (No exacta,estimacion)

Una manera eficiente de resolver las ecuaciones de Yules-Walker

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-Q) \\ r(1) & r(0) & \cdots & r(-Q+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(m) & r(m-1) & \cdots & r(m-Q) \\ r(m+1) & r(m) & \cdots & r(m-Q+1) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

En formulacion matricial $\underline{\underline{R}}\,\underline{a} = \sigma^2 \underline{1}$

La solucion para el vector a es de la forma $\underline{a} = \sigma^2 \underline{\underline{R}}^{-1} \underline{1}$

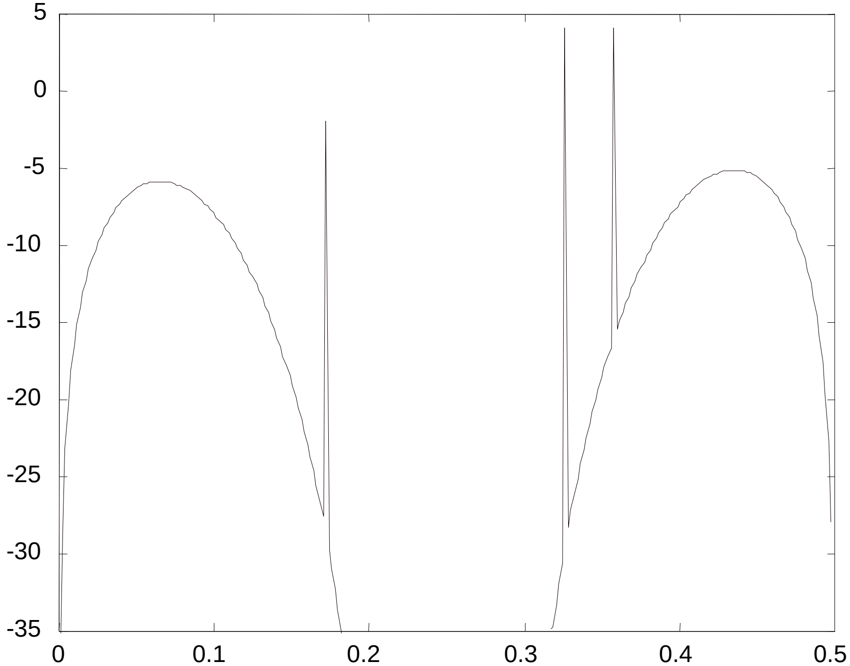
Sigma cuadrado se obtiene
Forzando el primer coeficiente
del vector a la unidad

$$1 = \underline{1}^H \underline{a} = \sigma^2 \left(\underline{1}^H \underline{\underline{R}}^{-1} \underline{1} \right) \Rightarrow \boxed{\sigma^2 = \frac{1}{\underline{1}^H \underline{\underline{R}}^{-1} \underline{1}}}$$

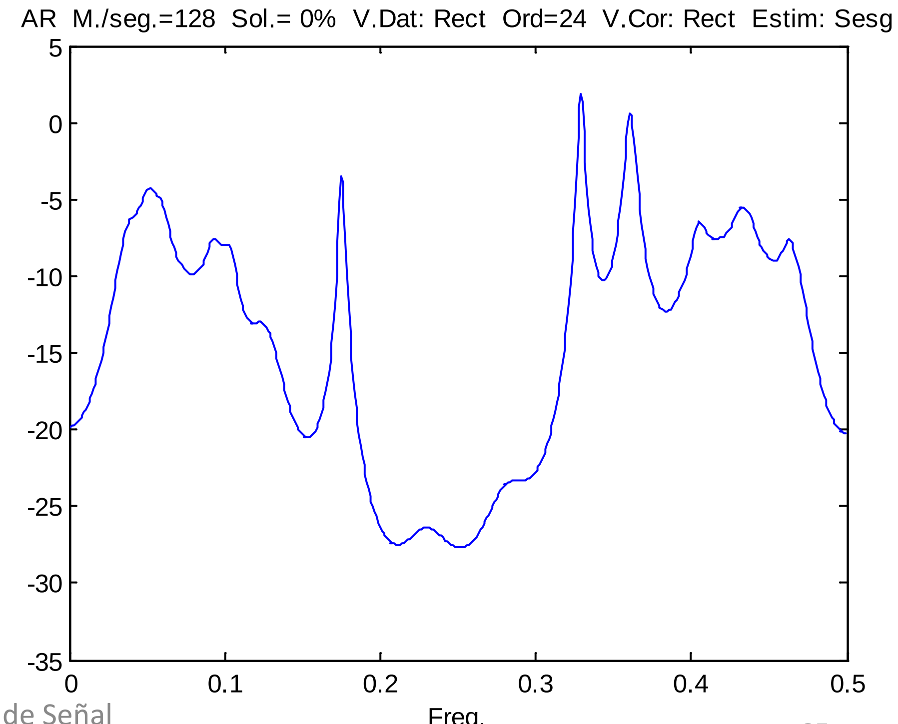
Y el vector de
coeficientes

$$\boxed{\underline{a} = \frac{\underline{\underline{R}}^{-1} \underline{1}}{\underline{1}^H \underline{\underline{R}}^{-1} \underline{1}}}$$

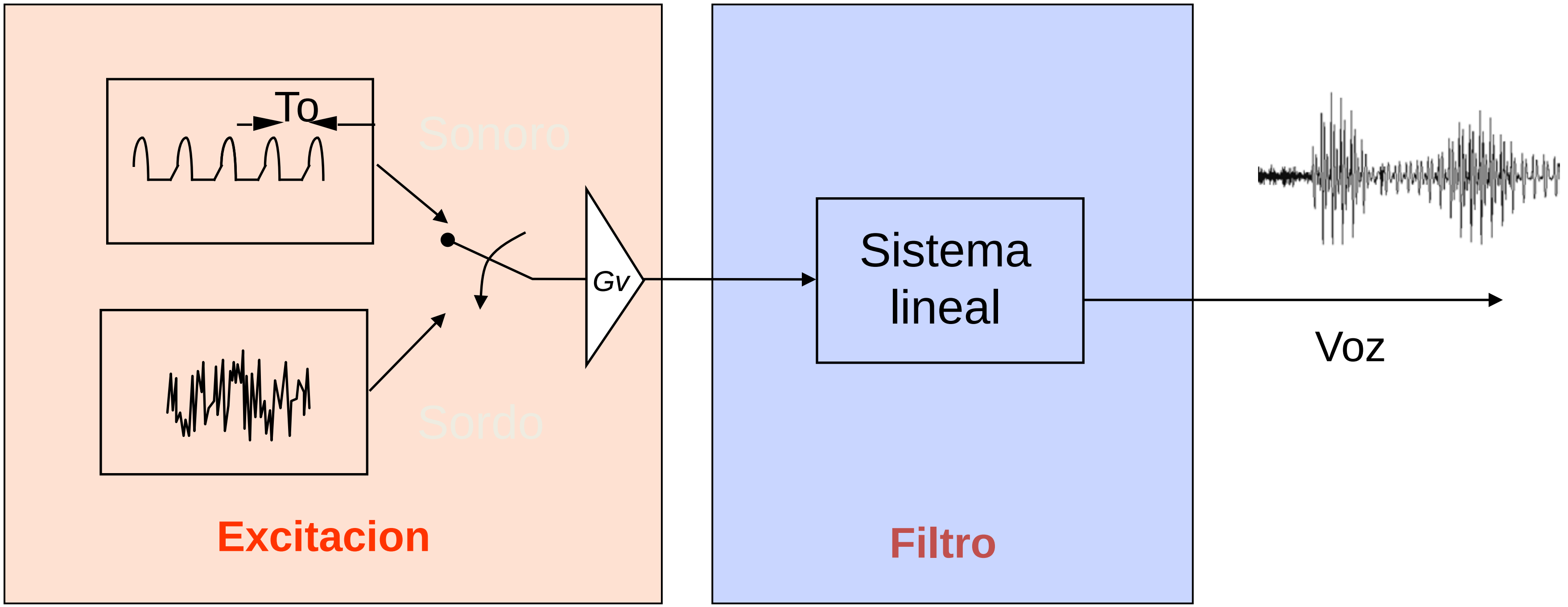
Espectro correcto



Estimacion AR(24)



Procesado LPC para Voz



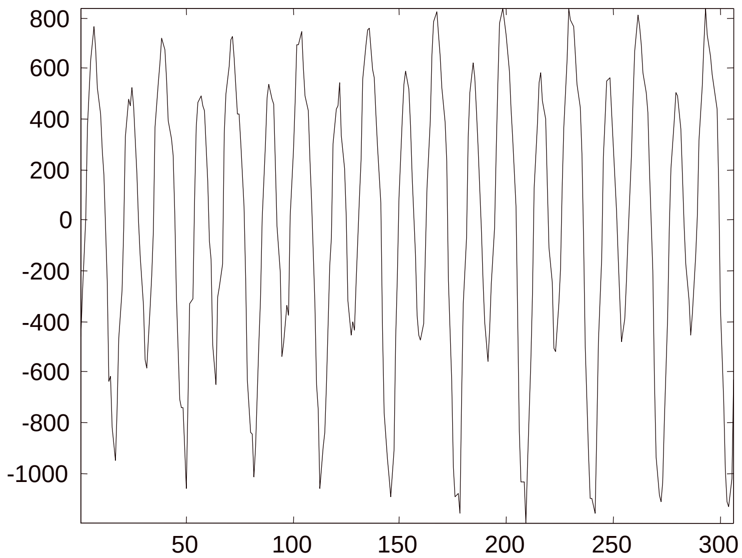
Fuente de presion

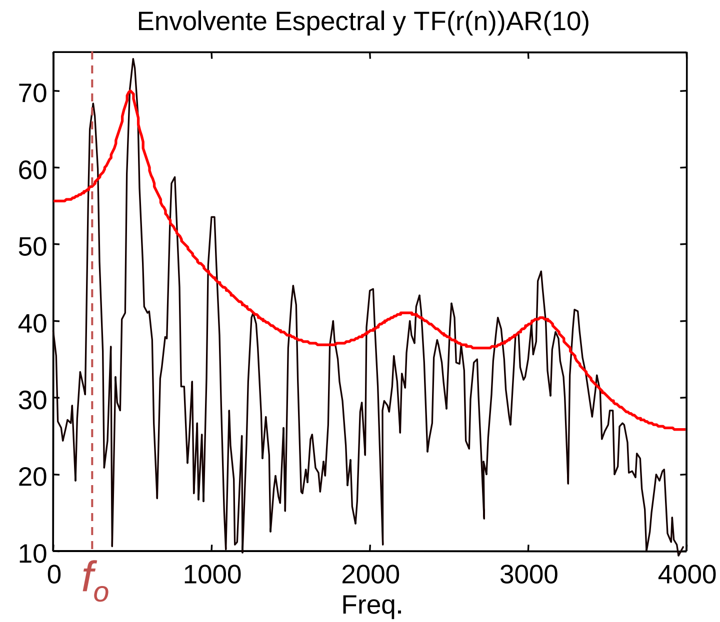
Tracto Vocal

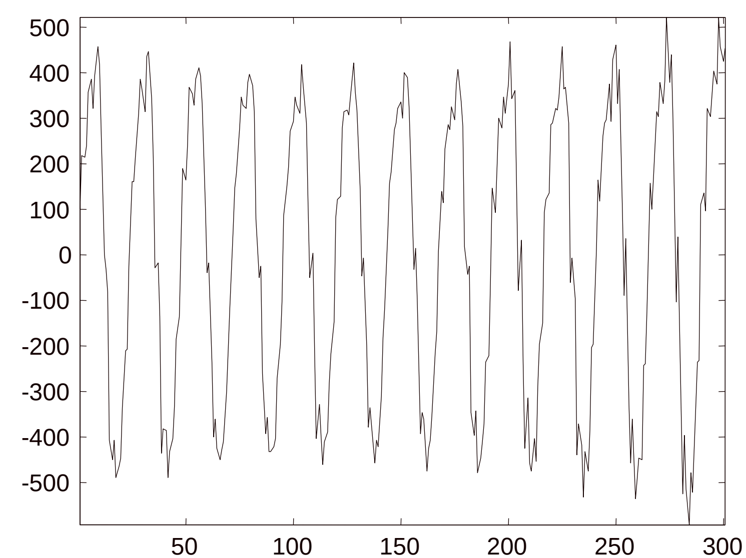
Sonoridad, tono e intensidad

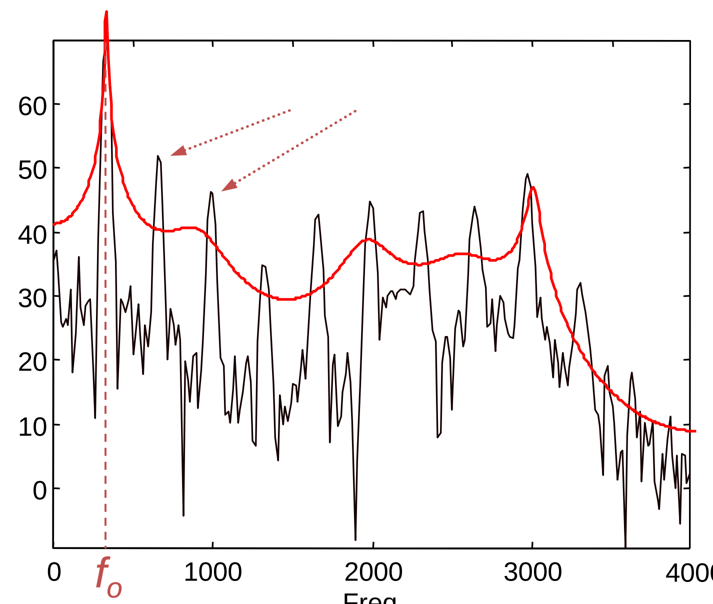
Formantes

Fonema /e/


800
600
400
200
0
-200
-400
-600
-800
-1000
50
100
150
200
250
300


Envolvente Espectral y TF(r(n))AR(10)
70
60
50
40
30
20
10
0
f_o
1000
2000
3000
4000
Freq.


500
400
300
200
100
0
-100
-200
-300
-400
-500
50
100
150
200
250
300


60
50
40
30
20
10
0
0
f_o
1000
2000
3000
4000
Freq.

¿Porque se le denomina envolvente espectral al estimador AR

Cuano la señal de voz es voiced o sonora, su espectro correcto son lineas espectrales (rayas) al periodo o melodia con que excita el tracto vocal la glotis, es decir, el espectro correcto es:

$$s_x(w) = \sum s(q)\partial\left(w - w_q\right)$$

El modelado AR siempre verifica la ecuacion

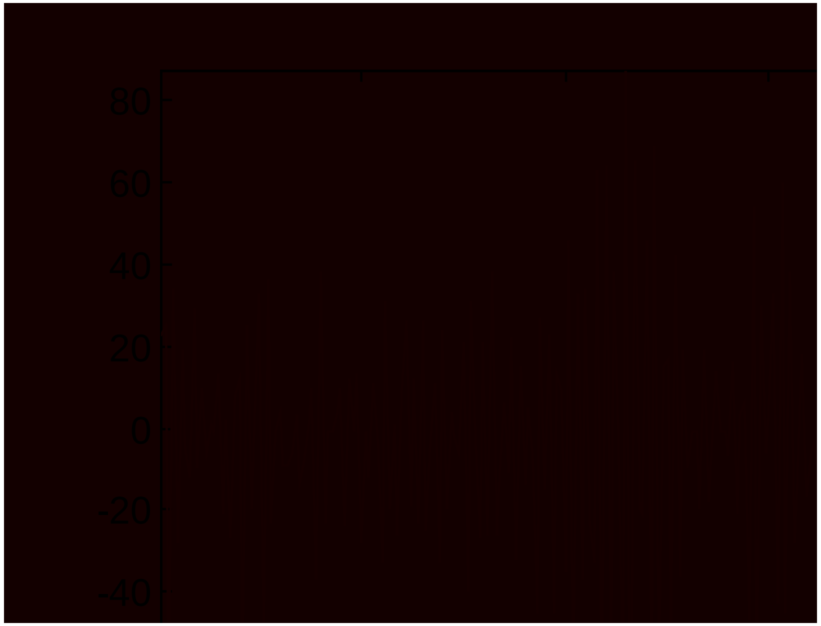
$$s_{salida}(w) = s_x(w)\left|A\left(e^{jwT}\right)\right|^2$$

$$P_{salida} = \frac{1}{2\pi T}\int_{-\pi/T}^{\pi/T} s_x(w)\left|A\left(e^{jwT}\right)\right|^2 dw$$

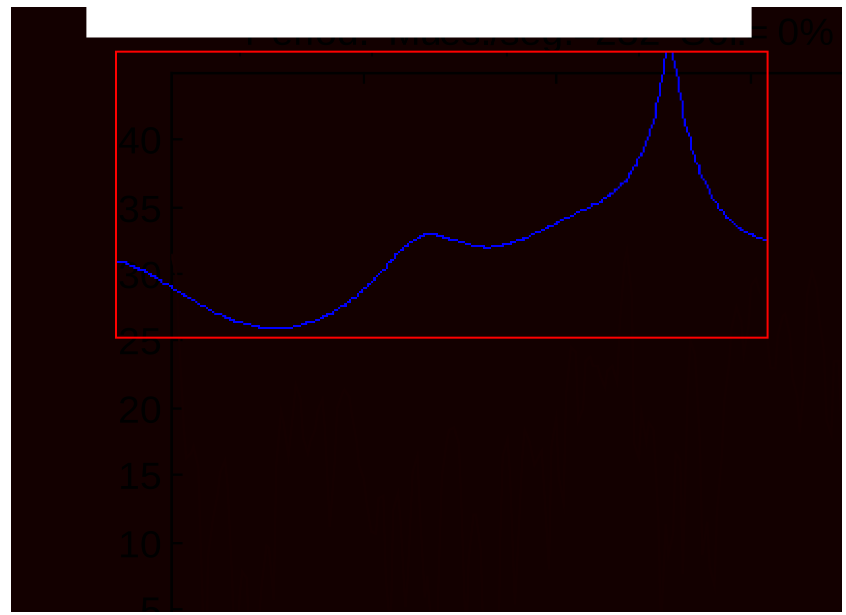
Como la potencia de salida ha de igualar a sigma cuadrado, esta claro que el modelo AR no reacciona cuanod el espectro de verdad es cero, tan solo necesita igualarlo cuando es distinto de cero

$$\boxed{P_{salida} = \sigma^2\left(\frac{1}{2\pi T}\int_{-\pi/T}^{\pi/T} \frac{s_x(w)}{\hat{s}_{AR}(w)} dw\right)}$$

Señal voz (un-voiced)



Transformada F de r(n) y estimacion AR o envolvente espectral



MODELOS ARMA(P,Q)

Las ecuaciones de un modelo ARMA seran:

$$x(n) = w(n) + \sum_{p=1}^{P} b(p) w(n-p) - \sum_{q=1}^{Q} a(q) x(n-q)$$

$$H(z) = \frac{1 + \sum_{p=1}^{P} b(p) z^{-p}}{1 + \sum_{q=1}^{Q} a(q) z^{-q}} = B(z)/A(z)$$

Es facil verificar que la relacion entre los coeficientes del modelo y la respuesta impulsional son:

$$\begin{bmatrix} 1 \\ b(1) \\ \vdots \\ b(P) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h(1) & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ h(P) & h(P-1) & \cdots & h(P-Q) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix}$$

Como se verifica que $R(z)=\sigma^2 H(z)H^*(1/z^*)$ $\quad o\ bien \quad$ $S(\omega)=\sigma_w^2 H(\omega)H^*(\omega)$

entonces

$$R(z)=\sigma_w^2\frac{B(z)}{A(z)}H^*(1/z^*) \quad o\ bien \quad R(z)A(z)=\sigma_w^2 H^*(1/z^*)B(z)$$

El primer termino es identico al caso AR, el segundo es la convolucion de la respuesta (anticausal) del sistema con el numerador. Escrito todo en forma matricial resulta ser:

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-Q) \\ r(1) & r(0) & \cdots & r(-Q+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(P) & r(P-1) & \cdots & r(P-Q) \\ \vdots & \vdots & \ddots & \vdots \\ r(P+1) & r(P) & \cdots & r(P+1-Q) \\ \vdots & \vdots & \ddots & \vdots \\ r(P+Q+1) & r(P+Q) & \cdots & r(P+1) \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & h(1) & \cdots & h(P) \\ 0 & 1 & \cdots & h(P-1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ b(P) \end{bmatrix}$$

Es decir, despues de P, desde P+1 a P+Q se dispone de Q ecuaciones con incognitas solo los coeficientes del denominador.

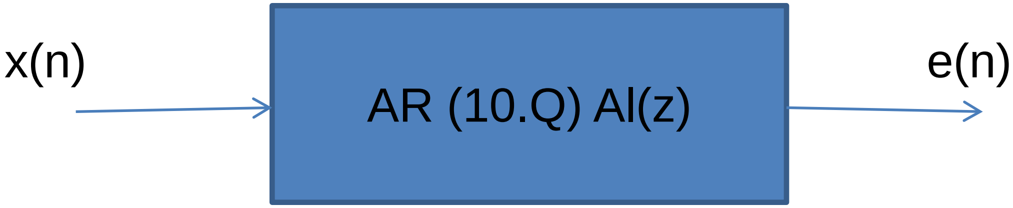
Asi pues los coeficientes del denominador se calculan en un modelo ARMA(P,Q) con las denominadas ecuaciones de Y-W extendidas

$$\begin{bmatrix} r(P+1) & r(P) & .. & r(P-Q+1) \\ r(P+2) & r(P+1) & .. & r(P-Q) \\ .. & .. & .. & .. \\ r(P+Q) & r(P+Q-1) & .. & r(P-1) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(Q) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Una vez obtenido el denominador, se filtra x(n) con el polinomio A(z) y se resuelve para el numerador como en un modelo MA.

Otras alternativas a la estimacion de modelos ARMA

Un metodo eficiente de estimacion ARMA se debe a Durbin. La idea es filtrar la señal original con un modelo AR de orden lo mas alto posible. Dado que el modelo AR obtenido se considera modela efectivamente bien el espectro del proceso (Ver mas adelante el metodo MEM), se puede considerar que la salida e(n) es blanca y coherente (derivada linealmente) de la señal dada.



Asi pues dados ahora x(n) y e(n), se resuelve el problema de identificacion del modelo por MSE

$$X(z)A(z) - E(z)B(z)$$

$$\xi = \sum \left| \underline{a}^H \underline{X}_n - \underline{b}^H \underline{E}_n \right|^2$$

$$sujeto \quad a:$$

$$\underline{a}^H \underline{1} = 1 \quad \underline{b}^H \underline{1} = 1$$

Tomandoel el gradiente con respecto a los dos vectores y con los dos multiplicadores para cada restriccion, se obtiene la solucion para los coeficientes de numerador y denominador

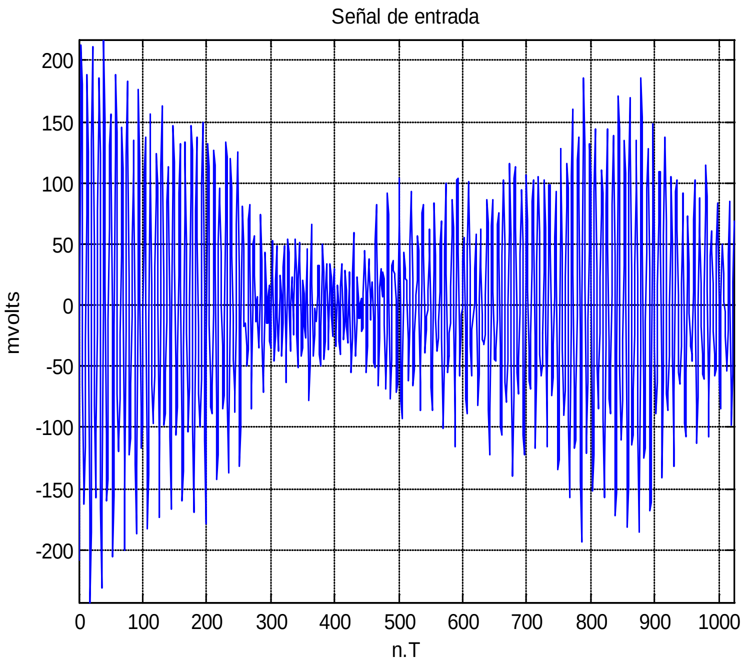
$$\nabla_{\underline{a}^H} \Rightarrow \left(\sum \underline{X}_n \underline{X}_n^H\right)\underline{a} - \left(\sum \underline{X}_n \underline{E}_n^H\right)\underline{b} = \mu \underline{1}$$

$$\nabla_{\underline{b}^H} \Rightarrow \left(\sum \underline{E}_n \underline{X}_n^H\right)\underline{a} - \left(\sum \underline{E}_n \underline{E}_n^H\right)\underline{b} = \lambda \underline{1}$$

$$\left(\sum \underline{X}_n \underline{X}_n^H\right)\underline{a} - \left(\sum \underline{X}_n \underline{E}_n^H\right)\left(\sum \underline{E}_n \underline{E}_n^H\right)^{-1}\left(\sum \underline{E}_n \underline{X}_n^H\right)\underline{a} = (\mu - \lambda)\underline{1}$$

$$\underline{a} = \frac{\left(\underline{\underline{XX}} - \underline{\underline{XE}}\,\underline{\underline{EE}}^{-1}\,\underline{\underline{EX}}\right)^{-1}\underline{1}}{\underline{1}^H\left(\underline{\underline{XX}} - \underline{\underline{XE}}\,\underline{\underline{EE}}^{-1}\,\underline{\underline{EX}}\right)^{-1}\underline{1}} \qquad \underline{b} = \frac{\left(\underline{\underline{EE}} - \underline{\underline{EX}}\,\underline{\underline{XX}}^{-1}\,\underline{\underline{XE}}\right)^{-1}\underline{1}}{\underline{1}^H\left(\underline{\underline{EE}} - \underline{\underline{EX}}\,\underline{\underline{XX}}^{-1}\,\underline{\underline{XE}}\right)^{-1}\underline{1}}$$

Note que el denominador de cada vector es simplemente normalizar el primer coeficiente a la unidad

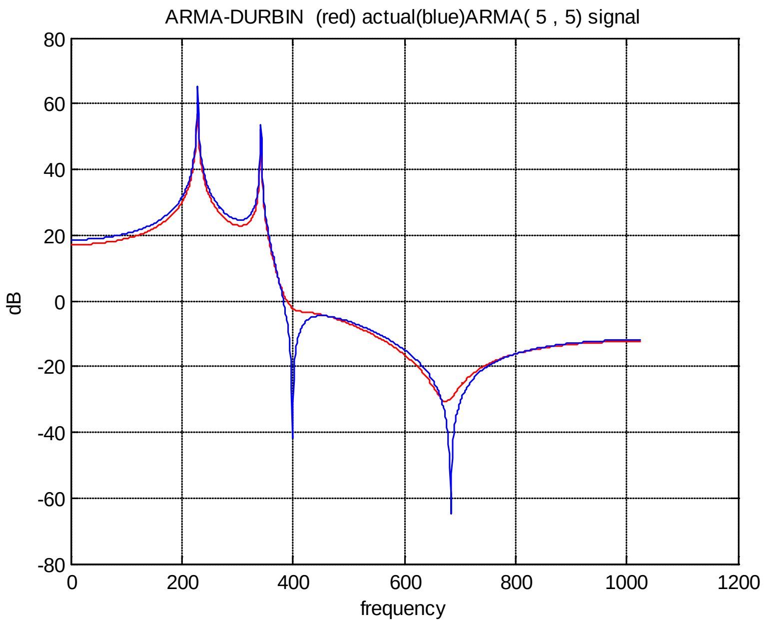


ARMA(5,5)

Numerator coefficients
1.0000 0.3160 1.3160 0.3160 1.0000

Denominator coefficients
1.0000 -2.5296 3.5250 -2.5245 0.9960

Estimador de Durbin orden 5,5 el predictor original se situo a orden 50

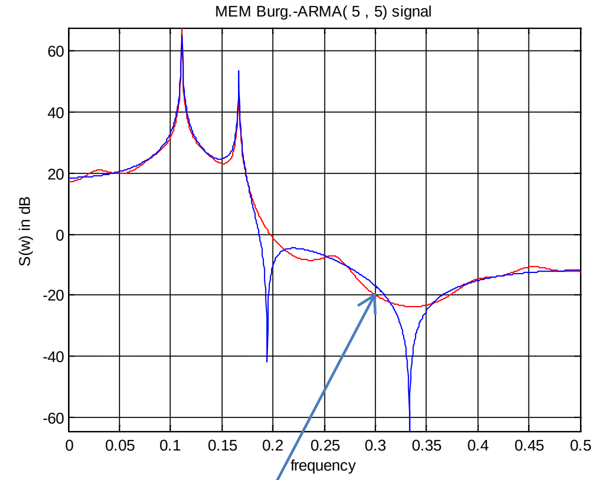


Durbin Method
Noise power
1.1509

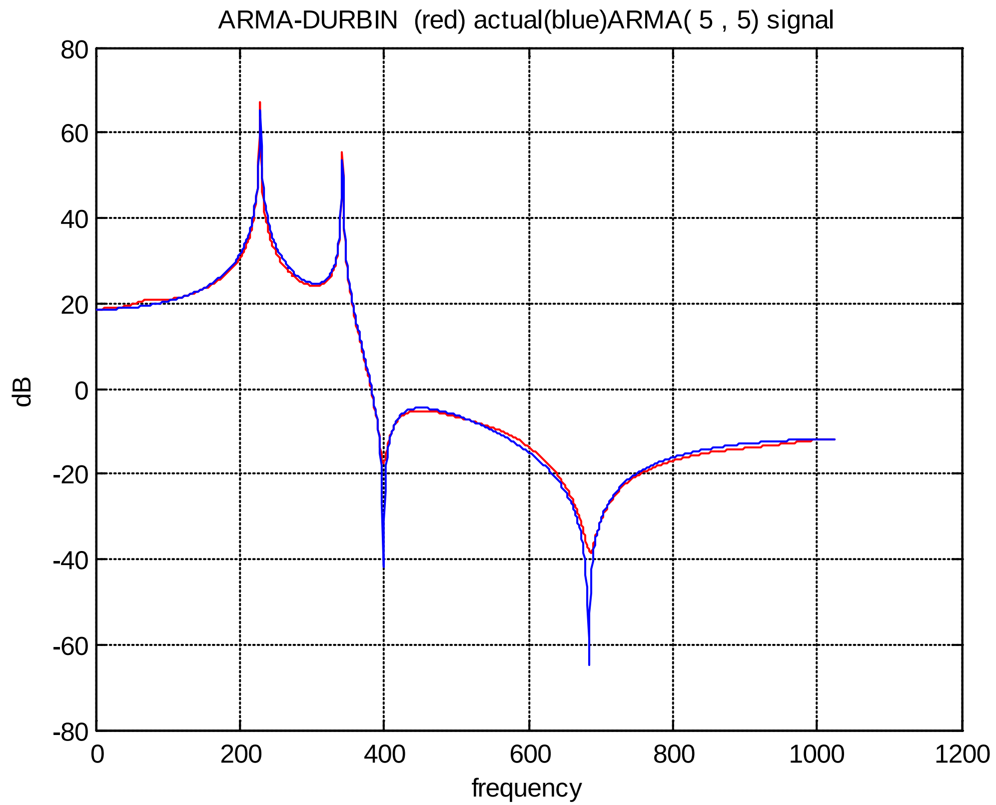
Numerator coefficients
1.0000 0.2198 1.1551 0.1688 0.7782

Coefficients denorminator
1.0000 -2.5267 3.5180 -2.5180 0.9919

Durbin (15,15)



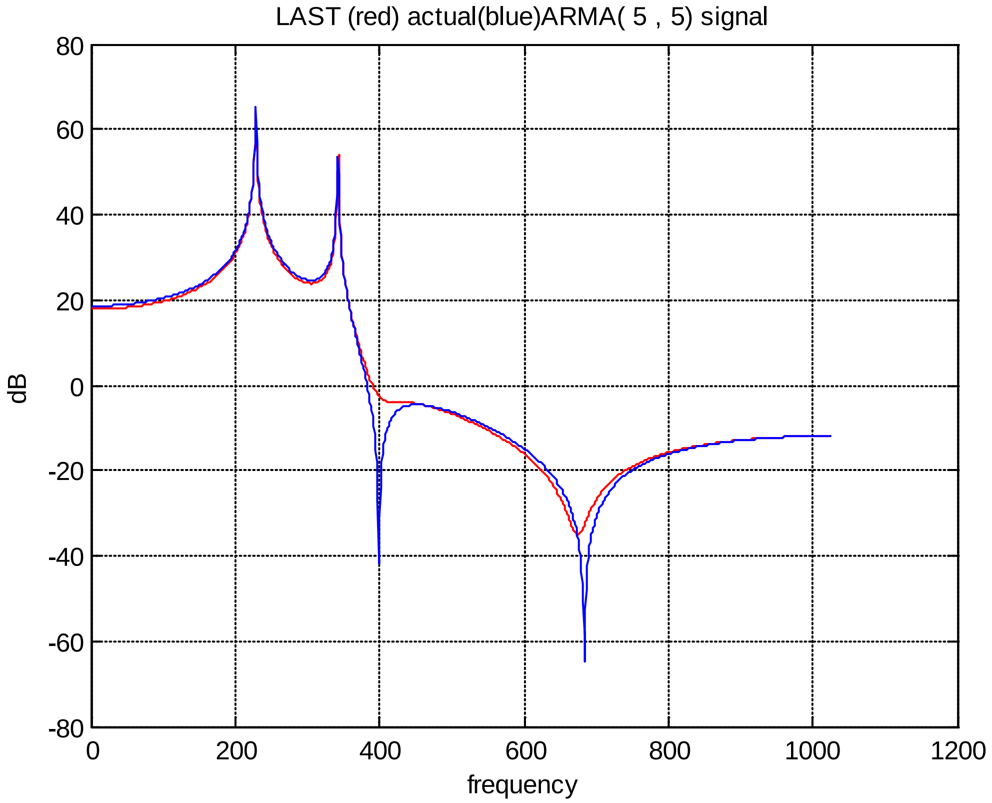
AR orden 20



METODO LAST

Numerator coefficients
1.0000 0.3160 1.3160 0.3160
1.0000

Denominator coefficients
1.0000 -2.5296 3.5250 -2.5245
0.9960



Noise power
1.0855

Numerator coefficients
1.0000 0.2688 1.2124 0.1896 0.8213

Coefficients denorminator
1.0000 -2.5286 3.5232 -2.5237 0.9958