



DEPARTAMENTO DE ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES

Processos Estocásticos

(Apontamentos de Processamento Digital de Sinal)

Fernando Santos Perdigão
Coimbra, 2007

Índice

1. Variáveis aleatórias.....	3
1.1 Média e Variância.....	8
1.2 Duas Variáveis Aleatórias	8
1.3 Sequências de Variáveis Aleatórias.....	13
2. Processos estocásticos	15
2.1 Estacionaridade em sentido lato (WSS)	17
2.2 Ergodicidade	18
2.3 Espectro de Potência.....	19
2.4 Filtragem de Processos Estocásticos	21
2.4.1 Processos regulares.....	26
2.4.3 Teorema da decomposição de Wold.....	29
3. Estimação Espectral de Potência	30
3.1 Noção de estimador	30
3.2 Estimativa da autocorrelação.....	31
3.2 Periodograma.....	32
3.3 Métodos de suavização	36
A. Método de Bartlett (médias de periodogramas).....	36
B. Método de Welch.....	37
C. Método de Blackman-Tuckey.....	38
3.4 Método da Máxima Entropia.....	39
4. Filtros Óptimos	42
4.1 Filtro FIR de Wiener	44
4.2 Aplicações	47
a) Cancelamento de eco acústico ou eléctrico	47
b) Redução de ruído.....	47
c) Cancelamento de ruído (melhoramento do sinal).....	48
d) Predição	49

1. Variáveis aleatórias

Função densidade de probabilidade

Definição: $F_x(a) = \Pr\{x \leq a\}$

Propriedades:

$$F_x(-\infty) = 0$$

$$F_x(+\infty) = 1$$

$$\Pr\{x > a\} = 1 - F_x(a)$$

$$\Pr\{a < x \leq b\} = F_x(b) - F_x(a)$$

Função densidade de probabilidade (PDF)

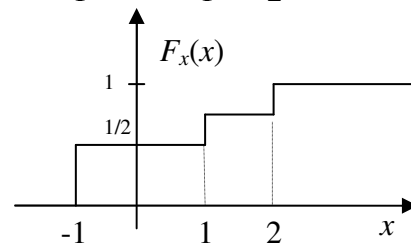
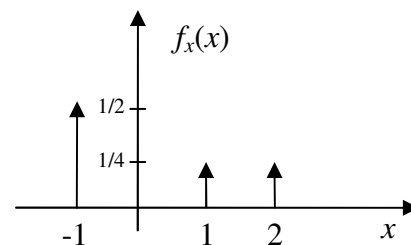
$$f_x(x) = \frac{dF_x(x)}{dx}$$

$$f_x(x) \geq 0; \int_{-\infty}^{+\infty} f_x(x) dx = 1$$

Exemplos:

Exemplo de PDF de variável aleatória (VA) discreta

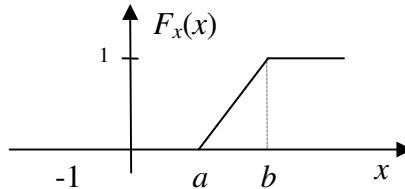
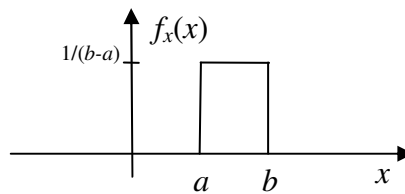
$$f_x(a) = \frac{1}{2} \delta(a+1) + \frac{1}{4} \delta(a-1) + \frac{1}{4} \delta(a-2)$$



Exemplo de VA contínua:
(densidade uniforme)

$$f_x(x) = \begin{cases} \frac{1}{b-a} & , \quad a \leq x < b \\ 0 & , \quad - \end{cases}$$

$$F_x(x) = \int_{-\infty}^x f_x(\alpha) d\alpha$$



Exemplos de distribuições usuais

-Gaussiana (normal) $f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) = \mathcal{N}(x; \mu, \sigma^2)$

-Lognormal $f_x(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln(x)-\mu)^2\right) \cdot u(x)$

-Rayleigh $f_x(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) \cdot u(x)$

-Maxwell $f_x(x) = \frac{2x^2}{\sigma^3\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \cdot u(x)$

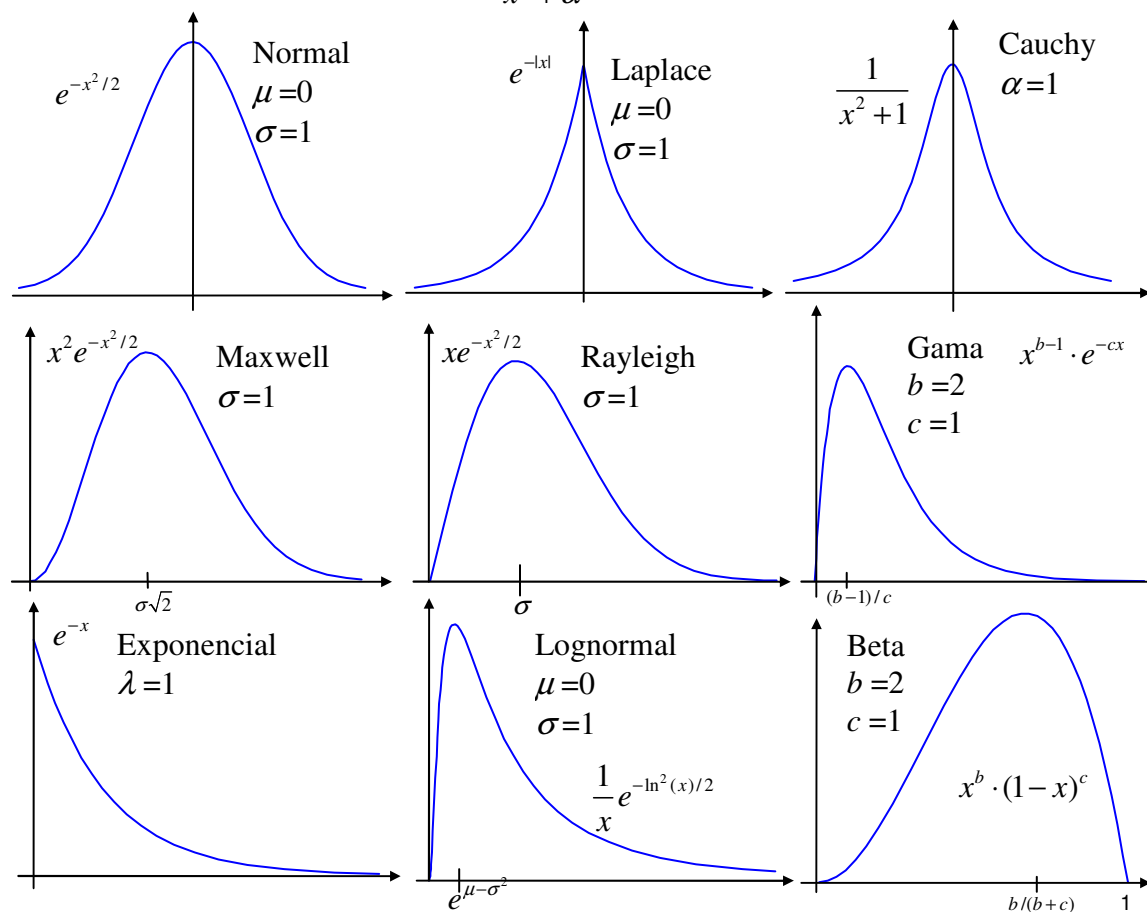
-Laplace $f_x(x) = \frac{1}{2\sigma} \exp\left(-\frac{1}{\sigma}|x-\mu|\right)$

-Gama (Erlang) $f_x(x) = \frac{c^b}{\Gamma(b)} x^{b-1} \cdot e^{-cx} \cdot u(x)$

-Beta $f_x(x) = \frac{1}{B(c+1, b+1)} x^b \cdot (1-x)^c \cdot u(x) \cdot u(1-x)$

-Exponencial $f_x(x) = \lambda e^{-\lambda x} \cdot u(x)$

-Cauchy $f_x(x) = \frac{\alpha/\pi}{x^2 + \alpha^2}$



Função de uma VA ([Papoulis, 91])

Se $y = g(x)$,

então $f_y(y) = \sum_n \frac{f_x(x_n)}{|g'(x_n)|}$, onde x_n são as raízes de $y = g(x)$.

Exemplos

1. $y = ax + b$; $y' = a$;

$$\text{Uma única raiz: } x = \frac{y-b}{a} \Rightarrow f_y(y) = \frac{1}{|a|} f_x\left(\frac{y-b}{a}\right)$$

-Se x é uniforme em $[0,1]$, $f_x(x) = u(x) \cdot u(1-x)$, y é uniforme em $[b, a+b]$.

-Se x é normal com média zero e var. unitária ($x \sim \mathcal{N}(0,1)$), y é também normal, com média b e variância a^2 .

- Se $x \sim \mathcal{N}(\mu, \sigma^2)$, $y = (x-\mu)/\sigma$ tem média nula e variância unitária.

2. $y = \frac{1}{x}$; $y' = -\frac{1}{x^2} = -y^2$;

$$\text{Uma única raiz: } x = \frac{1}{y} \Rightarrow f_y(y) = \frac{1}{y^2} f_x\left(\frac{1}{y}\right)$$

- Se x tem distribuição de Cauchy, y também.

3. $y = x^2$; $y' = 2x = 2\sqrt{y}$

Duas raízes se $y \geq 0$: $x = \pm\sqrt{y}$. Para $y < 0$, $f_y(y) = 0$.

$$\Rightarrow f_y(y) = \frac{1}{2\sqrt{y}} \left(f_x(\sqrt{y}) + f_x(-\sqrt{y}) \right) \cdot u(y)$$

Se $x \sim \mathcal{N}(0,1)$, isto é, $f_x(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$, então:

$$f_y(y) = \frac{1}{\sqrt{2\pi y}} \exp(-y/2) \cdot u(y) \quad (\text{Gama com } b=1/2, c=1/2, \text{ pois } \Gamma(\frac{1}{2}) = \sqrt{\pi}).$$

4. $y = e^x$; $y' = e^x = y$

Uma raiz se $y \geq 0$: $x = \ln(y)$. Para $y < 0$, $f_y(y) = 0$.

$$\Rightarrow f_y(y) = \frac{1}{y} f_x(\ln(y)) \cdot u(y)$$

Se $x \sim \mathcal{N}(0, \sigma^2)$, então $f_y(y) = \frac{1}{\sigma y \sqrt{2\pi}} \exp(-\frac{1}{2} \ln^2(y) / \sigma^2) \cdot u(y)$, isto é,

a distribuição de y é lognormal.

Histogramas

Dada uma amostra com N pontos x_n de uma VA x com PDF $f_x(x)$, determina-se quais os valores extremos, x_{\min} , x_{\max} . Divide-se o intervalo de x_{\min} a x_{\max} em N_b subintervalos (bins) de largura $\Delta_x = (x_{\max} - x_{\min}) / (N_b - 1)$. Conta-se o número de ocorrências em cada bin.

Para amostras grandes, obtém-se aproximação da PDF.

Exemplo:

Gerar 10000 números aleatórios de uma distribuição normal de média zero e variância unitária.

Fazer o histograma.

Em matlab:

```
xn=randn(1,10000);  
hist(xn,100) %com 100 bins
```

ou:

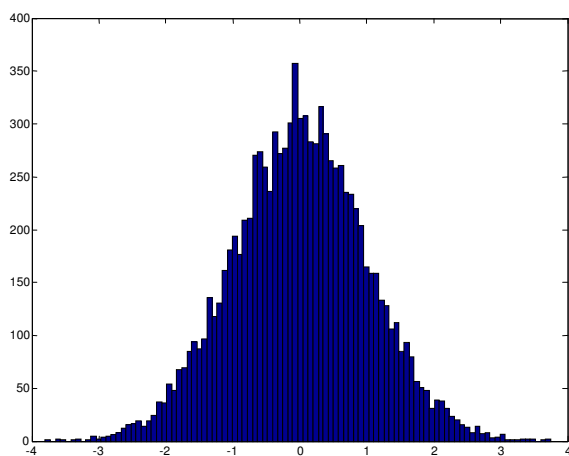
```
[count,bin]=hist(xn,100);  
plot(bin,count)
```

Comparação com a PDF esperada:

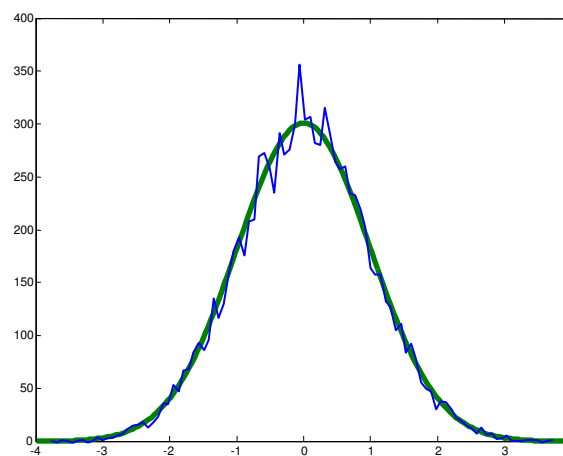
- A área do histograma é $N\Delta_x$ pois os bins são todos iguais e o nº total de contagens é N .
- A área da PDF é 1. Para justapor a PDF no histograma, multiplicar $f_x(x)$ por $N\Delta_x$.

Exemplo com Matlab:

```
x=linspace(-4,4,200);  
fx=exp(-0.5*x.^2)/sqrt(2*pi);  
Dx=bin(2)-bin(1);  
plot(bin,count,x,fx*10000*Dx);
```



a) Histograma com barras



b) Histograma com PDF justaposta.

Exemplo de função de VA com Matlab:

x uniforme entre 1 e 2: $f_x(x) = u(x-1) \cdot u(2-x)$
 $y = x^2$.

y irá estar distribuída entre 1 e 4. $f_x(\sqrt{y}) = u(\sqrt{y}-1) \cdot u(2-\sqrt{y}) = u(y-1) \cdot u(4-y)$

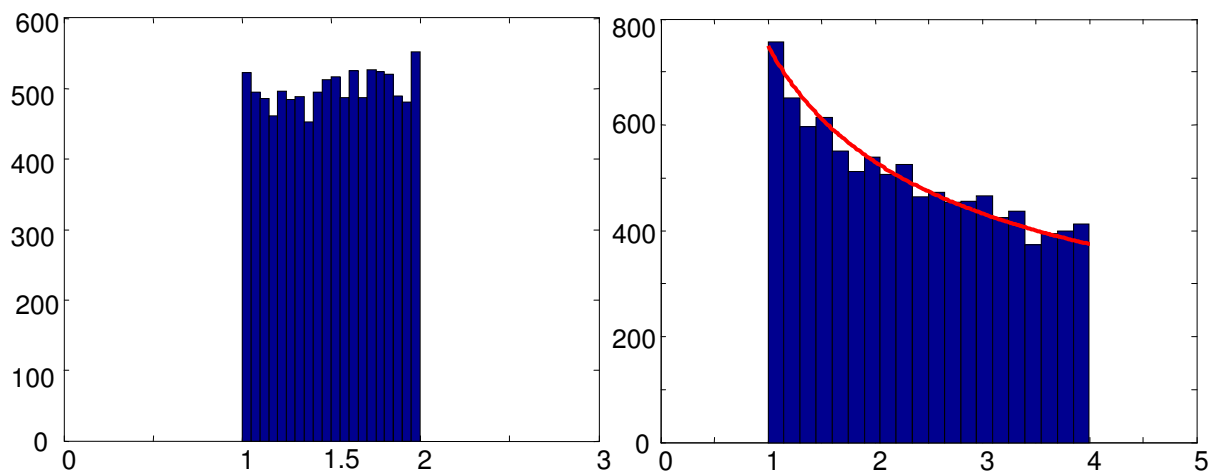
pois $u(2-\sqrt{y}) = 1$ se $2-\sqrt{y} \geq 0$ e 0 caso contrário. Ou seja, $u(2-\sqrt{y}) = 1$ se $y \leq 4$.

Assim,

$$f_y(y) = \frac{1}{2\sqrt{y}} u(y-1) u(4-y).$$

Verificação experimental:

```
xn=rand(1,10000)+1; %uniforme entre 1 e 2
hist(xn,20); %histograma com 20 bins.
yn=xn.^2;
hist(yn,20);
y=linspace(1,4,200);
fy=0.5./sqrt(y);
Dx=bin(2)-bin(1);
hold on, plot(y,fy*10000*Dx,'r'); hold off
```



1.1 Média e Variância

O valor esperado ou *média* da VA x é, por definição

$$\mu_x = E\{x\} = \int_{-\infty}^{+\infty} x \cdot f_x(x) \cdot dx$$

Para VAs discretas: $f_x(x) = \sum_i p_i \cdot \delta(x - x_i)$,

logo
$$\mu_x = E\{x\} = \sum_i p_i \int_{-\infty}^{+\infty} x \cdot \delta(x - x_i) \cdot dx = \sum_i p_i x_i$$

A média da VA x corresponde ao “centro de massa” da sua função densidade ((PDF).

A média de $y = g(x)$ é

$$E\{g(x)\} = \int_{-\infty}^{+\infty} g(x) \cdot f_x(x) \cdot dx .$$

Variância

$$\sigma_x^2 = \text{Var}(x) = E\{(x - \mu_x)^2\} = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f_x(x) \cdot dx$$

O operador “esperança matemática”, $E\{x\}$, é linear, isto é, dadas as VAs x e y ,

$$E\{ax + by\} = a \cdot E\{x\} + b \cdot E\{y\} .$$

Assim,

$$\sigma_x^2 = E\{(x - \mu_x)^2\} = E\{x^2 - 2x\mu_x + \mu_x^2\} = E\{x^2\} - 2\mu_x E\{x\} + \mu_x^2$$

ou

$$\sigma_x^2 = E\{x^2\} - \mu_x^2, \text{ ou ainda: } \sigma_x^2 = E\{x^2\} - E^2\{x\} ; \boxed{E\{x^2\} = \sigma_x^2 + \mu_x^2} .$$

Momentos: $m_n = E\{x^n\}$

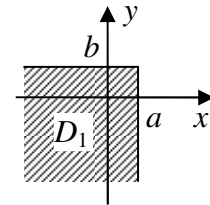
1.2 Duas Variáveis Aleatórias

A distribuição conjunta de duas VAs x e y , $F_{xy}(a, b)$, é a probabilidade do evento $\{x \leq a, y \leq b\}$:

$$F_{xy}(a, b) = \Pr\{x \leq a, y \leq b\} = \Pr\{(x, y) \in D_1\}$$

A função densidade conjunta é

$$f_{xy}(x, y) = \frac{\partial^2 F_{xy}(x, y)}{\partial x \partial y}$$



Estatísticas marginais:

$$f_x(x) = \int_{-\infty}^{+\infty} f_{xy}(x, y) \cdot dy \quad ; \quad f_y(y) = \int_{-\infty}^{+\infty} f_{xy}(x, y) \cdot dx$$

Momentos conjuntos

A **correlação** entre as VAs x e y é o momento conjunto de 2ª ordem

$$r_{xy} = E\{x \cdot y^*\}$$

Covariância:

$$c_{xy} = \text{Cov}(x, y) = E\{(x - \mu_x) \cdot (y - \mu_y)^*\};$$
$$c_{xy} = r_{xy} - \mu_x \cdot \mu_y^*$$

A covariância iguala a correlação se as médias das VAs forem zero.

Coefficiente de correlação:

$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{c_{xy}}{\sigma_x \cdot \sigma_y} = \frac{r_{xy} - \mu_x \cdot \mu_y^*}{\sigma_x \cdot \sigma_y}$$

O coeficiente de correlação é um número com módulo inferior a 1,

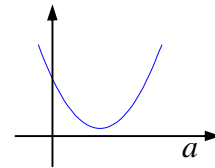
$$\rho_{xy}^2 \leq 1.$$

Prova: $(a(x - \mu_x) - (y - \mu_y))^2 \geq 0, \forall a,$

$$E\{(a(x - \mu_x) - (y - \mu_y))^2\} = a^2 \sigma_x^2 - 2a \cdot c_{xy} + \sigma_y^2 \geq 0.$$

Esta equação de 2º grau em a não pode ter raízes reais (quando muito duplas). O seu discriminante tem de ser não negativo:

$$4c_{xy}^2 - 4\sigma_x^2 \sigma_y^2 \leq 0 \text{ ou } \rho_{xy}^2 \leq 1.$$



Ou: $a^2 \sigma_x^2 - 2a \cdot c_{xy} + \sigma_y^2 = \sigma_x^2 \left(a - \frac{c_{xy}}{\sigma_x^2}\right)^2 + \sigma_y^2 - \frac{c_{xy}^2}{\sigma_x^2}$ é uma parábola virada para cima com centro em

$a = \frac{c_{xy}}{\sigma_x^2}$ e valor mínimo $\sigma_y^2 - \frac{c_{xy}^2}{\sigma_x^2}$, que não pode ser negativo. Logo

$$\sigma_y^2 - \frac{c_{xy}^2}{\sigma_x^2} \geq 0 \Rightarrow c_{xy}^2 \leq \sigma_x^2 \sigma_y^2 \Rightarrow \rho_{xy}^2 \leq 1.$$

Definição: Duas VAs dizem-se **não correlacionadas** se $\rho_{xy} = 0$ (ou $c_{xy} = 0$).

Significa neste caso que $r_{xy} = \mu_x \cdot \mu_y^* = E\{x\} \cdot E\{y^*\}$ (a correlação torna-se separável).

Definição: Duas VAs dizem-se **ortogonais** se $r_{xy} = 0$ e anota-se como $x \perp y$.

VAs x e y não correlacionadas não são necessariamente ortogonais, mas VAs de médias nulas não correlacionadas são sempre ortogonais.

Interpretação de VA como vectores

As VAs podem ser interpretadas como vectores num espaço abstracto. Assim $E\{xy\}$ corresponde ao **produto interno** dos dois vectores $E\{x^2\}$ e $E\{y^2\}$ ao quadrado dos seus módulos. Assim,

$$\frac{E\{xy\}}{\sqrt{E\{x^2\}E\{y^2\}}} = \cos(\theta)$$

é o cosseno do ângulo entre os vectores. Este valor será 1 se os vectores estiverem alinhados e será zero se forem perpendiculares (ortogonais). Assim, como $\cos^2(\theta) \leq 1$, $r_{xy}^2 \leq \sigma_x^2 \sigma_y^2$ é a chamada desigualdade de cosseno.

Exemplos de correlação de VAs :

Seja $x \sim \mathcal{N}(0,1)$ e $y = 2x + 3$.

As duas VAs estão obviamente totalmente *correlacionadas*. Teremos que:

$$\mu_x = 0 ; \sigma_x^2 = 1$$

$$\mu_y = E\{y\} = 2\mu_x + 3 = 3$$

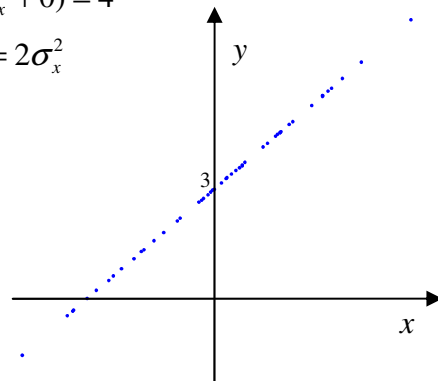
$$\sigma_y^2 = E\{(y - \mu_y)^2\} = E\{(2x)^2\} = 4E\{x^2\} = 4(\sigma_x^2 + 0) = 4$$

$$r_{xy} = E\{x \cdot y\} = E\{x \cdot (2x + 3)\} = 2E\{x^2\} + 3\mu_x = 2\sigma_x^2$$

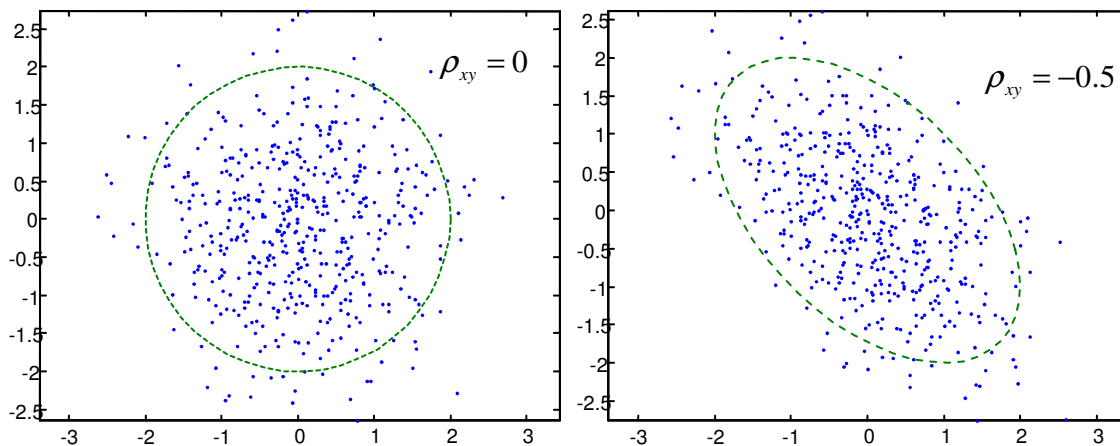
$$c_{xy} = r_{xy} - \mu_x \mu_y = r_{xy} = 2\sigma_x^2$$

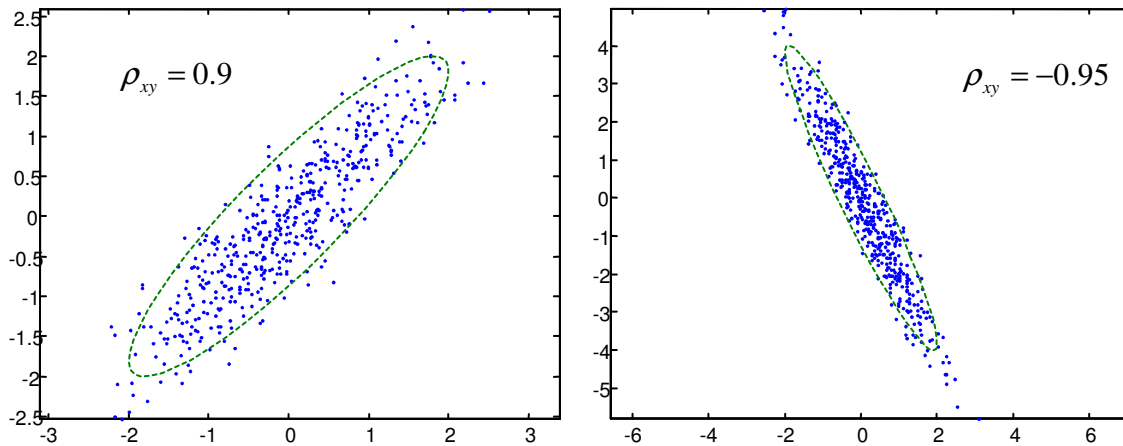
$$\rho_{xy} = \frac{c_{xy}}{\sigma_x \cdot \sigma_y} = \frac{2\sigma_x^2}{\sigma_x \cdot \sigma_y} = \frac{2\sigma_x}{\sigma_y} = 1 .$$

Figura: “scatter plot” de 100 pontos (x,y)
(todos os pontos estão na linha $y = 2x + 3$)



Scatter plots de VAs normais com vários coeficientes de correlação (500 pontos):





De uma forma geral¹, a não correlação entre VAs x e y manifesta-se por um espalhamento dos pontos (x,y) no interior de um círculo ou de uma elipse. À medida que o coeficiente de correlação aumenta (em módulo) este espalhamento concentra-se em elipses mais excêntricas. O eixo principal destas elipses tem inclinação positiva para coeficientes de correlação positivos e inclinação negativa para coeficientes de correlação negativos.

Independência estatística

Duas VAs x e y dizem-se estatisticamente independentes se a sua função densidade de probabilidade conjunta for separável, isto é, se

$$f_{xy}(x, y) = f_x(x) \cdot f_y(y)$$

- Este conceito é mais “forte” que o de VAs não correlacionadas, uma vez que a correlação é apenas uma estatística de 2ª ordem.
- Se as VAs são independentes então $E\{xy^*\} = E\{x\}E\{y^*\}$.
- Se as VAs são independentes então são não correlacionadas. O contrário não é necessariamente verdade, a não ser no caso de VAs conjuntamente normais.

PDF Gaussiana multidimensional

A função densidade Gaussiana a duas dimensões é mais facilmente definida como um caso particular ($N=2$) da função Gaussiana multidimensional. Seja \mathbf{x} um vector de N VAs. Se estas VAs forem conjuntamente normais, então a sua PDF é

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_{\mathbf{x}}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}_{\mathbf{x}}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

onde $\mathbf{C}_{\mathbf{x}}$ é a matriz de covariância de \mathbf{x} e $\boldsymbol{\mu}$ o vector de médias.

¹ Isto é verdade para distribuições normais, embora estes “elipsóides de concentração” [Therrien, 92, Cap. 2] sejam igualmente úteis para outras distribuições, para dar uma ideia do espalhamento da PDF.

No caso de $N=2$, com $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ e \mathbf{C}_x , dada por

$$\mathbf{C}_x = E \left\{ \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \begin{bmatrix} x - \mu_x & y - \mu_y \end{bmatrix} \right\} = E \left\{ \begin{bmatrix} (x - \mu_x)^2 & (x - \mu_x)(y - \mu_y) \\ (y - \mu_y)(x - \mu_x) & (y - \mu_y)^2 \end{bmatrix} \right\}$$

$$= \begin{bmatrix} \sigma_x^2 & c_{xy} \\ c_{xy} & \sigma_y^2 \end{bmatrix};$$

$$|\mathbf{C}_x| = \sigma_x^2 \sigma_y^2 - c_{xy}^2 = \sigma_x^2 \sigma_y^2 (1 - \rho_{xy}^2);$$

$$\mathbf{C}_x^{-1} = \begin{bmatrix} \sigma_x^2 & c_{xy} \\ c_{xy} & \sigma_y^2 \end{bmatrix}^{-1} = \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho_{xy}^2)} \begin{bmatrix} \sigma_y^2 & -c_{xy} \\ -c_{xy} & \sigma_x^2 \end{bmatrix},$$

então

$$f_{xy}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp \left(-\frac{1}{2(1-\rho_{xy}^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho_{xy} \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right)$$

Daqui se verifica que se $\rho_{xy} = 0$ então esta PDF torna-se separável:

$$f_{xy}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left(-\frac{1}{2} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right)$$

$$= f_x(x) \cdot f_y(y)$$

Logo: VAs conjuntamente **normais** e não correlacionadas são independentes.

Caso geral: Quando as VAs são conjuntamente normais e a matriz de covariância é diagonal, então as VAs são independentes.

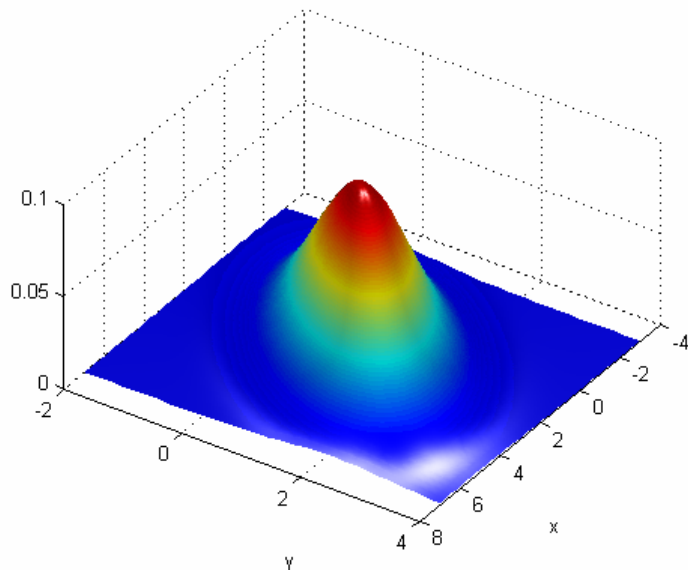
Fazer $f_{xy}(x,y)=C^{te}$, equivale a tornar o expoente da Gaussiana constante, o que resulta numa equação de uma elipse (elipsóide para $N>2$) que será centrada se as VAs forem não correlacionadas.

Exemplo de uma Gaussiana com:

$$\sigma_x^2 = 2 ; \sigma_y^2 = 1$$

$$\mu_x = 2 ; \mu_y = 1;$$

$$\rho_{xy} = 1/2$$

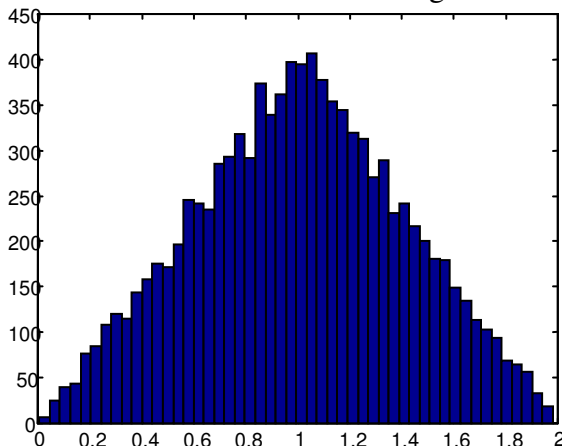


Independência e convolução

Se duas VAs x e y são independentes, então a sua soma, $z = x + y$, tem função densidade que é a convolução das PDFs de x e de y :

$$f_{xy}(x, y) = f_x(x)f_y(y)$$
$$z = x + y \Rightarrow f_z(z) = f_x(x) * f_y(y)$$

Exemplo: x e y independentes e uniformes entre 0 e 1. Histograma de $x+y$ com 10000 números:



Propriedade correspondente para VAs não correlacionadas:

Se x e y são VAs **não correlacionadas**, então a variância da sua soma iguala a soma das variâncias das VAs:

$$z = x + y$$
$$c_{xy} = 0 \Rightarrow \sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

1.3 Sequências de Variáveis Aleatórias

Um vector aleatório define-se como $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

onde as suas componentes x_i , $i = 1..n$, são VAs. A função de distribuição conjunta define-se como

$$F_{\mathbf{x}}(\mathbf{a}) = F_{\mathbf{x}}(a_1, a_2, \dots, a_n) = \Pr\{x_1 < a_1, x_2 < a_2, \dots, x_n < a_n\}$$

e a densidade conjunta como

$$f_{\mathbf{x}}(\mathbf{x}) = \frac{\partial^n F_{\mathbf{x}}(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_n}.$$

A média é o vector

$$\boldsymbol{\mu} = E\{\mathbf{x}\} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}.$$

Teremos agora matrizes de correlação,

$$\mathbf{R}_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^{*T}\} = E\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1^* & x_2^* & \cdots & x_n^* \end{bmatrix} \right\} = E\left\{ \begin{bmatrix} |x_1|^2 & x_1 x_2^* & \cdots & x_1 x_n^* \\ x_2 x_1^* & |x_2|^2 & & \\ \vdots & & \ddots & \\ x_n x_1^* & & & |x_n|^2 \end{bmatrix} \right\},$$

com elementos

$$r_{ij} = E\{x_i x_j^*\},$$

e a matriz de covariância

$$\mathbf{C}_{\mathbf{x}} = E\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^{*T}\} = \mathbf{R}_{\mathbf{x}} - \boldsymbol{\mu}\boldsymbol{\mu}^{*T}$$

com elementos $c_{ij} = E\{(x_i - \mu_i)(x_j - \mu_j)^*\} = r_{ij} - \mu_i \mu_j^*.$

As n VAs dizem-se **mutuamente não correlacionadas** se $c_{ij} = 0$ para $i \neq j$. Neste caso a matriz de covariância é diagonal. Se as VAs forem não correlacionadas e conjuntamente normais então são independentes.

A matriz de covariância iguala a matriz de correlação se todas as n VAs tiverem média nula.

Uma matriz de correlação é não negativamente definida (os seus valores próprios nunca são negativos).

2. Processos estocásticos

O valor de uma VA x é um número associado ao resultado de uma experiência aleatória.
O valor de um processo estocástico é um sinal $x[n]$ associado a uma experiência aleatória.

Exemplo [Hayes,96]:

Ao lançamento de um dado associa-se uma VA A cujo valor é um inteiro entre 1 e 6, correspondente à face do dado virada para cima. Forma-se depois o sinal

$$x[n] = A \cos(\omega_0 n)$$

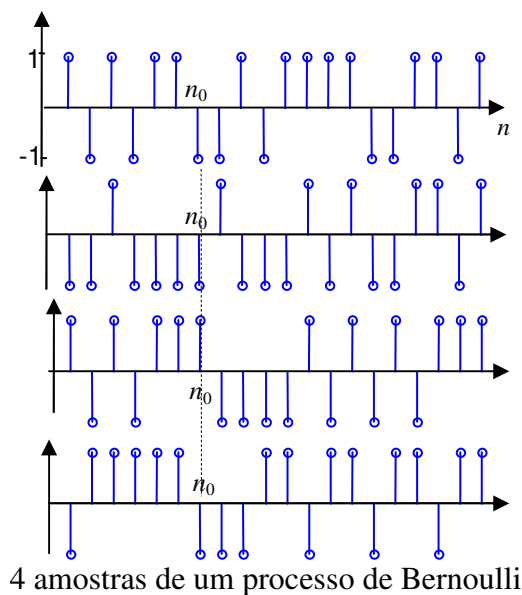
Cria-se assim um processo estocástico (PE) que consiste num conjunto de 6 sinais sinusoidais diferentes e igualmente prováveis.

Processo de Bernoulli

Uma moeda é lançada repetidamente e de forma independente. Associa-se ao valor $x[n]=1$ quando o n -ésimo resultado é “cara” e $x[n] = -1$ quando é “coroa”.

Uma nova experiência resulta num sinal $x[n]$ diferente, mas com as mesmas propriedades.

Para um dado instante, $n = n_0$, obtemos para cada experiência uma VA com valor $x[n_0]$.



A **média** de um PE $x[n]$ é a média da VA $x[n]$ para cada instante n . Depende, portanto do tempo e é indicada como

$$\mu_x[n] = E\{x[n]\}.$$

(Trata-se de uma sequência determinística).

A **variância** de um PE obtém-se de forma idêntica:

$$\sigma_x^2[n] = E\left\{|x[n] - \mu_x[n]|^2\right\}$$

Para um par de índices (k, l) do PE, define-se a **autocorrelação** e a **autocovariância**:

$$r_x[k, l] = E\left\{x[k] \cdot x^*[l]\right\}$$

$$c_x[k, l] = E\left\{(x[k] - \mu_x[k]) \cdot (x[l] - \mu_x[l])^*\right\}$$

que se relacionam com as duas VAs $x[k]$ e $x[l]$. Quando $k=l=n$ a autocorrelação é média quadrática da VA $x[n]$ e a autocovariância a sua variância.

Dados dois PE podemos igualmente definir a **correlação cruzada** entre eles:

$$r_{xy}[k, l] = E\left\{x[k] \cdot y^*[l]\right\}$$

bem como a **covariância cruzada**:

$$c_{xy}[k, l] = E\left\{(x[k] - \mu_x[k]) \cdot (y[l] - \mu_y[l])^*\right\}$$

$$= r_{xy}[k, l] - \mu_x[k] \mu_y^*[l]$$

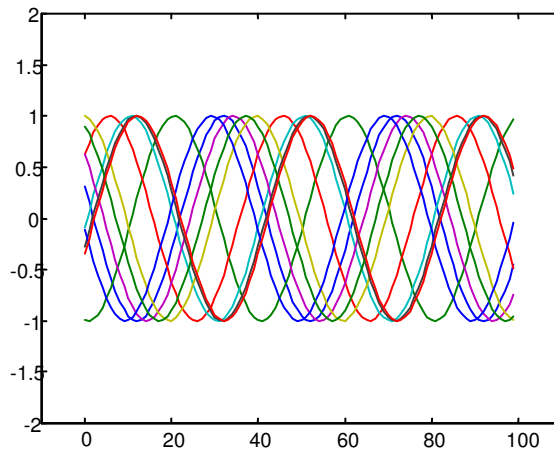
PE harmónico

$$x[n] = A \sin(\omega_0 n + \varphi)$$

onde A e ω_0 são constantes de φ é uma VA uniformemente distribuída entre $-\pi$ e π .

10 amostras deste PE:

```
fi=pi*(2*rand(1,10)-1);
A=1;
w0=pi/20;
x=zeros(10,100);
n=0:99;
for k=1:10
    x(k,:) = A*sin(w0*n+fi(k));
end
plot(n,x);
```



Média: com $g(\varphi) = \sin(\omega_0 n + \varphi) \Rightarrow E\{\sin(\omega_0 n + \varphi)\} = \int_{-\pi}^{\pi} g(\varphi) \frac{1}{2\pi} d\varphi = 0$.

Autocorrelação:

$$\begin{aligned}
 r_x[k, l] &= E\{x[k] \cdot x^*[l]\} \\
 &= A^2 E\{\sin(\omega_0 k + \varphi) \sin(\omega_0 l + \varphi)\} \\
 &= -\frac{A^2}{4} E\left\{\left(e^{j(\omega_0 k + \varphi)} - e^{-j(\omega_0 k + \varphi)}\right)\left(e^{j(\omega_0 l + \varphi)} - e^{-j(\omega_0 l + \varphi)}\right)\right\} \\
 &= -\frac{A^2}{4} E\left\{e^{j(\omega_0(k+l)+2\varphi)} - e^{j\omega_0(k-l)} - e^{-j\omega_0(k-l)} + e^{-j(\omega_0(k+l)+2\varphi)}\right\} \\
 &= \frac{A^2}{2} E\{\cos(\omega_0(k-l))\} - \frac{A^2}{2} E\{\cos(\omega_0(k+l)+2\varphi)\}
 \end{aligned}$$

A média de um cosseno (2º termo) é nula. O 1º termo é constante pois não depende de φ . Logo

$$r_x[k, l] = \frac{A^2}{2} \cos(\omega_0(k-l))$$

Neste caso a autocorrelação depende apenas da diferença entre os índices e não da origem temporal:

$$\begin{aligned}
 r_x[k, l] &= r_x[k-l, 0] \\
 &= r_x[n+k-l, n] \quad , \quad \forall n
 \end{aligned}$$

\Rightarrow PE estacionário (em sentido lato).

2.1 Estacionaridade em sentido lato (WSS)

1. Média constante: $\mu_x[n] = \mu_x$
2. Autocorrelação depende apenas da diferença dos índices $r_x[k, l] = r_x[k-l]$
3. Variância finita: $c_x[0] = r_x[0] - |\mu_x|^2 < \infty$

\rightarrow a autocorrelação passa a ser definida com apenas um índice: $r_x[k]$.

$$\rightarrow r_x[k] = E\{x[n+k]x^*[n]\} \quad , \quad \forall n$$

$$\rightarrow c_x[k] = E\{(x[n+k] - \mu_x) \cdot (x[n] - \mu_x)^*\} = r_x[k] - |\mu_x|^2 \quad , \quad \forall n$$

Propriedades

$$r_x[k] = r_x^*[-k] \quad (\text{autocorrelação é par para PE reais})$$

$$r_x[0] = E\{|x[n]|^2\} \geq 0 \quad (\text{potência do PE})$$

$$|r_x[k]| \leq r_x[0] \quad (\text{módulo do valor máximo é a potência})$$

Se $r_x[k_0] = r_x[0]$, então $r_x[k]$ é periódica com período k_0 .

Dados dois PE estacionários WSS, então

$$\begin{aligned} r_{xy}[k] &= r_{yx}^*[-k] \\ |r_{xy}[k]|^2 &\leq r_x[0] \cdot r_y[0] \\ |c_{xy}[k]|^2 &\leq c_x[0] \cdot c_y[0] \end{aligned}$$

2.2 Ergodicidade

Um PE WSS diz-se ergódico se as suas propriedades poderem ser estimadas a partir de **uma única realização** do processo.

Para estimar a média $\mu_x[n]$ de um PE dadas K realizações, $x_i[n]$, podemos usar o seguinte estimador

$$\hat{\mu}_x[n] = \frac{1}{K} \sum_{i=1}^K x_i[n] .$$

Dada apenas UMA realização do PE pretende-se estimar a média através da média temporal em N pontos

$$m_N = \hat{\mu}_x(N) = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

Se $\lim_{N \rightarrow \infty} E\{|\hat{\mu}_x(N) - \mu_x|^2\} = 0$

então o PE $x[n]$ diz-se ergódico para a média.

Para a convergência no sentido da média quadrática é necessário que a média da VA m_N tenda para a média e que a sua variância tenda para zero à medida que mais pontos são tomados:

1. $\lim_{N \rightarrow \infty} E\{\hat{\mu}_x(N)\} = \mu_x$
2. $\lim_{N \rightarrow \infty} \text{Var}\{\hat{\mu}_x(N)\} = 0$

Mas

$$\begin{aligned} \text{Var}\{\hat{\mu}_x(N)\} &= E\{|\hat{\mu}_x(N) - \mu_x|^2\} \\ &= E\left\{\left|\frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu_x)\right|^2\right\} \\ &= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} E\{(x[m] - \mu_x)(x[n] - \mu_x)^*\} \\ &= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} c_x[m-n] \end{aligned}$$

Pode colocar-se na forma:

$$\text{Var}\{\hat{\mu}_x(N)\} = \frac{1}{N} \sum_{k=-N+1}^{N-1} \left(1 - \frac{|k|}{N}\right) c_x[k]$$

Uma condição necessária e suficiente para que um PE WSS seja ergódico para a média é que

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=0}^{N-1} c_x[k] = 0$$

Uma condição suficiente para que um PE WSS seja ergódico para a média é que

$$\lim_{k \rightarrow \infty} c_x[k] = 0.$$

Uma consequência desta condição é que

$$\lim_{k \rightarrow \infty} r_x[k] = |\mu_x|^2.$$

Exemplo [Mitra, 2002]

Determinar a média, a variância e autocovariância de um processo estocástico real que tem

autocorrelação $r_x[k] = \frac{10 + 21k^2}{1 + 3k^2}.$

O quadrado da média é $\mu_x^2 = \lim_{k \rightarrow \infty} r_x[k] = 7$. Logo a média é $\mu_x = \sqrt{7}$ ou $\mu_x = -\sqrt{7}$. A média quadrática é $r_x[0] = 10$ e a variância é $\sigma_x^2 = 10 - 7 = 3$.

A autocovariância é $c_x[k] = r_x[k] - \mu_x^2 = \frac{10 + 21k^2}{1 + 3k^2} - 7 = \frac{3}{1 + 3k^2}$ que tende para zero quando k tende para infinito.

2.3 Espectro de Potência

O espectro de potência de um PE WSS é a transformada de Fourier em tempo discreto (DTFT) da autocorrelação

$$P_x(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} r_x[n] \cdot e^{-j\omega n}$$

Como a autocorrelação é complexa simétrica, então o espectro de potência é **real**:

$$P_x(e^{j\omega}) = P_x^*(e^{j\omega})$$

Para PE **reais** a autocorrelação é **real e par**, logo o espectro de potência é **real e par**:

$$P_x(e^{j\omega}) = P_x(e^{-j\omega}).$$

Além disso o espectro de potência não pode ser negativo:

$$\boxed{P_x(e^{j\omega}) \geq 0}.$$

Potência total: $r_x[0] = E\{|x[n]|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_x(e^{j\omega}) d\omega.$

Se o PE $x[n]$ não tiver média nula, então o espectro de potência tem um impulso na origem.

Exemplo (processo autorregressivo de 1ª ordem)

Se $r_x[n] = a^{|n|}, a > 0,$

o espectro de potência é

$$\begin{aligned} P_x(e^{j\omega}) &= \sum_{n=-\infty}^{+\infty} a^{|n|} e^{-j\omega n} = \sum_{n=-\infty}^{-1} a^{-n} e^{-j\omega n} + \sum_{n=0}^{\infty} a^n e^{-j\omega n} \\ &= \frac{ae^{j\omega}}{1 - ae^{j\omega}} + \frac{1}{1 - ae^{-j\omega}} = \frac{1 - a^2}{1 - 2a \cos(\omega) + a^2} \end{aligned}$$

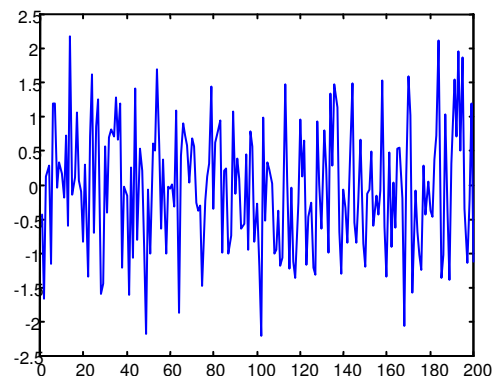
Ruído branco

Se a autocovariância de um PE for um impulso na origem: $c_x[n] = \sigma_x^2 \delta[n]$, o processo diz-se ruído branco pois o seu espectro de potência é constante (além de um eventual impulso em $\omega=0$ se a média do processo não for nula):

$$P_x(e^{j\omega}) = \sigma_x^2 \quad \text{ou} \quad P_x(e^{j\omega}) = \sigma_x^2 + 2\pi |\mu_x|^2 \delta(\omega), |\omega| < \pi$$

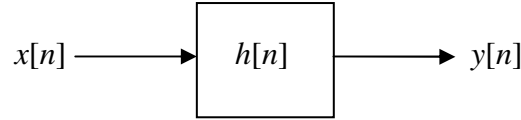
Ruído branco corresponde a uma sequência de VAs não correlacionadas com igual variância σ_x^2 . Uma vez que o ruído branco é definido apenas em termos do momento de 2ª ordem, existe uma infinidade de processos ruído branco: Gaussiano, Bernoulli, etc.

Exemplo de uma amostra de ruído branco Gaussiano de média zero e variância unitária: (sequência de 200 VAs normais não correlacionadas).



2.4 Filtragem de Processos Estocásticos

Frequentemente a entrada de sistemas LTI são processos estocásticos WSS. Quais as propriedades do PE de saída?



$$y[n] = x[n] * h[n] \triangleq \sum_{k=-\infty}^{+\infty} h[k] \cdot x[n-k]$$

Média da saída:

$$\mu_y = E\{y[n]\} = \sum_{k=-\infty}^{+\infty} h[k] \cdot E\{x[n-k]\} = \mu_x \sum_{k=-\infty}^{+\infty} h[k]$$

logo

$$\boxed{\mu_y = \mu_x H(e^{j0})}.$$

Correlação cruzada:

$$\begin{aligned} r_{yx}[m] &= E\{y[n]x^*[n-m]\} = E\left\{\sum_{k=-\infty}^{+\infty} h[k] \cdot x[n-k] \cdot x^*[n-m]\right\} \\ &= \sum_{k=-\infty}^{+\infty} h[k] \cdot E\{x[n-k] \cdot x^*[n-m]\} \\ &= \sum_{k=-\infty}^{+\infty} h[k] \cdot r_x[m-k] \\ &= h[m] * r_x[m] = r_{xy}^*[-m] \end{aligned}$$

Autocorrelação da saída:

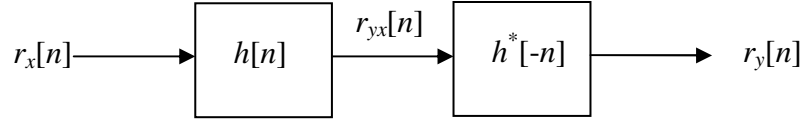
$$\begin{aligned} r_y[m] &= E\{y[n]y^*[n-m]\} = E\left\{\sum_{k=-\infty}^{+\infty} h[k] \cdot x[n-k] \cdot y^*[n-m]\right\} \\ &= \sum_{k=-\infty}^{+\infty} h[k] \cdot E\{x[n-k] \cdot y^*[n-m]\} \\ &= \sum_{k=-\infty}^{+\infty} h[k] \cdot r_{xy}[m-k] \\ &= h[m] * r_{xy}[m] = h[m] * r_{yx}^*[-m] \\ &= h[m] * h^*[-m] * r_x[m] \end{aligned}$$

Definindo a autocorrelação determinística da resposta a impulso,

$$r_h[m] = h[m] * h^*[-m] = \sum_{k=-\infty}^{+\infty} h[k] \cdot h^*[m+k]$$

resulta

$$r_y[m] = r_h[m] * r_x[m] .$$



Resultados em termos de espectro de potência

$$\begin{aligned}
 r_{yx}[m] &= h[m] * r_x[m] & \Leftrightarrow & P_{yx}(e^{j\omega}) = H(e^{j\omega})P_x(e^{j\omega}) \\
 r_{xy}[m] &= h^*[-m] * r_x[m] & \Leftrightarrow & P_{xy}(e^{j\omega}) = H^*(e^{j\omega})P_x(e^{j\omega}) \\
 r_y[m] &= h[m] * r_{xy}[m] & \Leftrightarrow & P_y(e^{j\omega}) = H(e^{j\omega})P_{xy}(e^{j\omega}) \\
 r_y[m] &= h[m] * h^*[-m] * r_x[m] & \Leftrightarrow & P_y(e^{j\omega}) = |H(e^{j\omega})|^2 P_x(e^{j\omega})
 \end{aligned}$$

Como

$$P_y(e^{j\omega}) = |H(e^{j\omega})|^2 P_x(e^{j\omega}) ,$$

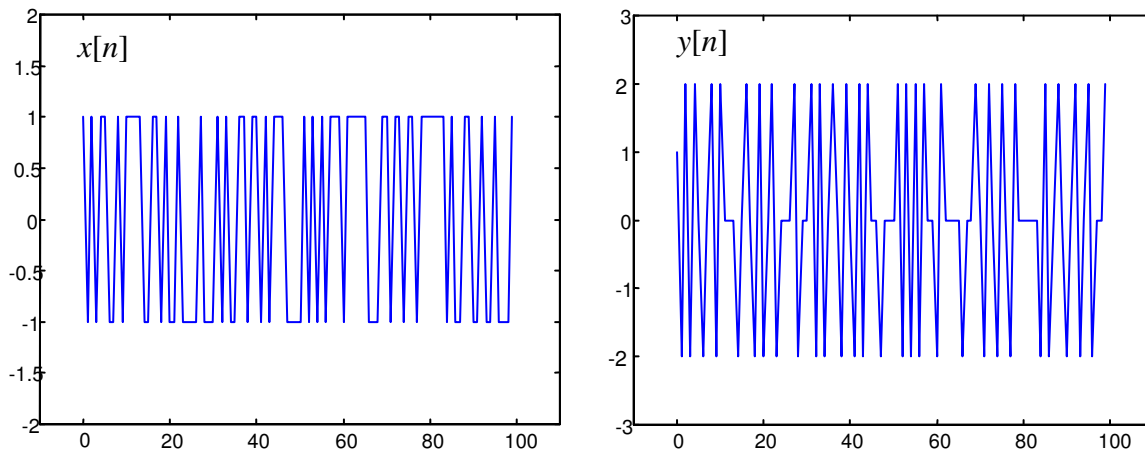
a potência da saída é:

$$E\{|y[n]|^2\} = r_y[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 P_x(e^{j\omega}) d\omega .$$

Exemplo 1:

Faz-se a 1ª diferença de um processo Bernoulli: $y[n] = x[n] - x[n-1]$. O PE $y[n]$ é ruído branco?

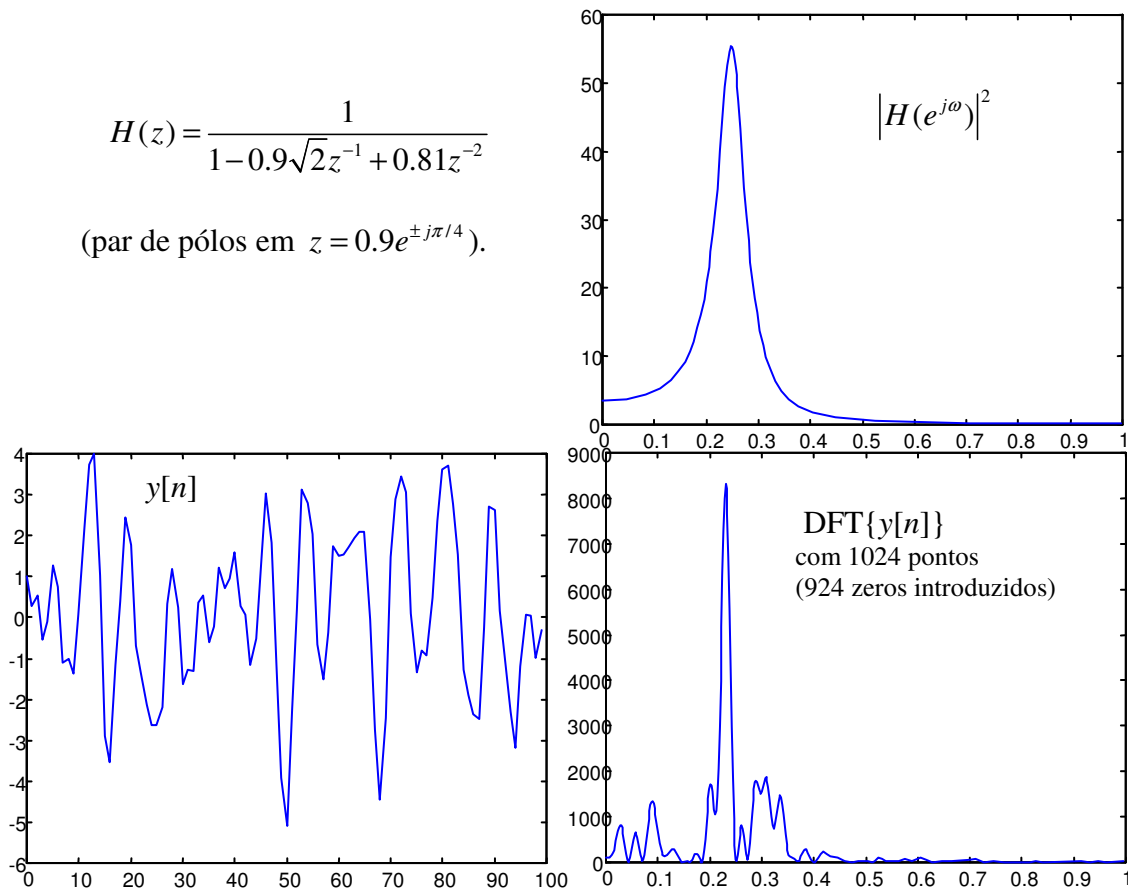
Neste caso $h[n] = \delta[n] - \delta[n-1]$ e $H(e^{j\omega}) = 1 - e^{j\omega}$. Como $r_x[n] = \delta[n]$ (ruído branco de variância unitária), então $P_x(e^{j\omega}) = 1$ e $P_y(e^{j\omega}) = (1 - e^{j\omega})(1 - e^{-j\omega}) = 2(1 - \cos(\omega)) = 4\sin^2(\frac{\omega}{2})$. Logo já não é branco. A média continua zero mas a variância é 2.



A autocorrelação é igual à autocorrelação determinística de $h[n]$ que é $2\delta[n] - (\delta[n-1] + \delta[n+1])$. Podemos “colorir” mais este processo se usarmos, por exemplo o seguinte filtro de 2ª ordem:

$$H(z) = \frac{1}{1 - 0.9\sqrt{2}z^{-1} + 0.81z^{-2}}$$

(par de pólos em $z = 0.9e^{\pm j\pi/4}$).



A correlação entre amostras é agora bem evidente.

Exemplo 2 [Hayes,96]:

Pretende-se gerar um PE com espectro de potência $P_x(e^{j\omega}) = \frac{5+4\cos(2\omega)}{10+6\cos(\omega)}$, por filtragem de ruído branco de média nula e variância unitária. Qual o filtro a usar?

$$\text{Como } P_x(e^{j\omega}) = \frac{5+2(e^{j2\omega} + e^{-j2\omega})}{10+3(e^{j\omega} + e^{-j\omega})} \text{ então } P_x(z) = \frac{5+2z^2+2z^{-2}}{10+3z+3z^{-1}} = \frac{z^{-2}(2z^4+5z^2+2)}{z^{-1}(3z^2+10z+3)}.$$

O numerador tem raízes -2 e -1/2 em z^2 . O denominador raízes -3 e -1/3 em z (existem sempre pares de raízes recíprocas). Logo podemos factorizar $P_x(z)$ da seguinte forma

$$P_x(z) = \frac{2z^{-2}(z^2+2)(z^2+\frac{1}{2})}{3z^{-1}(z+3)(z+\frac{1}{3})} = \frac{(1+2z^{-2})(2z^2+1)}{(1+3z^{-1})(3z+1)} = H(z)H(z^{-1})$$

Para termos um filtro $H(z)$ estável, devemos escolher o pólo em $z = -1/3$. Quantos aos zeros é indiferente, por exemplo

$$H(z) = \frac{2z^2+1}{3z+1} = \frac{2z^2(1+\frac{1}{2}z^{-2})}{3z(1+\frac{1}{3}z^{-1})} = \frac{2}{3}z \frac{1+\frac{1}{2}z^{-2}}{1+\frac{1}{3}z^{-1}}$$

Podemos adicionar ao filtro um atraso unitário sem que isso altere o espectro de potência final, ficando assim o filtro não só estável como também causal:

$$H(z) = \frac{2}{3} \frac{1+\frac{1}{2}z^{-2}}{1+\frac{1}{3}z^{-1}}.$$

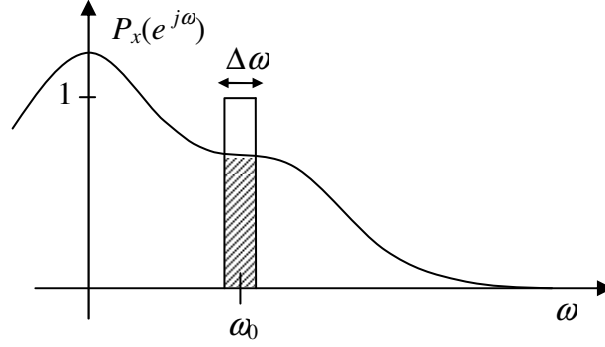
Positividade do espectro de potência

A partir da relação $E\{|y[n]|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(e^{j\omega})|^2 P_x(e^{j\omega}) d\omega \geq 0$, se definirmos um filtro passa

banda com uma banda de passagem de ganho unitário muito estreita, esta relação tem de se manter, qualquer que seja a frequência central do filtro, ω_0 , e a largura de banda, $\Delta\omega$. Como a potência da saída é

$$\frac{1}{2\pi} \int_{-\omega_0-\Delta\omega/2}^{\omega_0+\Delta\omega/2} P_x(e^{j\omega}) d\omega \geq 0$$

significa que qualquer espectro de potência $P_x(e^{j\omega})$ não pode ser negativo: $P_x(e^{j\omega}) \geq 0$.



Matriz de autocorrelação

Um PE pode ser visto como uma sequência de VAs. Se forem definidas $N+1$ VAs do processo, por exemplo,

$$\mathbf{x} = [x[0] \quad x[1] \quad \cdots \quad x[N]]^T$$

então a correspondente matriz de autocorrelação tem as propriedades de uma matriz de correlação com elementos $r_{ij} = E\{x[i]x^*[j]\}$. Mas se o processo for WSS, significa que a matriz de autocorrelação tem elementos $r_{ij} = r_x[i-j]$ e portanto apresenta a seguinte forma:

$$\mathbf{R}_x = E\{\mathbf{xx}^{*T}\} = \begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-N] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x[-1] \\ r_x[N] & \cdots & r_x[1] & r_x[0] \end{bmatrix} = \begin{bmatrix} r_x[0] & r_x^*[1] & \cdots & r_x^*[N] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x^*[1] \\ r_x[N] & \cdots & r_x[1] & r_x[0] \end{bmatrix}$$

isto é, todas as diagonais da matriz são iguais: é uma **matriz Toeplitz**. Se o processo for real então a matriz além de Toeplitz é simétrica.

Sabemos que os valores próprios da matriz são não negativos, mas além disso, verificam a seguinte propriedade (para processos WSS de média nula):

$$\min_{\omega} P_x(e^{j\omega}) \leq \lambda_i \leq \max_{\omega} P_x(e^{j\omega}).$$

A não negatividade da matriz de autocorrelação pode ser verificada através da potência total de saída de um filtro FIR. Seja N a ordem do filtro FIR onde os $N+1$ coeficientes do filtro se colocam no vector $\mathbf{h} = [h[0] \quad h[1] \quad \cdots \quad h[N]]^T$.

A potência de saída é

$$E\{|y[n]|^2\} = r_y[0]$$

Neste caso

$$r_y[m] = h[m] * r_x[m] * h^*[-m]$$

ou

$$\begin{aligned}
r_y[m] &= h[m] * \left(\sum_{k=0}^N r_x[m+k] \cdot h^*[k] \right) \\
&= \sum_{l=0}^N \sum_{k=0}^N h[l] \cdot r_x[m+k-l] \cdot h^*[k]
\end{aligned}$$

Logo

$$\begin{aligned}
r_y[0] &= \sum_{l=0}^N \sum_{k=0}^N h[l] \cdot r_x[k-l] \cdot h^*[k] \\
&= \mathbf{h}^{*T} \mathbf{R}_x \mathbf{h}
\end{aligned}$$

Um escalar definido da forma $\mathbf{h}^{*T} \mathbf{R}_x \mathbf{h}$ chama-se factor quadrático. Neste caso o factor quadrático é a potência de saída, logo um valor positivo ou zero, qualquer que sejam os coeficientes do filtro, ou seja \mathbf{h} . Isto só pode acontecer se a matriz for semi definida positiva, o que se anota por $\mathbf{R}_x \geq 0$. Os valores próprios de \mathbf{R}_x são positivos ou zero.

2.4.1 Processos regulares

O espectro de potência, $P_x(e^{j\omega})$, de um processo estacionário é uma função **real, positiva e periódica** em ω . Se for uma função contínua (sem impulsos), pode mostrar-se que $P_x(z)$ pode ser factorizado no seguinte produto com pares de raízes recíprocas,

$$P_x(z) = \sigma_0^2 H(z) H^*(1/z^*)$$

e diz-se que o processo é regular. Significa que um processo regular pode ser definido através da saída de um filtro causal e estável, $H(z)$, quando a entrada é ruído branco com $P_v(e^{j\omega}) = \sigma_0^2$. O seu espectro de potência é então

$$P_x(e^{j\omega}) = \sigma_0^2 |H(e^{j\omega})|^2.$$

Se $H(z)$ for uma expressão racional, podem definir-se 3 tipos de processos regulares consoante $H(z)$ tem zeros e pólos (finitos), só zeros ou só pólos.

Processo autoregressivos (AR):

$$H(z) = \frac{b_0}{\sum_{k=0}^p a_k z^{-k}} \quad P_x(e^{j\omega}) = \sigma_0^2 \frac{|b_0|^2}{|A(e^{j\omega})|^2}.$$

Processo “moving average”, (MA):

$$H(z) = \sum_{k=0}^q b_k z^{-k} \quad P_x(e^{j\omega}) = \sigma_0^2 |B(e^{j\omega})|^2$$

Processo ARMA:

$$H(z) = \frac{\sum_{k=0}^q b_k z^{-k}}{\sum_{k=0}^p a_k z^{-k}} \quad P_x(e^{j\omega}) = \sigma_0^2 \left| \frac{B(e^{j\omega})}{A(e^{j\omega})} \right|^2.$$

2.4.2 Equações de Yule-Walker

Através da equação de diferenças do filtro $H(z)$ podem derivar-se relações recursivas para a sequência de autocorrelação do processo.

Para o caso mais genérico, o processo ARMA, como

$$\frac{\sum_{k=0}^q b_k z^{-k}}{\sum_{k=0}^p a_k z^{-k}} = \frac{X(z)}{V(z)} \Rightarrow X(z) \sum_{k=0}^p a_k z^{-k} = V(z) \sum_{k=0}^q b_k z^{-k},$$

teremos a seguinte equação de diferenças:

$$\sum_{k=0}^p a_k x[n-k] = \sum_{k=0}^q b_k v[n-k], \quad a_0=1,$$

onde $v[n]$ é o processo ruído branco de entrada do filtro. Multiplicando por $x^*[n-l]$ e tomando a média, resulta

$$\sum_{k=0}^p a_k r_x[l-k] = \sum_{k=0}^q b_k r_{vx}[l-k]$$

Mas como $r_{vx}[m] = h^*[-m] * r_v[m] = \sigma_0^2 h^*[-m]$ (a entrada $v[n]$ é ruído branco com $r_v[m] = \sigma_0^2 \delta[m]$) então

$$\sum_{k=0}^p a_k r_x[l-k] = \sigma_0^2 \sum_{k=0}^q b_k h^*[k-l].$$

Apenas no caso de processos **AR** é que é possível obter um sistema de equações lineares para obter os coeficientes do filtro. Neste caso $b_k=0$ para $k \neq 0$, e $h[0]=b_0$, pelo que a equação para $l \geq 0$ é:

$$\sum_{k=0}^p a_k r_x[l-k] = \sigma_0^2 |b_0|^2 \delta[l], \quad l \geq 0; \quad a_0=1.$$

Matricialmente,

$$\left[\begin{array}{c|ccc} r_x[0] & r_x[-1] & \cdots & r_x[-p] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x[-1] \\ r_x[p] & \cdots & r_x[1] & r_x[0] \end{array} \right] \left[\begin{array}{c} 1 \\ a_1 \\ \vdots \\ a_p \end{array} \right] = \sigma_0^2 |b_0|^2 \left[\begin{array}{c} 1 \\ 0 \\ \vdots \\ 0 \end{array} \right]$$

Para $l \geq 1$ teremos:

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-p+1] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x[-1] \\ r_x[p-1] & \cdots & r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix}$$

Se soubermos que um processo é AR significa que podemos conhecer o filtro associado (os coeficientes a_k) resolvendo estas equações de Yule-Walker. Existem algoritmos eficientes para resolver este sistema de equações, por exemplo o algoritmo de Levinson-Durbin.

Exemplo

Um PE é AR(2), isto é, $p=2$. Sabendo que $r_x[0]=1$, $r_x[1]=1/2$ e $r_x[2]=0$, qual é o espectro de potência do processo? Qual o valor de $r_x[3]$?

Basta conhecer o filtro $H(z) = \frac{b_0}{1 + a_1 z^{-1} + a_2 z^{-2}}$. Através das equações de Yule-Walker obtemos

$$\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = - \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

Da equação de Yule Walquer com $l=0$, sabemos que $\sum_{k=0}^p a_k r_x[-k] = \sigma_0^2 |b_0|^2$, ou

$$r_x[0] + a_1 r_x[-1] + a_2 r_x[-2] = \sigma_0^2 |b_0|^2 \Rightarrow 1 - \frac{2}{3} \cdot \frac{1}{2} + 0 = \sigma_0^2 |b_0|^2 = \frac{2}{3}.$$

Podemos admitir que $\sigma_0^2 = 1$ vindo $|b_0|^2 = 2/3$. Logo o espectro de potência do processo é

$$P_x(e^{j\omega}) = \frac{2/3}{\left|1 - \frac{2}{3}e^{-j\omega} + \frac{1}{3}e^{-j2\omega}\right|^2}.$$

Podemos agora saber toda a sequência de autocorrelação do processo, pois

$$\sum_{k=0}^2 a_k r_x[l-k] = |b_0|^2 \delta[l] \text{ para } l \geq 0.$$

Para $l=3$, $r_x[3] + a_1 r_x[2] + a_2 r_x[1] = 0$, ou

$$r_x[3] - 0 + \frac{1}{3} \cdot \frac{1}{2} = 0 \Rightarrow r_x[3] = -1/6.$$

Nota: A potência do processo é $\sigma_x^2 = r[0] = 1$. Logo

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{2/3}{\left|1 - \frac{2}{3}e^{-j\omega} + \frac{1}{3}e^{-j2\omega}\right|^2} d\omega = 1$$

Através dos coeficientes do filtro podemos também obter a variância do processo (de media nula). Basta tomar as equações de Yule-Walker tendo como incógnitas os valores de autocorrelação. Neste caso teríamos

$$\begin{bmatrix} r[0] & r[1] & r[2] \\ r[1] & r[0] & r[1] \\ r[2] & r[1] & r[0] \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} b_0^2 \\ 0 \\ 0 \end{bmatrix} \quad \text{ou} \quad \begin{cases} r[0] + a_1 r[1] + a_2 r[2] = b_0^2 \\ r[1] + a_1 r[0] + a_2 r[1] = 0 \\ r[2] + a_1 r[1] + a_2 r[0] = 0 \end{cases}$$

logo

$$\begin{bmatrix} 1 & a_1 & a_2 \\ a_1 & 1+a_2 & 0 \\ a_2 & a_1 & 1 \end{bmatrix} \begin{bmatrix} r[0] \\ r[1] \\ r[2] \end{bmatrix} = \begin{bmatrix} b_0^2 \\ 0 \\ 0 \end{bmatrix}.$$

O determinante da matriz do sistema é $\Delta = (1+a_2) - a_1^2 + a_2(a_1^2 - a_2(1+a_2))$ e

$$r[0] = \frac{1}{\Delta} \begin{vmatrix} b_0^2 & a_1 & a_2 \\ 0 & 1+a_2 & 0 \\ 0 & a_1 & 1 \end{vmatrix} = \frac{b_0^2(1+a_2)}{\Delta} = \frac{b_0^2(1+a_2)}{1+a_2 - a_1^2 + a_2(a_1^2 - a_2(1+a_2))}.$$

Assim a variância de um processo AR de ordem 2 com $H(z) = b_0/(1+a_1z^{-1} + a_2z^{-2})$ é

$$r[0] = \sigma_x^2 = \frac{b_0^2(1+a_2)}{1+a_2 - a_1^2 + a_2(a_1^2 - a_2(1+a_2))} = \frac{b_0^2(a_2+1)}{(a_2-1)[a_1^2 - (a_2+1)^2]}.$$

Neste caso, com $b_0^2 = \frac{2}{3}$, $a_1 = -2/3$ e $a_2 = 1/3$, obtemos $r[0]=1$, $r[1]=1/2$ e $r[2]=0$. Fazendo $a_2=0$, obtemos a variância de um processo AR de ordem 1:

$$\sigma_x^2 = \frac{b_0^2}{1-a_1^2}.$$

2.4.3 Teorema da decomposição de Wold

Um importante teorema (**teorema da decomposição de Wold**) diz que qualquer processo pode ser definido através da soma de dois processos, um regular e outro “previsível” (com espectro de potência composto por impulsos),

$$x[n] = x_r[n] + x_p[n],$$

onde $x_r[n]$ é regular e $x_p[n]$ é previsível e ortogonal a $x_r[n]$: $E\{x_r[n]x_p^*[n]\} = 0$. A forma geral do espectro de potência de processos WSS é assim dada através de

$$P_x(e^{j\omega}) = P_r(e^{j\omega}) + \sum_k c_k \delta(\omega - \omega_k).$$

3. Estimação Espectral de Potência

3.1 Noção de estimador

Considere-se o problema de estimar o valor de um parâmetro, θ , através de uma sequência de VAs, x_n , $n = 1..N$. Seja a estimativa $\hat{\theta}_N$, que é também uma VA. Se o valor esperado da estimativa for igual ao verdadeiro valor, θ , diz-se que a estimativa é **não-polarizada**. A polarização (bias) da estimativa é

$$B = \theta - E\{\hat{\theta}_N\}.$$

Se a estimativa for **não polarizada**, então $B=0$ e $E\{\hat{\theta}_N\} = \theta$.

Se a estimativa for polarizada mas a polarização tender para zero à medida que o número de observações N tende para infinito, então diz-se que a estimativa é **assimptoticamente não polarizada**. Se a estimativa converge para o verdadeiro valor segundo um dado critério, por exemplo em termos de média quadrática,

$$\lim_{N \rightarrow \infty} E\{|\hat{\theta}_N - \theta|^2\} = 0$$

diz-se que a estimativa é **consistente**. Assim, se $\hat{\theta}_N$ é não polarizada, a variância da estimativa deve tender para zero à medida que o número de observações aumenta.

Podemos “medir” esta tendência através da desigualdade de Tchebycheff (ver exemplo seguinte):

$$\Pr\{|\hat{\theta}_N - \theta| \geq \varepsilon\} \leq \frac{\text{Var}\{\hat{\theta}_N\}}{\varepsilon^2}.$$

Um exemplo clássico de estimador é a média amostrada,

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=1}^N x_n,$$

onde se assume que as N VAs são não correlacionadas e igualmente distribuídas com a mesma média μ_x . Trata-se de uma estimativa **não polarizada** pois

$$E\{\hat{\mu}_N\} = \frac{1}{N} \sum_{n=1}^N E\{x_n\} = \mu_x.$$

Sendo as VAs não correlacionadas, a variância da soma das VAs é igual à soma das variâncias:

$$\text{Var}\{N\hat{\mu}_N\} = \text{Var}\left\{\sum_{n=1}^N x_n\right\} = \sum_{n=1}^N \sigma_x^2 = N\sigma_x^2 = N^2 \text{Var}\{\hat{\mu}_N\}. \text{ Logo}$$

$$\text{Var}\{\hat{\mu}_N\} = E\{(\hat{\mu}_N - \mu_x)^2\} = \frac{\sigma_x^2}{N}$$

Trata-se portanto de uma estimativa consistente da média.

Exemplo

O número de observações N para que a estimativa da média esteja dentro de um intervalo de 10% do valor correcto com 99% de certeza é dado por

$$\Pr\{|\hat{\mu}_N - \mu_x| \leq 10^{-1} \mu_x\} \geq 1 - \frac{\sigma_x^2}{N 10^{-2} \mu_x^2} = 0.99$$

Logo $N = 10^4 \frac{\sigma_x^2}{\mu_x^2}$. Por exemplo, para que a média amostrada de uma sequência de VAs com $\mu_x = 2$ e $\sigma_x^2 = 1$ esteja no intervalo entre 1.9 e 2.1 com 99% de certeza, devem tomar-se 2500 amostras (VAs).

3.2 Estimativa da autocorrelação

Usualmente dispomos apenas de N amostras de um processo WSS ergódico, por exemplo para $n=0..N-1$, pelo que a estimativa da autocorrelação pode ser obtida através da soma finita

$$\hat{r}_x[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n+k] \cdot x^*[n], \quad k=0..N-1$$

sendo os valores para índices negativos obtidos pela simetria complexa conjugada: $\hat{r}_x[-k] = \hat{r}_x^*[k]$.

Podemos assim obter valores da estimativa apenas para $|k| < N$. Além disso, o somatório anterior tem apenas um termo para $k=N-1$, que usa os valores extremos de $x[n]$:

$\hat{r}_x[N-1] = \frac{1}{N} x[N-1] \cdot x^*[0]$. Devido a este facto, esta estimativa é **polarizada**, embora seja **assimptoticamente não polarizada**. De facto,

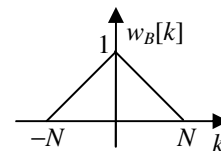
$$\begin{aligned} E\{\hat{r}_x[k]\} &= \frac{1}{N} \sum_{n=0}^{N-1-k} E\{x[n+k] \cdot x^*[n]\} = \frac{1}{N} \sum_{n=0}^{N-1-k} r_x[k] = \frac{1}{N} r_x[k] \sum_{n=0}^{N-1-k} 1 \\ &= \frac{N-k}{N} r_x[k] \end{aligned}$$

Considerando os índices para $|k| < N$, podemos dizer que

$$E\{\hat{r}_x[k]\} = w_B[k] \cdot r_x[k],$$

onde $w_B[k]$ é a janela de Bartlett (janela triangular) centrada na origem e definida como

$$w_B[k] = \begin{cases} \frac{N-|k|}{N}, & |k| < N \\ 0, & |k| \geq N \end{cases}.$$



Podemos também definir a estimativa da autocorrelação através de uma convolução. Basta definir a amostra conhecida do processo como o sinal

$$x_N[n] = w_R[n] \cdot x[n]$$

onde $w_R[n]$ é a janela rectangular de N amostras. Podemos assim definir a estimativa da autocorrelação com limites infinitos:

$$\begin{aligned}\hat{r}_x[k] &= \frac{1}{N} \sum_{n=-\infty}^{+\infty} x_N[n+k] \cdot x_N^*[n] \\ &= \frac{1}{N} x_N[k] * x_N^*[-k]\end{aligned}$$

Se usarmos esta expressão para obter o valor esperado da estimativa da autocorrelação, resulta o seguinte:

$$\begin{aligned}E\{\hat{r}_x[k]\} &= \frac{1}{N} \sum_{n=-\infty}^{+\infty} w_R[n+k] w_R[n] E\{x[n+k] \cdot x_N^*[n]\} \\ &= \frac{1}{N} (w_R[k] * w_R[-k]) \cdot r_x[k]\end{aligned}$$

Significa que $w_B[k] = \frac{1}{N} (w_R[k] * w_R[-k])$ (resultado conhecido).

A estimativa não polarizada da autocorrelação é

$$\hat{r}_{xNP}[k] = \frac{1}{N-k} \sum_{n=0}^{N-1-k} x[n+k] \cdot x^*[n] = \frac{N}{N-|k|} \hat{r}_x[k], \quad |k| < N.$$

3.2 Periodograma

O espectro de potência é a transformada de Fourier da sequência de autocorrelação. Portanto a estimação do espectro de potência é equivalente à estimação da autocorrelação.

A transformada de Fourier (DTFT) da estimativa da autocorrelação é conhecida com o nome de **periodograma**:

$$\hat{P}_{per}(e^{j\omega}) = \sum_{k=-(N-1)}^{N-1} \hat{r}_x[k] \cdot e^{-jk\omega}$$

Em termos da amostra do processo, $x_N[n]$, como $\hat{r}_x[k] = \frac{1}{N} x_N[k] * x_N^*[-k]$, temos que

$$\hat{P}_{per}(e^{j\omega}) = \frac{1}{N} X_N(e^{j\omega}) \cdot X_N^*(e^{j\omega}) = \frac{1}{N} |X_N(e^{j\omega})|^2$$

Portanto, o periodograma é proporcional ao módulo quadrado da DTFT de $x_N[n]$.

As propriedades do periodograma como estimador espectral de potência derivam das propriedades da estimativa da autocorrelação: o periodograma é um estimador polarizado mas assintoticamente não polarizado:

$$\begin{aligned} E\{\hat{P}_{per}(e^{j\omega})\} &= \text{DTFT}\{E\{\hat{r}_x[k]\}\} = \text{DTFT}\{w_B[k] \cdot r_x[k]\} \\ &= \frac{1}{2\pi} W_B(e^{j\omega}) \circledast P_x(e^{j\omega}) \end{aligned}$$

onde $W_B(e^{j\omega})$ é a transformada de Fourier da janela triangular,

$$W_B(e^{j\omega}) = \frac{1}{N} \frac{\sin^2(N\omega/2)}{\sin^2(\omega/2)} = \frac{1}{N} |W_R(e^{j\omega})|^2.$$

uma vez que, como vimos, $w_B[k] = \frac{1}{N} (w_R[k] * w_R[-k])$. Quando N tende para infinito o espectro da janela tende para um impulso, logo o periodograma tende para o espectro de potência do processo.

O valor esperado do periodograma é a convolução periódica do espectro de potência do processo com a DTFT da janela triangular (logo polarizado).

Exemplo

Considere-se o processo harmónico embebido em ruído branco

$$x[n] = A \sin(\omega_0 n + \varphi) + v[n]$$

onde φ é uma VA uniformemente distribuída entre $-\pi$ e π e $v[n]$ é ruído branco de variância σ_v^2 . A autocorrelação deste processo é

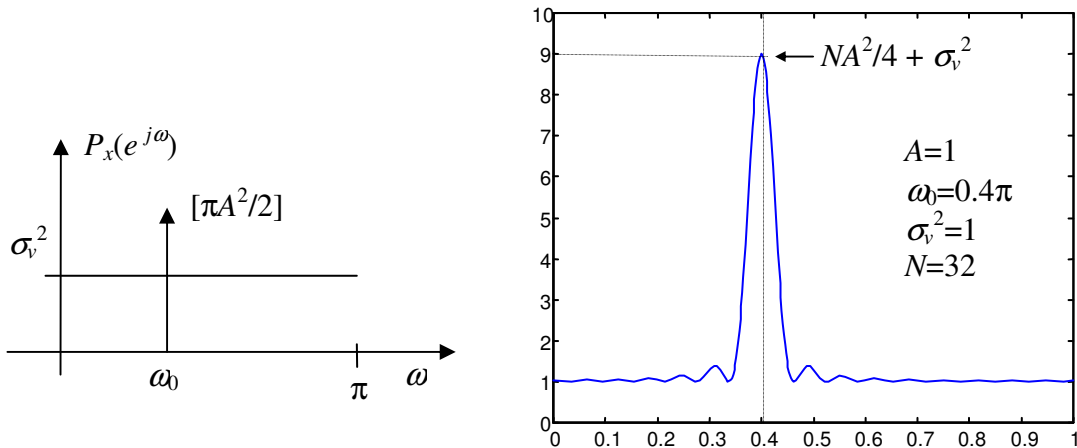
$$r_x[k] = \frac{A^2}{2} \cos(\omega_0 k) + \sigma_v^2 \delta[k]$$

e portanto, o seu espectro de potência é

$$P_x(e^{j\omega}) = \pi \frac{A^2}{2} [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] + \sigma_v^2, \quad |\omega| < \pi$$

A esperança do periodograma é

$$\begin{aligned} E\{\hat{P}_{per}(e^{j\omega})\} &= \frac{1}{2\pi} W_B(e^{j\omega}) \circledast P_x(e^{j\omega}) \\ &= \frac{A^2}{4} [W_B(e^{j(\omega - \omega_0)}) + W_B(e^{j(\omega + \omega_0)})] + \sigma_v^2 \end{aligned}$$



Variância do periodograma

No caso de ruído branco Gaussiano é possível conhecer a variância do periodograma. O resultado é o seguinte:

$$\text{Var}\{\hat{P}_{per}(e^{j\omega})\} = \sigma_x^4 = P_x^2(e^{j\omega}).$$

Vemos assim que a variância do periodograma não tende para zero mesmo quando N tende para infinito. Significa que o periodograma **não** é uma estimativa consistente do espectro de potência. Para outros processos o resultado aproximado é

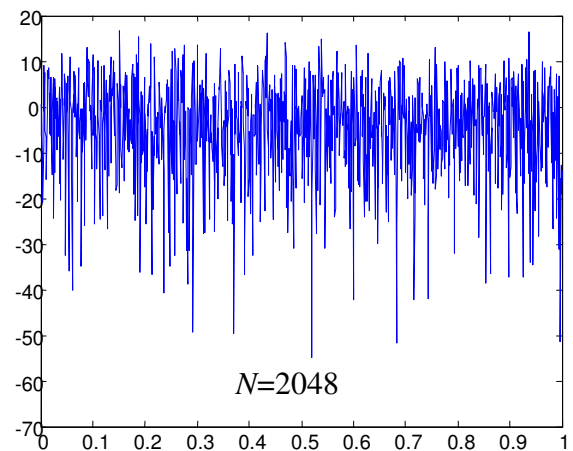
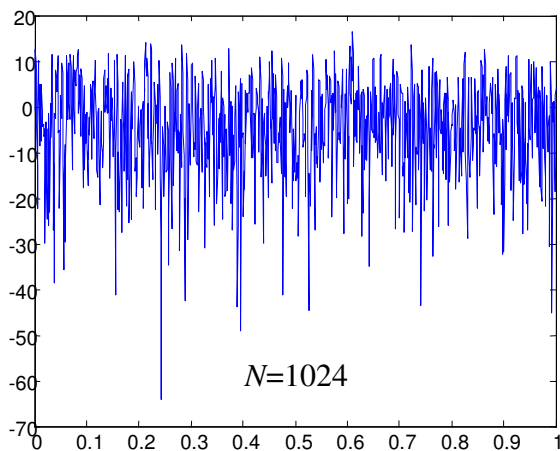
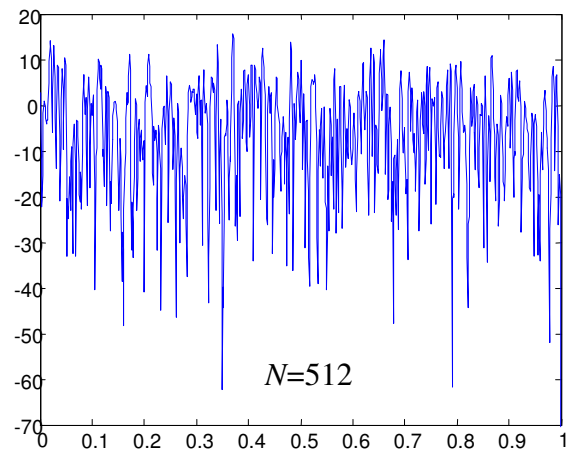
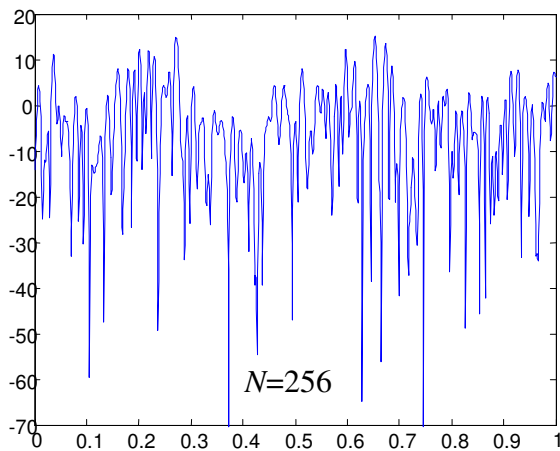
$$\text{Var}\{\hat{P}_{per}(e^{j\omega})\} \approx P_x^2(e^{j\omega}),$$

mantendo-se a conclusão anterior.

Exemplo

Periodograma de ruído branco Gaussiano ($\sigma_v^2=1$) com 256, 512, 1024 e 2048 amostras, avaliado através de DFTs de 2048 pontos.

```
X=fft(randn(1,256),2048); P1=(1/256)*abs(X(1:1025)).^2;
X=fft(randn(1,512),2048); P2=(1/512)*abs(X(1:1025)).^2;
X=fft(randn(1,1024),2048); P3=(1/1024)*abs(X(1:1025)).^2;
X=fft(randn(1,2048),2048); P4=(1/2048)*abs(X(1:1025)).^2;
wn=(0:1024)/1024;
plot(wn,db(P1),wn,db(P2),wn,db(P3),wn,db(P4))
```

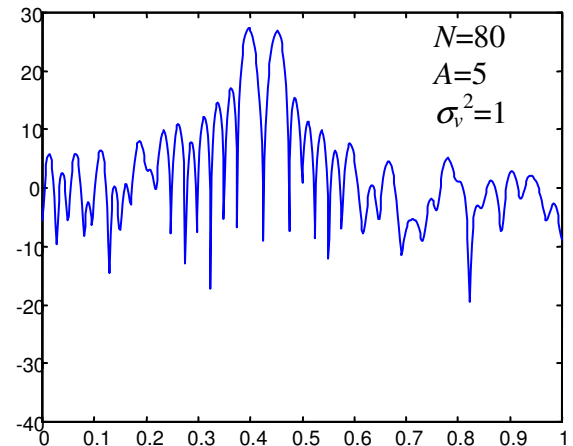
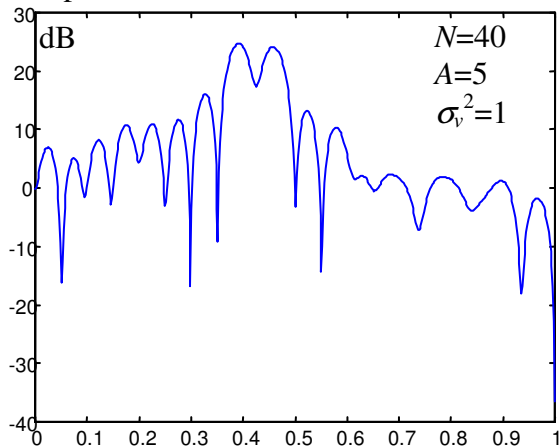


A **resolução do periodograma** depende da largura do lóbulo principal da janela de Bartlett, que neste caso (largura $2N$) é $4\pi/N$, a mesma da janela rectangular de N amostras.

Considere-se o caso de um processo harmónico que consiste em duas sinusóides de igual amplitude e de fase aleatória, embebidas em ruído branco de variância unitária,

$$x[n] = A \sin(\omega_1 n + \varphi_1) + A \sin(\omega_2 n + \varphi_2) + v[n]$$

com $\omega_1 = 0.4\pi$, $\omega_2 = 0.45\pi$ e $A=5$. Apenas com $N=80$ amostras do processo é possível resolver completamente os dois harmónicos.



Periodograma modificado

Se aplicarmos às amostras do processo uma janela diferente da janela rectangular obtemos um periodograma modificado. Embora a janela rectangular tenha o lóbulo principal mais estreito que todas as outras, os lóbulos secundários podem mascarar as componentes fracas e de banda estreita do processo.

O periodograma modificado é dado por

$$\hat{P}_M(e^{j\omega}) = \frac{1}{NU} \left| \text{DTFT} \{x[n] \cdot w[n]\} \right|^2$$

onde $w[n]$ é uma janela genérica, N a sua largura da janela e U a sua energia normalizada:

$$U = \frac{1}{N} \sum_{n=0}^{N-1} |w[n]|^2 = \frac{1}{2\pi N} \int_{-\pi}^{\pi} |W(e^{j\omega})|^2 d\omega.$$

O valor esperado do periodograma modificado é

$$E\{\hat{P}_M(e^{j\omega})\} = \frac{1}{2\pi NU} P_x(e^{j\omega}) \otimes |W(e^{j\omega})|^2.$$

A constante U permite normalizar a estimativa e torná-la assintoticamente não polarizada. De facto, se $x[n]$ for ruído branco de variância unitária ($P_x(e^{j\omega})=1$) então o valor esperado da estimativa é $|W(e^{j\omega})|^2 / NU$ e tende para um impulso quando $N \rightarrow \infty$.

A variância do periodograma modificado é a mesma do periodograma. A resolução depende da janela usada.

3.3 Métodos de suavização

A fim de obter uma estimativa consistente do espectro de potência, vários métodos foram propostos.

A. Método de Bartlett (médias de periodogramas)

O método consiste em dividir as N amostras conhecidas do processo em K sequências de comprimento L ($N = KL$):

$$x_k[n] = x[n + kL] \quad , \quad k = 0..K-1.$$

Fazer a média dos K periodogramas de comprimento L .

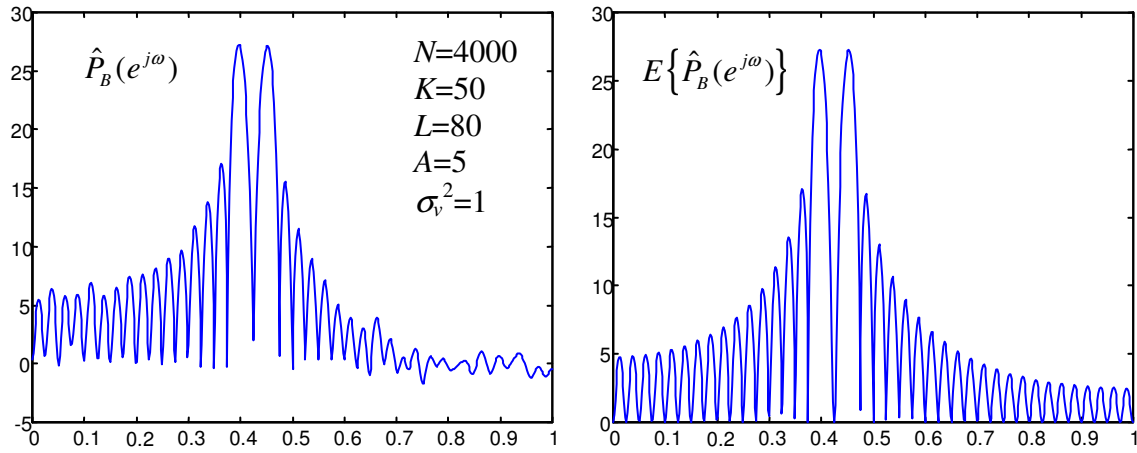
$$\hat{P}_B(e^{j\omega}) = \frac{1}{KL} \sum_{k=0}^{K-1} \left| \text{DTFT} \{x_k[n]\} \right|^2$$

O valor esperado não se altera em relação ao do periodograma, contudo a variância desta estimativa é agora

$$\text{Var}\{\hat{P}_B(e^{j\omega})\} = \frac{1}{K} \text{Var}\{\hat{P}_{per}(e^{j\omega})\} \approx \frac{1}{K} P_x^2(e^{j\omega})$$

Assim, à medida que N tende para infinito, o mesmo acontece com K e com L , tornando esta estimativa consistente.

Exemplo anterior com $N=4000$, $K=50$, $L=80$:



B. Método de Welch

Idêntico ao anterior, mas com sobreposição das sub-sequências e usando o periodograma modificado.

$$\hat{P}_W(e^{j\omega}) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{P}_M^{(k)}(e^{j\omega})$$

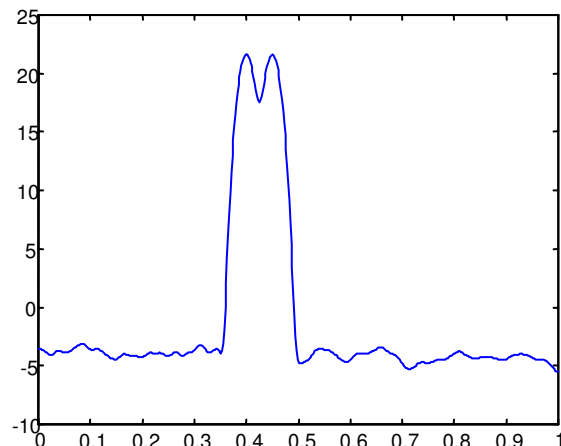
onde $\hat{P}_M^{(k)}(e^{j\omega})$ é o periodograma modificado de cada secção (sub-sequência) k .

Exemplo com 4000 amostras, $L=80$, janela de Hamming e sobreposição de 10 amostras ($K=57$).

Nota:

O comando PSD no Matlab usa este método com a janela de Hanning. Ex:

`Px = psd(x, 1024, 1, 80, 10);`



C. Método de Blackman-Tuckey

Este método tem em conta que os valores implícitos da autocorrelação são mal estimados para índices perto de N . Por exemplo o valor $\hat{r}_x[N-1] = \frac{1}{N} x[N-1] \cdot x^*[0]$ é estimado tomando apenas duas amostras.

Assim, uma forma de reduzir a variância do periodograma consiste em reduzir a contribuição destas estimativas no periodograma, o que pode ser feito aplicando uma janela à sequência da autocorrelação:

$$\hat{P}_{BT}(e^{j\omega}) = \text{DTFT}\{\hat{r}_x[n] \cdot w[n]\}.$$

Usualmente a janela $w[n]$ é nula para $|n| > M$ onde $M < N$. isto é, desprezam-se as últimas $M-N$ amostras da autocorrelação. A janela tem de ser simétrica de forma que o produto com $\hat{r}_x[n]$ continue uma sequência par/complexa simétrica para garantir a positividade da estimativa do espectro de potência.

Como $E\{\hat{r}_x[k]\} = w_B[k] \cdot r_x[k]$, temos que o valor esperado desta estimativa é

$$\begin{aligned} E\{\hat{P}_{BT}(e^{j\omega})\} &= \text{DTFT}\{E\{\hat{r}_x[k]\} \cdot w[n]\} \\ &= \text{DTFT}\{r_x[k] \cdot w_B[k] \cdot w[n]\} \\ &= \frac{1}{2\pi} P_x(e^{j\omega}) \otimes W_B(e^{j\omega}) \otimes W(e^{j\omega}) \end{aligned}$$

Fazendo um certo número de simplificações, pode mostrar-se que a variância desta estimativa é aproximadamente

$$\text{Var}\{\hat{P}_{BT}(e^{j\omega})\} \approx P_x^2(e^{j\omega}) \frac{1}{N} \sum_{k=-M}^M w^2[k]$$

Se a janela for rectangular, então $\text{Var}\{\hat{P}_{BT}(e^{j\omega})\} \approx P_x^2(e^{j\omega}) \frac{2M+1}{N}$.

Existe assim um compromisso entre polarização e variância: para reduzir a variância da estimativa M deve ser tão pequeno quanto possível, embora neste caso a suavização espectral e resolução em frequência, isto é a polarização, venham aumentadas.

3.4 Método da Máxima Entropia

Uma das limitações dos métodos clássicos é o facto de se assumir implicitamente que a sequência de autocorrelação é nula para $|k| \geq N$.

No método da máxima entropia a sequência de autocorrelação é **extrapolada** para os índices $k \geq N$ de forma a maximizar a taxa de entropia do processo.

Entropia é uma medida da incerteza de um processo estocástico. Define-se a taxa de entropia de um processo como a média, por amostra, dessa incerteza.

A taxa entropia de um processo com espectro de potência $P_x(e^{j\omega})$ (de média nula) é dada por [Papoulis, 91],

$$\Delta H = \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) d\omega + C^{te}.$$

Suponhamos que de um processo estocástico (de média nula) é conhecida a sequência de autocorrelação apenas para $|k| \leq p$. Pretende-se obter uma estimativa do espectro de potência do processo extrapolando os valores da autocorrelação para $|k| > p$ e de forma que esta extrapolação corresponda a admitir o menos possível sobre as amostras do processo necessárias para obter essa extrapolação. Isto é, pretende-se maximizar a taxa de entropia para $|k| > p$. Teremos então

$$P_x(e^{j\omega}) = \sum_{k=-p}^p r_x[k] \cdot e^{-j\omega k} + \sum_{|k|>p} r_e[k] \cdot e^{-j\omega k}$$

onde $r_e[k]$ corresponde à sequência de autocorrelação extrapolada. Para maximizar a taxa de entropia para $|k| > p$, deverá ser

$$\begin{aligned} \frac{\partial \Delta H}{\partial r_e[k]} &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial r_e[k]} \ln P_x(e^{j\omega}) d\omega \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(e^{j\omega})} \frac{\partial P_x(e^{j\omega})}{\partial r_e[k]} d\omega \\ &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{e^{-j\omega k}}{P_x(e^{j\omega})} d\omega \\ &= 0, \quad |k| > p \end{aligned}$$

Poderíamos igualmente ter derivado em ordem a $r_e[-k]$, vindo então que

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(e^{j\omega})} e^{j\omega k} d\omega = 0, \quad |k| > p.$$

Tomando a transformada inversa de Fourier desta expressão, verificamos que se trata de uma sequência limitada ao intervalo $|k| \leq p$. Seja ela $q[k]$. Teremos que

$$\frac{1}{P_x(e^{j\omega})} = \sum_{k=-p}^p q[k] \cdot e^{-j\omega k},$$

e portanto

$$\hat{P}_{MEM}(e^{j\omega}) = \frac{1}{\sum_{k=-p}^p q[k] \cdot e^{-j\omega k}}$$

Factorizado esta expressão, chegamos à conclusão que a estimativa de máxima entropia é, afinal, o espectro de potência de um processo autoregressivo (AR):

$$\boxed{\hat{P}_{MEM}(e^{j\omega}) = \frac{|b_0|^2}{\left| \sum_{k=0}^p a_k e^{-j\omega k} \right|^2}}, \quad a_0=1.$$

Como vimos anteriormente, os coeficientes a_k podem ser obtidos através das equações de Yule-Walker:

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-p+1] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x[-1] \\ r_x[p-1] & \cdots & r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p] \end{bmatrix}$$

que correspondem à equação

$$\sum_{k=0}^p a_k r_x[l-k] = 0, \quad l > 0$$

Esta é também a equação que define a extrapolação da sequência de autocorrelação. Como $a_0=1$, temos a recursão

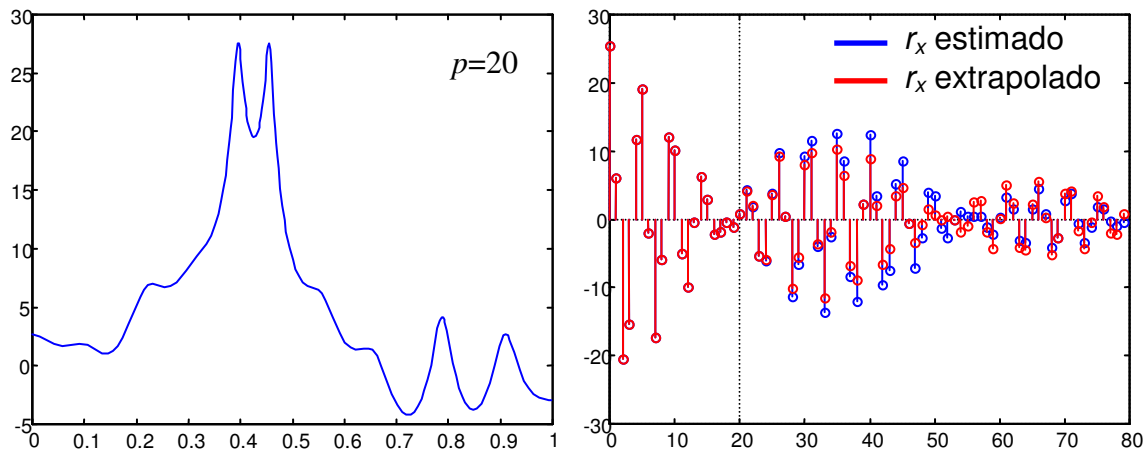
$$r_x[l] = - \sum_{k=1}^p a_k r_x[l-k], \quad l > 0$$

O coeficiente b_0 deve ser escolhido de forma a potência total do processo coincida com a desta estimativa, isto é,

$$\sum_{k=0}^p a_k r_x[-k] = |b_0|^2.$$

Exemplo.

Considere-se o processo harmónico anterior com $N=80$ pontos de onde se estimam 21 valores de autocorrelação. A estimativa de máxima entropia com $p=20$ é a que se apresenta na figura seguinte. Mostram-se também a sequência de autocorrelação estimada e a sequência extrapolada. Notar que para $l \leq p$ as duas sequências coincidem.



Código em Matlab que gerou estas figuras:

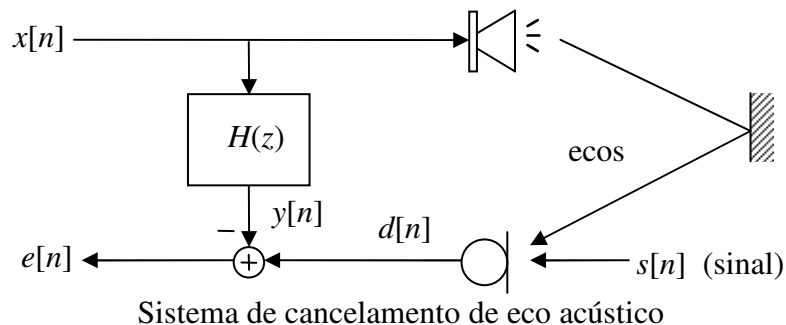
```
N=80;
A=5;
w1=0.4*pi;
w2=0.45*pi;
rand('state',0);
randn('state',0);
fil=rand(1);
fi2=rand(1);
v=randn(1,N);
n=0:N-1;
%amostra do processo:
x=A*sin(w1*n+fil)+A*sin(w2*n+fi2)+v;
%autocorrelação
rx=xcorr(x,'biased'); %sequência par; divisão por N.
%Estimativa MEM com p=20
p=20
r=rx(N:N+p); % p+1 valores da estimativa da autocorrelação, r[0]..r[p]
%Equações de Yule-Walker:
R=toeplitz(r(1:p)); %matriz do sistema com valores r[0]...r[p-1]
a=-R\r(2:p+1); %solução de R*a = -b, onde b=[r[1],...,r[p]]'
a=[1;a]; %colocar a0=1
b02=r'*a; %quadrado do ganho: b0^2 = sum(ak*r[k]), k=0..p.
%Módulo quadrado em [0,pi[ com 512 pontos:
P=abs(freqz(sqrt(b02),a)).^2;
wn=(0:511)/512; %freq. normalizada em [0,pi[
plot(wn,10*log10(P))
%sequência de autocorrelação extrapolada com 80 pontos
re=zeros(N,1);
re(1:p+1)=r; %valores tomados (não extrapolados)
for m=p+1:N-1
    re(m+1)= - a(2:p+1)'*re((m-1:-1:m-p)+1);
end
stem((0:N-1),rx(N:end)), hold on, stem((0:N-1),re,'r'), hold off
```

4. Filtros Óptimos

Filtros óptimos são filtros obtidos em condições estocásticas através da minimização da média quadrática do erro de modelação.

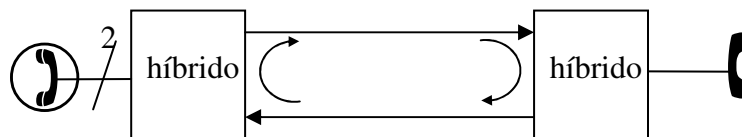
Exemplos de Aplicações

a) Cancelamento de eco acústico ou eléctrico



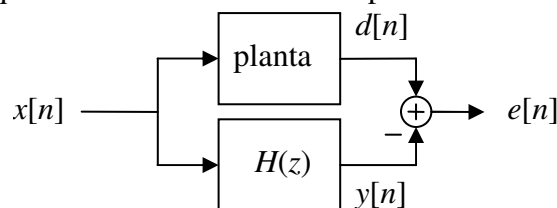
Neste exemplo de aplicação pretende-se eliminar ou reduzir o eco acústico em sistemas de “mãos livres” através de um filtro $H(z)$ que identifique e simule o sistema acústico que produz o eco. Usualmente o filtro é “adaptativo”, isto é actualizado com grande frequência.

Mesmo nos sistemas telefónicos convencionais pode acontecer o chamado eco “eléctrico” (em comunicações via satélite ou a grande distância que provocam um atraso considerável do sinal) devido à imperfeição do sistema que passa de 4 ↔ 2 fios (híbrido).



Eco eléctrico (em comunicações de longa distância (satélite))

Um problema semelhante é o problema da identificação de sistemas: pretende-se modelar um sistema desconhecido (planta) através do sistema linear $H(z)$. Os coeficientes do filtro $H(z)$ são escolhidos de forma a minimizar o média quadrática do erro de modelação, entre o sinal desejado, $d[n]$, e a saída do sistema quando a entrada comum é o processo estocástico $x[n]$.

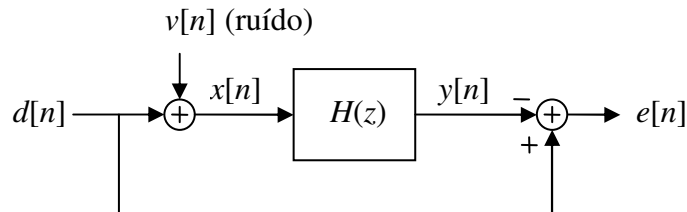


Identificação de sistemas

b) Redução de ruído

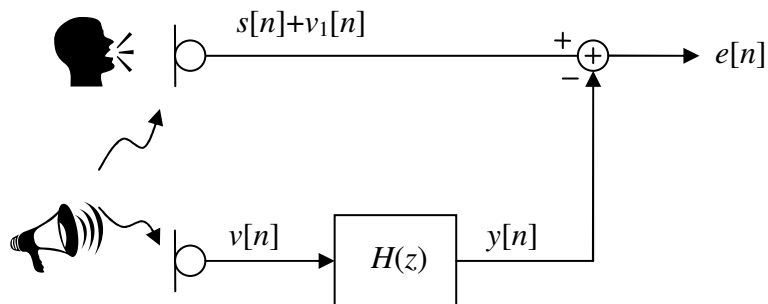
Em muitas situações os sinais observados estão contaminados por ruído. Pretende-se neste caso estimar o sinal desejado, $d[n]$, através das observações ruidosas $x[n]$.

Assumindo que o ruído é não correlacionado com o sinal desejado, é possível melhorar a relação sinal/ruído conhecendo as estatísticas do ruído.



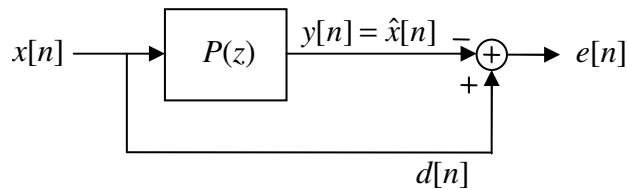
c) Cancelamento de ruído

Um exemplo ilustrativo é o caso de obtenção de um sinal de áudio num ambiente muito ruidoso (por exemplo num cockpit de um avião, devido ao ruído dos motores). Neste caso são captados dois sinais com dois microfones: um perto da boca do locutor (capta o sinal principal) e outro afastado, de tal forma que capte apenas o ruído ambiente (sinal de referência). Como o ruído no sinal principal está correlacionado com o sinal de referência, é possível estimar o ruído presente no sinal principal e assim cancelar este efeito, ou, pelo menos, melhorar a relação sinal/ruído.



d) Predição

Neste caso pretende-se fazer uma predição de um sinal $x[n]$ à custa das amostras anteriores desse sinal (através de um preditor $P(z)$). A estimação será ótima, segundo o critério da média quadrática, se o erro de predição tiver a mínima variância possível.

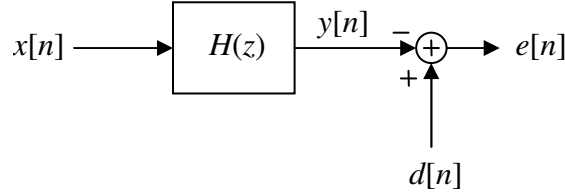


Se o preditor for da forma $P(z) = -\sum_{k=1}^p a_k z^{-k}$, então o sistema global, com entrada $x[n]$ e saída $e[n]$ é o filtro FIR (designado de predição linear)

$$H(z) = 1 - P(z) = 1 + \sum_{k=1}^p a_k z^{-k} = \sum_{k=0}^p a_k z^{-k} \quad , \quad a_0 = 1.$$

4.1 Filtro FIR de Wiener

As diferentes situações identificadas anteriormente podem generalizar-se no diagrama da figura seguinte.



Vejamos como se pode obter o chamado filtro óptimo de Wiener no caso de se tratar de um sistema FIR de ordem p :

$$H(z) = \sum_{k=0}^p h_k z^{-k}.$$

Teremos que

$$e[n] = d[n] - y[n]$$

e

$$y[n] = \sum_{k=0}^p h_k x[n-k],$$

logo

$$e[n] = d[n] - \sum_{k=0}^p h_k x[n-k].$$

Assumimos que $x[n]$ e $d[n]$ são processos estacionários (WSS) com sequências de autocorrelação conhecidas bem como a correlação cruzada entre $d[n]$ e $x[n]$.

Pretende-se calcular os $p+1$ coeficientes h_k do filtro $H(z)$, para $k=0..p$, usando como critério a minimização da média quadrática do erro $e[n]$. Para isso definimos

$$\mathcal{E} = E \left\{ |e[n]|^2 \right\}$$

e as derivadas deste valor em ordem aos coeficientes do filtro têm de ser nulas. Se os processos forem reais, teremos que $\mathcal{E} = E \{ e^2[n] \}$ e que

$$\frac{\partial \mathcal{E}}{\partial h_k} = E \left\{ 2e[n] \cdot \frac{\partial e[n]}{\partial h_k} \right\} = 0, \text{ para } k = 0..p.$$

Como $\partial e[n] / \partial h_k = -x[n-k]$ obtemos as condições

$$E \{ e[n] \cdot x[n-k] \} = 0 \quad , \quad k = 0..p.$$

No caso de processos complexos, pode mostrar-se que as condições resultantes são semelhantes e são as seguintes:

$$E\{e[n] \cdot x^*[n-k]\} = 0 \quad , \quad k = 0..p$$

Diz-se que a solução óptima se obtém quando as observações são ortogonais (ou perpendiculares) ao erro, o que é conhecido como o **princípio da ortogonalidade** ou da perpendicularidade. Pode também mostrar-se que a solução obtida usando estas condições corresponde a um mínimo (e não a um máximo) do erro quadrático. A solução obtém-se introduzindo na expressão anterior a definição do erro:

$$\begin{aligned} E\left\{\left(d[n] - \sum_{m=0}^p h_m x[n-m]\right) x^*[n-k]\right\} \\ = E\{d[n]x^*[n-k]\} - \sum_{m=0}^p h_m E\{x[n-m] \cdot x^*[n-k]\} \end{aligned} \quad k = 0..p.$$

Através da definição de correlação cruzada e da autocorrelação, obtemos, finalmente

$$r_{dx}[k] = \sum_{m=0}^p h_m \cdot r_x[k-m] \quad , \quad k = 0..p.$$

Trata-se de um sistema de $p+1$ equações a $p+1$ incógnitas h_k , sendo estas equações conhecidas como **equações de Wiener-Hopf**. Matricialmente teremos

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-p] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x[-1] \\ r_x[p] & \cdots & r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_p \end{bmatrix} = \begin{bmatrix} r_{dx}[0] \\ r_{dx}[1] \\ \vdots \\ r_{dx}[p] \end{bmatrix}$$

ou

$$\mathbf{R}_x \cdot \mathbf{h} = \mathbf{r}_{dx} \quad ,$$

cujas solução óptima é

$$\mathbf{h} = \mathbf{R}_x^{-1} \cdot \mathbf{r}_{dx}.$$

A matriz do sistema é a matriz de autocorrelação de $x[n]$, que é, como sabemos Toeplitz (todas as diagonais são iguais). No caso de o processo $x[n]$ ser real, esta matriz é também simétrica pois nesse caso $r_x[k] = r_x[-k]$.

Falta apenas conhecer o erro médio quadrático mínimo. Temos que

$$\begin{aligned}
\varepsilon &= E\{|e[n]|^2\} = E\{e[n] \cdot e^*[n]\} \\
&= E\left\{e[n] \left(d^*[n] - \sum_{m=0}^p h_m^* x^*[n-m]\right)\right\} \\
&= E\{e[n]d^*[n]\} - \sum_{m=0}^p h_m^* E\{e[n] \cdot x^*[n-m]\}
\end{aligned}$$

Pelo princípio da ortogonalidade, o 2º termo desta última expressão tem de ser nulo na solução óptima, resultando que o erro médio quadrático mínimo é

$$\begin{aligned}
\varepsilon_{\min} &= E\{e[n] \cdot d^*[n]\} \\
&= E\left\{\left(d[n] - \sum_{m=0}^p h_m x[n-m]\right) d^*[n]\right\} \\
&= r_d[0] - \sum_{m=0}^p h_m \cdot r_{xd}[-m] \\
&= r_d[0] - \sum_{m=0}^p h_m \cdot r_{dx}^*[m] = r_d[0] - \mathbf{r}_{dx}^{*T} \cdot \mathbf{h}
\end{aligned}$$

ou ainda

$$\boxed{\varepsilon_{\min} = r_d[0] - \mathbf{r}_{dx}^{*T} \cdot \mathbf{R}_x^{-1} \cdot \mathbf{r}_{dx}}$$

Exemplo

Se $r_x[k] = \alpha^{|k|}$ e $r_{dx}[k] = \beta^{|k|}$ (com $|\alpha| < 1$ e $|\beta| < 1$), qual o filtro óptimo de Wiener de ordem 1? A solução obtém-se através da equação

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \beta \end{bmatrix}$$

vindo

$$h_0 = \frac{1}{1-\alpha^2} \begin{vmatrix} 1 & \alpha \\ \beta & 1 \end{vmatrix} = \frac{1-\alpha\beta}{1-\alpha^2} ; h_1 = \frac{1}{1-\alpha^2} \begin{vmatrix} 1 & 1 \\ \alpha & \beta \end{vmatrix} = \frac{\beta-\alpha}{1-\alpha^2}$$

e

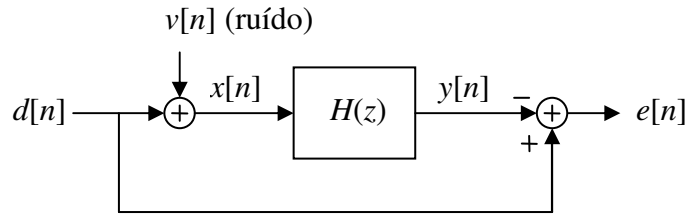
$$H(z) = \frac{1-\alpha\beta}{1-\alpha^2} \left(1 - \frac{\alpha-\beta}{1-\alpha\beta} z^{-1}\right).$$

4.2 Aplicações

a) Cancelamento de eco acústico ou eléctrico

Nesta situação admite-se que o sinal $s[n]$ é não correlacionado com $x[n]$ (ver figura anterior). Assim $d[n]$ contém os “ecos” de $x[n]$ que são evidenciados pela correlação cruzada $r_{dx}[k]$. Conhecendo a sequência de autocorrelação $r_x[k]$ e a correlação cruzada $r_{dx}[k]$, obtemos a solução óptima através de $\mathbf{h} = \mathbf{R}_x^{-1} \cdot \mathbf{r}_{dx}$.

b) Redução de ruído



Nesta situação assume-se que o sinal de interesse, $d[n]$, é não correlacionado com o ruído, $v[n]$, de forma que, sendo $x[n] = d[n] + v[n]$ então

$$\begin{aligned} r_{dx}[k] &= E\{d[n+k]x^*[n]\} \\ &= E\{d[n+k](d^*[n] + v^*[n])\} \\ &= r_d[k] \end{aligned}$$

Por outro lado

$$\begin{aligned} r_x[k] &= E\{x[n+k]x^*[n]\} \\ &= E\{(d[n+k] + v[n+k])(d^*[n] + v^*[n])\} \\ &= r_d[k] + r_v[k] \end{aligned}$$

A equação a resolver, $\mathbf{R}_x \cdot \mathbf{h} = \mathbf{r}_{dx}$, é agora

$$\boxed{(\mathbf{R}_d + \mathbf{R}_v) \cdot \mathbf{h} = \mathbf{r}_d}.$$

A matriz $(\mathbf{R}_d + \mathbf{R}_v) = \mathbf{R}_x$ pode ser estimada através de $x[n]$. Se for possível estimar a autocorrelação do ruído $v[n]$ (por exemplo quando $d[n]=0$), então obtemos também $r_d[k]$ através de $r_d[k] = r_x[k] - r_v[k]$.

Exemplo

Considere-se o caso em que a autocorrelação dos dados é $r_d[k] = \alpha^{|k|}$, $\alpha=0.8$ e $v[n]$ é ruído branco de variância $\sigma_v^2 = 1$. Calcular a relação sinal/ruído na entrada e na saída do filtro de Wiener de ordem 1.

Como a matriz de autocorrelação do ruído é diagonal, então

$$\mathbf{R}_d = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}; \mathbf{R}_v = \sigma_v^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \mathbf{R}_x = \mathbf{R}_d + \mathbf{R}_v = \begin{bmatrix} 1+\sigma_v^2 & \alpha \\ \alpha & 1+\sigma_v^2 \end{bmatrix};$$

A equação a resolver é

$$\begin{bmatrix} 1+\sigma_v^2 & \alpha \\ \alpha & 1+\sigma_v^2 \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix}.$$

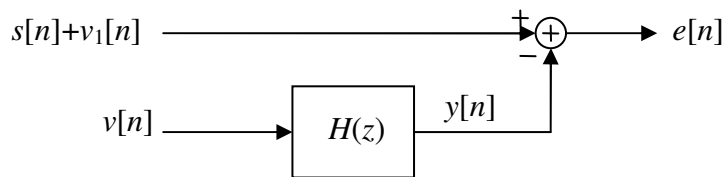
A solução é

$$\begin{bmatrix} h_0 \\ h_1 \end{bmatrix} = \frac{1}{(1+\sigma_v^2)^2 - \alpha^2} \begin{bmatrix} 1+\sigma_v^2 - \alpha^2 \\ \alpha\sigma_v^2 \end{bmatrix}.$$

Com os valores dados, resulta $h_0=0.4048$ e $h_1=0.2381$.

A saída do filtro de Wiener é $y[n] = h[n] * (d[n] + v[n]) = d'[n] + v'[n]$. A relação sinal/ ruído na entrada é $\sigma_d^2 / \sigma_v^2 = 1$, enquanto que na saída do filtro (ver página 24) é $\sigma_{d'}^2 / \sigma_{v'}^2 = \mathbf{h}^* \mathbf{R}_d \mathbf{h} / \mathbf{h}^* \mathbf{R}_v \mathbf{h} = 0.3747 / 0.2205 = 1.6992$. A relação sinal ruído melhorou cerca de $10 \log_{10}(1.6992) \approx 2.3$ dB.

c) Cancelamento de ruído (melhoramento do sinal)



Nesta situação $d[n] = s[n] + v_1[n]$ e $x[n] = v[n]$ (ruído). O sinal $v_1[n]$ não é conhecido mas sabe-se que está correlacionado com o sinal de referência $v[n]$. Sem sinal ($s[n]=0$), vemos que o filtro de Wiener estima o ruído $v_1[n]$ que contamina o sinal $s[n]$. O mesmo se passa na presença do sinal $s[n]$, uma vez que não existe correlação entre este e o ruído, isto é:

$$\begin{aligned} d[n] &= s[n] + v_1[n], \\ x[n] &= v[n]; \end{aligned}$$

$$\begin{aligned}
r_{dx}[k] &= E\{d[n+k]x^*[n]\} \\
&= E\{(s[n+k] + v_1[n])v^*[n]\} \\
&= 0 + r_{v_1v}[k]
\end{aligned}$$

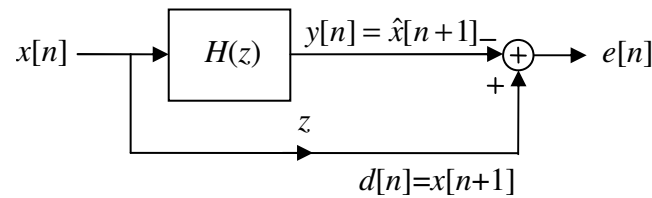
Significa que a correlação cruzada entre o sinal principal e o sinal de referência resulta na correlação entre os dois ruídos.

Como $x[n]=v[n]$, a matriz do sistema a resolver é $\mathbf{R}_x = \mathbf{R}_v$. Logo a equação é

$$\mathbf{R}_v \mathbf{h} = \mathbf{r}_{v_1v}$$

d) Predição

Usualmente considera-se um preditor da amostra actual em função das p amostras anteriores. Uma situação análoga consiste em prever a amostra $x[n+1]$ em função das $p+1$ amostras anteriores. Teremos então o esquema da figura de acordo com o esquema genérico.



Assim

$$\begin{aligned}
r_{dx}[k] &= E\{d[n+k]x^*[n]\} \\
&= E\{x[n+1+k]x^*[n]\} \\
&= r_x[k+1] = r_d[k]
\end{aligned}$$

Logo a equação do sistema é agora

$$\mathbf{R}_x \mathbf{h} = \mathbf{r}_d$$

As equações de Wiener-Hopf transforma-se nas equações de Yuke-Walker:

$$\begin{bmatrix} r_x[0] & r_x[-1] & \cdots & r_x[-p] \\ r_x[1] & r_x[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_x[-1] \\ r_x[p] & \cdots & r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_p \end{bmatrix} = \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[p+1] \end{bmatrix}$$

Nota: podemos considerar, em vez do avanço de uma amostra, o avanço ou atraso de m amostras. No segundo membro teremos nesse caso valores $r_x[\pm m]$.

Exemplo

Se $x[n]$ for o processo AR da saída do filtro $H_1(z) = \frac{1}{1 - 2\rho \cos(\theta)z^{-1} + \rho^2 z^{-2}}$ quando a entrada é ruído branco (de média nula e variância unitária), verificar que o preditor óptimo de ordem $p=2$ de $x[n]$ é $P(z) = 2\rho \cos(\theta)z^{-1} - \rho^2 z^{-2}$.

Pretendemos prever $x[n]$ à custa da estimativa $y[n] = \hat{x}[n] = h_1 x[n-1] + h_2 x[n-2]$. Logo $P(z) = h_1 z^{-1} + h_2 z^{-2}$. Neste caso, como $d[n] = x[n]$, $r_{dx}[k] = r_x[k]$. Queremos estimar h_k para $k=1$ e $k=2$. Obtermos então o sistema

$$r_x[k] = \sum_{m=1}^p h_m \cdot r_x[k-m] = h_1 r_x[k-1] + h_2 r_x[k-2] \quad , \quad k = 1, 2$$

ou:

$$\begin{bmatrix} r_x[0] & r_x[1] \\ r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} r_x[1] \\ r_x[2] \end{bmatrix} = \mathbf{r}.$$

Mas como $x[n]$ é AR de ordem 2, verifica as equações de Yule-Walker:

$$\begin{bmatrix} r_x[0] & r_x[1] \\ r_x[1] & r_x[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -r_x[1] \\ -r_x[2] \end{bmatrix} = -\mathbf{r},$$

pelo que $h_1 = -a_1 = 2\rho \cos(\theta)$ e $h_2 = -a_2 = -\rho^2$. O filtro com entrada $x[n]$ e saída $e[n]$ é o filtro de erro, $A_1(z) = 1 - P(z) = 1 - 2\rho \cos(\theta)z^{-1} + \rho^2 z^{-2}$, que é exactamente o **filtro inverso** de $H_1(z)$. Significa que a predição linear de processos AR pode ser exacta, desde que o filtro de Wiener tenha ordem igual ou superior à do processo AR. O erro de média quadrática mínimo é

$$\mathcal{E}_{\min} = r_x[0] - \mathbf{h}^{*T} \mathbf{r} = \sum_{k=0}^p a_k r_x[k], \quad a_0 = 1.$$

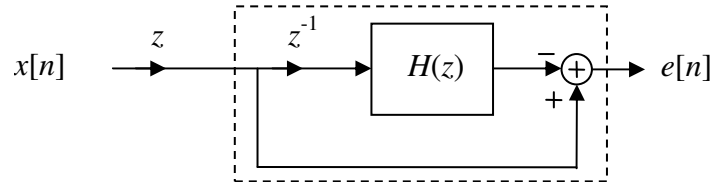
Vejamos agora o caso de um preditor das amostras $x[n+1]$ (o caso discutido anteriormente). A equação a resolver é agora (com $r[k] = r_x[k]$):

$$\begin{bmatrix} r[0] & r[1] & r[2] \\ r[1] & r[0] & r[1] \\ r[2] & r[1] & r[0] \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} r[1] \\ r[2] \\ r[3] \end{bmatrix}.$$

Como vimos, podemos calcular a sequência de autocorrelação de um processo AR se conhecermos os coeficientes a_k (ver exemplo sobre processos AR). Se fizermos isso, resulta que a solução desta equação é $h_0 = -a_1$, $h_1 = -a_2$ e $h_2 = 0$. Significa que o filtro óptimo é $H(z) = 2\rho \cos(\theta) - \rho^2 z^{-1}$. Mas, neste caso o filtro de erro vai ser dado através de $E(z) = zX(z) - H(z)X(z) = (z - H(z))X(z)$. Logo

$$\begin{aligned}
 A_2(z) &= E(z) / X(z) = z - H(z) \\
 &= z - h_0 - h_1 z^{-1} \\
 &= z(1 - h_0 z^{-1} - h_1 z^{-2}) = z(1 + a_1 z^{-1} + a_2 z^{-2}) \\
 &= zA_1(z)
 \end{aligned}$$

Equivale a redesenhar o sistema para a seguinte forma onde $z^{-1}H(z)$ é o preditor anterior.



■

Apêndice 1 - Factorização Espectral de Funções Racionais

Pretende-se obter uma forma genérica de factorizar uma expressão de um espectro de potência regular, na forma $P(z)=H(z)H(z^{-1})$, de maneira a obter um filtro $H(z)=B(z)/A(z)$, causal, estável e de coeficientes reais. Toma-se a solução de fase mínima (zeros e pólos no interior do círculo de raio unitário). Seja então um filtro $H(z)$ com M zeros z_k e N pólos p_k :

$$H(z) = A \frac{\prod_{k=1}^M (1 - z_k z^{-1})}{\prod_{k=1}^N (1 - p_k z^{-1})}. \quad (A1)$$

Teremos

$$P(z) = H(z)H(z^{-1}) = A^2 \frac{\prod_{k=1}^M (1 - z_k z^{-1})(1 - z_k z)}{\prod_{k=1}^N (1 - p_k z^{-1})(1 - p_k z)} \quad (A2)$$

ou seja

$$P(z) = A^2 \frac{\prod_{k=1}^M (1 + z_k^2 - z_k(z + z^{-1}))}{\prod_{k=1}^N (1 + p_k^2 - p_k(z + z^{-1}))} \quad (A3)$$

O espectro de potência fica definido em termos de polinómios de coeficientes reais em $\cos(\omega)$:

$$P(e^{j\omega}) = A^2 \frac{\prod_{k=1}^M (1 + z_k^2 - 2z_k \cos(\omega))}{\prod_{k=1}^N (1 + p_k^2 - 2p_k \cos(\omega))} \quad (A4)$$

Se existirem termos em $\cos(n\omega)$ usamos a expressão

$$\cos(n\omega) = 2\cos((n-1)\omega)\cos(\omega) - \cos((n-2)\omega) \quad (A5)$$

(ou os polinómios de Chebychev², $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$ onde $T_0(x)=1$ e $T_1(x)=x$) para passar a potências de $\cos(\omega)$. Definindo agora $x = \cos(\omega)$ obtemos

$$P(x) = A^2 \frac{\prod_{k=1}^M (1 + z_k^2 - 2z_k x)}{\prod_{k=1}^N (1 + p_k^2 - 2p_k x)} \quad (A6)$$

² ver, por exemplo: http://en.wikipedia.org/wiki/Chebyshev_polynomials

As raízes do numerador e denominador em função de x são calculadas a partir da expressão do espectro de potência. Dada uma raiz de x do numerador, obtemos o zero z_k através da equação $z_k^2 - 2z_kx + 1 = 0$, vindo:

$$z_k = x \pm \sqrt{x^2 - 1}. \quad (\text{A7})$$

As duas soluções são inversas uma da outra, pelo que se escolhe aquela que tiver modulo inferior a um (para que o sistema seja de fase mínima). Para o denominador faz-se o mesmo, obtendo-se os pólos:

$$p_k = x \pm \sqrt{x^2 - 1}. \quad (\text{A8})$$

Depois de calcular as raízes z_k e p_k , pode colocar-se a expressão na forma (A2). O factor A pode ser obtido comparando (A4) com a expressão original do espectro de potência.

Exemplo 1

$$P(e^{j\omega}) = \frac{5 + 4\cos(2\omega)}{10 + 6\cos(\omega)}$$

Como $\cos(2\omega) = 2\cos^2(\omega) - 1 = 2x^2 - 1$, passamos a expressão anterior a

$$P(x) = \frac{5 + 4(2x^2 - 1)}{10 + 6x} = \frac{8x^2 + 1}{10 + 6x}.$$

Raízes do numerador: $x_1 = +j\frac{1}{2\sqrt{2}}; x_2 = -j\frac{1}{2\sqrt{2}}$

Raízes do denominador: $x_1 = -\frac{5}{3}$

Zeros: $z_1 = x_1 \pm \sqrt{x_1^2 - 1} = j\frac{1}{2\sqrt{2}} \pm j\frac{3}{2\sqrt{2}} \rightarrow z_1 = -j\frac{1}{\sqrt{2}}$

$z_2 = x_2 \pm \sqrt{x_2^2 - 1} = -j\frac{1}{2\sqrt{2}} \pm j\frac{3}{\sqrt{8}} \rightarrow z_2 = j\frac{1}{\sqrt{2}}$

Pólos: $p_1 = x_1 \pm \sqrt{x_1^2 - 1} = -\frac{5}{3} \pm \frac{4}{3} \rightarrow p_1 = -\frac{1}{3}$

Logo

$$H(z) = A \frac{(1 - j\frac{1}{\sqrt{2}}z^{-1})(1 + j\frac{1}{\sqrt{2}}z^{-1})}{1 + \frac{1}{3}z^{-1}} = A \frac{1 + \frac{1}{2}z^{-2}}{1 + \frac{1}{3}z^{-1}}$$

Para $z=1$, temos $H(z=1) = A \frac{\frac{3}{2}}{\frac{4}{3}} = \frac{9}{8}A$.

Mas $P(e^{j0}) = \frac{5+4}{10+6} = \frac{9}{16} = P(z=1) \Rightarrow (\frac{9}{8}A)^2 = \frac{9}{16} \Rightarrow A^2 = \frac{4}{9}$. Escolhendo $A = \frac{2}{3}$, resulta

$$H(z) = \frac{2}{3} \frac{1 + \frac{1}{2}z^{-2}}{1 + \frac{1}{3}z^{-1}}$$

(a solução da página 24).

Exemplo 2

Seja
$$P(e^{j\omega}) = \frac{81}{4} \frac{17 - 8\cos(\omega)}{(5 + 4\cos(\omega))(91 - 60\cos(\omega) + 18\cos(2\omega))}.$$

Com $\cos(2\omega) = 2x^2 - 1$, temos que

$$P(x) = \frac{81}{4} \frac{17 - 8x}{(5 + 4x)(101 - 60x + 18(2x^2 - 1))} = \frac{81}{4} \frac{17 - 8x}{(5 + 4x)(73 - 60x + 36x^2)}$$

Raízes do numerador: $x = \frac{17}{8}$

Raízes do denominador: $x_1 = -\frac{5}{4}; x_{2,3} = -\frac{5}{6} \pm j \frac{2}{\sqrt{3}}$

Zeros: $z_1 = x \pm \sqrt{x^2 - 1} = \frac{17}{8} \pm \frac{15}{8} \rightarrow z_1 = \frac{1}{4}$

Pólos: $p_1 = x_1 \pm \sqrt{x_1^2 - 1} = -\frac{5}{4} \pm \frac{3}{4} \rightarrow p_1 = -\frac{1}{2}$
 $p_2 = \frac{1}{3} e^{j\pi/3}; p_3 = p_2^* = \frac{1}{3} e^{-j\pi/3}$

Logo

$$H(z) = A \frac{1 - \frac{1}{4} z^{-1}}{(1 + \frac{1}{2} z^{-1})(1 - \frac{1}{3} z^{-1} + \frac{1}{9} z^{-2})}$$

$$H(z=1)^2 = A^2 \left[\frac{1 - \frac{1}{4}}{(1 + \frac{1}{2})(1 - \frac{1}{3} + \frac{1}{9})} \right]^2 = P(e^{j0}) = \frac{81}{4} \frac{9}{(9)(91 - 60 + 18)} \Rightarrow A^2 = 1$$

Fazendo $A=1$, vem

$$H(z) = \frac{1 - \frac{1}{4} z^{-1}}{(1 + \frac{1}{2} z^{-1})(1 - \frac{1}{3} z^{-1} + \frac{1}{9} z^{-2})}.$$

Exemplo 3

Seja $H(z) = \frac{1 + z^{-2}}{1 - \frac{1}{2} z^{-1}}$ (com um par de zeros em $z = \pm j$). Neste caso temos que

$$P(z) = H(z)H(z^{-1}) = \frac{(1 + z^{-2})(1 + z^2)}{(1 - \frac{1}{2} z^{-1})(1 - \frac{1}{2} z)} = \frac{2 + z^{-2} + z^2}{\frac{5}{4} - \frac{1}{2}(z + z^{-1})}$$

e a expressão a factorizar será

$$P(e^{j\omega}) = |H(e^{j\omega})|^2 = \frac{2 + 2\cos(2\omega)}{\frac{5}{4} - \cos(\omega)}.$$

Com $x = \cos(\omega)$, $P(x) = \frac{4x^2}{\frac{5}{4} - x}$. Se aplicarmos o método a este caso, temos um zero duplo em $x=0$. Zeros duplos correspondem a zeros z_k na circunferência de raio unitário e têm que ser complexas conjugadas (caso de coeficientes reais). Neste caso $z_k = x \pm \sqrt{x^2 - 1} = \pm j$. Quanto ao pólo, temos $p_1 = \frac{5}{4} - \frac{3}{4} = \frac{1}{2}$, vindo (com $A=1$)

$$H(z) = \frac{(1 - jz^{-1})(1 + jz^{-1})}{1 - \frac{1}{2} z^{-1}} = \frac{1 + z^{-2}}{1 - \frac{1}{2} z^{-1}}$$

como deveria ser. []