Filtragem Adaptativa

Charles Casimiro Cavalcante

charles@gtel.ufc.br

Grupo de Pesquisa em Telecomunicações Sem Fio – GTEL
Programa de Pós-Graduação em Engenharia de Teleinformática
Universidade Federal do Ceará – UFC
http://www.gtel.ufc.br/∼charles

"Filtros adaptativos, os quais têm como meta transformar os sinais portadores de informação em versões 'limpas' ou 'melhoradas', ajustam suas características de acordo com os sinais encontrados. Eles formam os exemplos mais simples de algoritmos no campo de aprendizado de máquinas."

Philip A. Regalia, 2005

IEEE Control Systems Magazine, Agosto de 2005

Conteúdo do curso

1. Introdução
2. Revisão de Processos Estocásticos
3. Filtragem Linear Ótima
4. Algoritmos Recursivos no Tempo
5. Método dos Mínimos Quadrados
6. Estruturas Alternativas de Filtragem Adaptativa
7. Tópicos Avançados

Parte IV

Filtragem Linear Ótima

Filtragem

Processamento de sinais para extração ou modificação de certas características do sinal

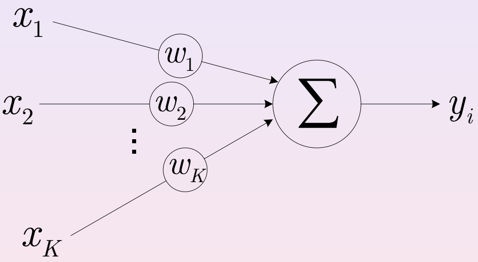
Clássico *versus* moderno

- Problema de filtragem "moderna": Wiener e Kolmogorov nos anos 40
- Processar, de maneira ótima, sinais aleatórios que ocupam (geralmente) mesma faixa de freqüência
- Técnicas clássicas de PDS não satisfazem
- Otimização: escolha de critério que ressalte as características de interesse do sinal segundo uma estrutura de processamento

Escolhas

- Estrutura: **linear** - grande potencial de aplicação e simplicidade de análise
- Critério: minimização do erro médio quadrático (**MMSE**)
- Cenário: sinal de treinamento disponível (**supervisionado**)

Estrutura básica



$$\overbrace{\boxed{y_i = \sum_{j=1}^{K} w_i x_i}}^{\text{combinação linear dos dados e parâmetros}}$$

$$y_i = \mathbf{w}^T \mathbf{x}$$

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_K \end{bmatrix}^T$$

$$\mathbf{w} = \begin{bmatrix} w_1 & w_2 & \dots & w_K \end{bmatrix}^T$$

Meta

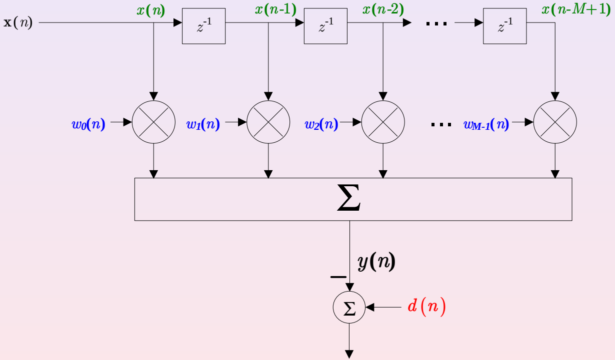
Problema

Otimizar $\mathbf{w}$, isto é, calcular $\mathbf{w}$ para que $y_i = d_i$ (sinal desejado)

- Solução para $\mathbf{w}$ ótimo: filtragem de Wiener
- Esse problema dá origem a duas configurações fundamentais que nos interessam

Primeira estrutura: filtragem temporal, em que $x_1, x_2, \ldots, x_K$ são amostras temporais de um sinal $x(n)$

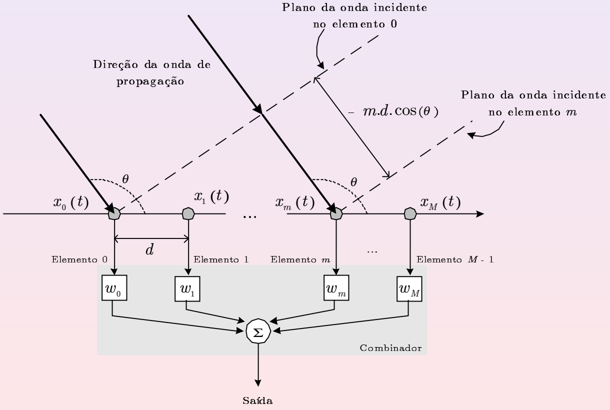
Problema: equalização de canais (Lucky, 1965)



$x(n)$ e $d(n)$ são processos estocásticos estacionários discretos

Segunda estrutura: filtragem espacial, em que $x_1, x_2, \ldots, x_K$ são amostras espaciais de um sinal $x(n)$ incidindo no conjunto de sensores (arranjo)

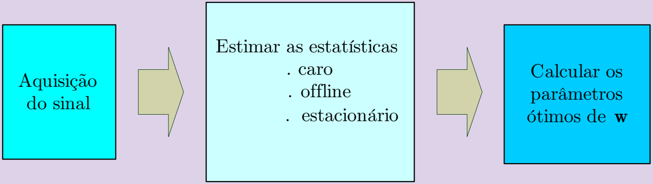
Problema: antenas adaptativas (Widrow, 1960)



Filtragem × otimização

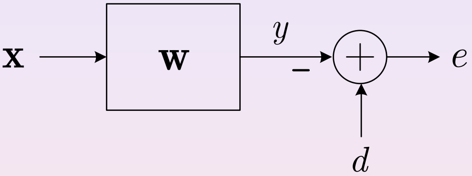
Filtragem ótima

- Boa base teórica
- Calcular $\mathbf{w}$ conhecendo as estatísticas dos sinais envolvidos
- Tipicamente o modelo abaixo



Modelagem

Sistema geral



1. Combinador linear: $\mathbf{x}$ composto de amostras espaciais
2. Filtro FIR transversal: $\mathbf{x}$ composto de amostras temporais

Modelagem - cont.

Considerando a filtragem temporal, temos então

Meta

Minimizar $\mathbb{E}\left\{e^2(n)\right\}$

- Filtro de comprimento M
- $e(n) = d(n) - y(n)$
- $\mathbf{x}(n) = \begin{bmatrix} x(n) & \cdots & x(n-M+1) \end{bmatrix}^T$
- $\mathbf{w} = \begin{bmatrix} w_0 & w_1 & \cdots & w_{M-1} \end{bmatrix}^T$

Considerações:

- Sinal $x(n)$ é estacionário e de média nula

$$\Rightarrow r(i,j) = \mathbb{E}\left\{x(n-i)x(n-j)\right\} = r(i-j)$$

- Sinal $d(n)$ é estacionário, de média nula e com variância igual a σ_d^2

Derivando filtro ótimo

Do modelo, temos então:

$$\begin{aligned} e(n) &= d(n) - y(n) \\ &= d(n) - \mathbf{w}^T \mathbf{x}(n) \end{aligned} \tag{68}$$

e

$$\begin{aligned} e^2(n) &= \big(d(n) - \mathbf{w}^T \mathbf{x}(n)\big) \cdot \big(d(n) - \mathbf{w}^T \mathbf{x}(n)\big)^T \\ &= d^2(n) - 2\mathbf{w}^T \mathbf{x}(n) d(n) + \mathbf{w}^T \mathbf{x}(n) \mathbf{x}^T(n) \mathbf{w} \end{aligned} \tag{69}$$

Derivando filtro ótimo - cont.

Aplicando o operando esperança...

$$\mathbb{E}\left\{e^2(n)\right\} = \underbrace{\mathbb{E}\left\{d^2(n)\right\}}_{\substack{\text{variância} \\ \text{de } d(n)}} - 2\mathbf{w}^T \underbrace{\mathbb{E}\left\{\mathbf{x}(n)d(n)\right\}}_{\substack{\text{correlação} \\ \text{cruzada}}} + \mathbf{w}^T \underbrace{\mathbb{E}\left\{\mathbf{x}(n)\mathbf{x}^T(n)\right\}}_{\substack{\text{matrix de} \\ \text{autocorrelação}}} \mathbf{w} \tag{70}$$

Assim, temos

$$\mathbb{E}\left\{e^2(n)\right\} = \sigma_d^2 - 2\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \mathbf{w}^T\mathbf{R}_{\mathbf{x}}\mathbf{w} \tag{71}$$

Como a equação é quadrática em relação aos parâmetros $\mathbf{w}$, existe somente um ponto de mínimo (máximo)

Derivando filtro ótimo - cont.

Achar o ponto ótimo é então equivalente a encontrar o ponto onde a função tem o seu mínimo, ou seja

$$\nabla_{\mathbf{w}} \mathbb{E}\left\{e^2(n)\right\} = 0 \tag{72}$$

Para simplificar a notação, podemos chamar $\mathbb{E}\left\{e^2(n)\right\} = \varepsilon$. Logo, a Eq. (72) nos leva à derivação da Eq. (71) em relação aos parâmetros $\mathbf{w}$, ou seja

$$\nabla_{\mathbf{w}} \varepsilon = \frac{\partial \varepsilon}{\partial \mathbf{w}} = \begin{bmatrix} \frac{\partial \varepsilon}{\partial w_0} & \frac{\partial \varepsilon}{\partial w_1} & \cdots & \frac{\partial \varepsilon}{\partial w_{M-1}} \end{bmatrix}^T \tag{73}$$

Derivando filtro ótimo - cont.

Então, temos

$$-2\mathbf{p}_{\mathbf{x}d} + 2\mathbf{R}_{\mathbf{x}}\mathbf{w} = 0 \quad (74)$$

Da qual, após multiplicação à esquerda por $\mathbf{R}_{\mathbf{x}}^{-1}$, obtém a equação do filtro ótimo:

$$\boxed{\mathbf{w}_{\text{opt}} = \mathbf{R}_{\mathbf{x}}^{-1} \cdot \mathbf{p}_{\mathbf{x}d}} \quad (75)$$

A Equação (75) é chamada então de *Equação de Wiener-Hopf*, ou conjunto de equações normais, e é por vezes escrita como

$$\sum_{i=0}^{M-1} w_{\text{opt},i} r_x(i-k) = p_{xd}(k), \quad k = 0, 1, \ldots, M-1 \quad (76)$$

em que $r(i-k)$ é a correlação para os instantes i e j e $p_{xd}(k)$ é a correlação cruzada entre $x(n-k)$ e $d(n)$.

Valor do erro quadrático mínimo

De posse do filtro ótimo, podemos então calcular o valor mínimo do erro médio quadrático, ou seja,

$$\mathbf{w} = \mathbf{w}_{\text{opt}} \Rightarrow \mathbb{E}\left\{e^2(n)\right\}\Big|_{\text{mínimo}}$$

Substituindo a Eq. (75) em (71), obtém-se

$$\begin{aligned}
\varepsilon_{\text{min}} &= \sigma_d^2 - 2\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \mathbf{w}^T\mathbf{R}_{\mathbf{x}}\mathbf{w} \\
&= \sigma_d^2 - 2\mathbf{p}_{\mathbf{x}d}^T\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p}_{\mathbf{x}d} + \mathbf{p}_{\mathbf{x}d}^T\mathbf{R}_{\mathbf{x}}^{-1}\overbrace{\cancel{\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{-1}}}^{\mathbf{I}}\mathbf{p}_{\mathbf{x}d} \qquad (77) \\
&\boxed{= \sigma_d^2 - \mathbf{p}_{\mathbf{x}d}^T\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p}_{\mathbf{x}d}}
\end{aligned}$$

Características do cálculo filtro ótimo

- Depende **somente das estatísticas de segunda ordem** dos sinais envolvidos
- Em geral, estimativas precisas de $\mathbf{R}_{\mathbf{x}}$ e $\mathbf{p}_{\mathbf{x}d}$ não são disponíveis na prática
- Considerando-se a ergodicidade, é possível utilizar médias temporais para estima-las
- Supõe-se que a inversa da matriz $\mathbf{R}_{\mathbf{x}}$ existe. Na prática, resolve-se o sistema linear $\mathbf{R}_{\mathbf{x}}\mathbf{w}_{\text{opt}} = \mathbf{p}_{\mathbf{x}d}$.
- Caso do combinador linear: mesma solução. A diferença reside no cálculo das correlações
- Extensivo ao caso complexo. Definição do gradiente em relação a parâmetros complexos. Mesma solução!

Princípio da ortogonalidade

Uma questão interessante é verificada por meio da Eq. (72). Ela implica que o gradiente deve ser nulo em relação a **todos** os parâmetros w_i, ou seja

$$\nabla_{\mathbf{w}} \mathbb{E}\left\{e^2(n)\right\} = 0 \Rightarrow \mathbb{E}\left\{\frac{\partial e^2(n)}{\partial w_i}\right\} = 0 \;\forall i = 0, \ldots, M-1 \quad (78)$$

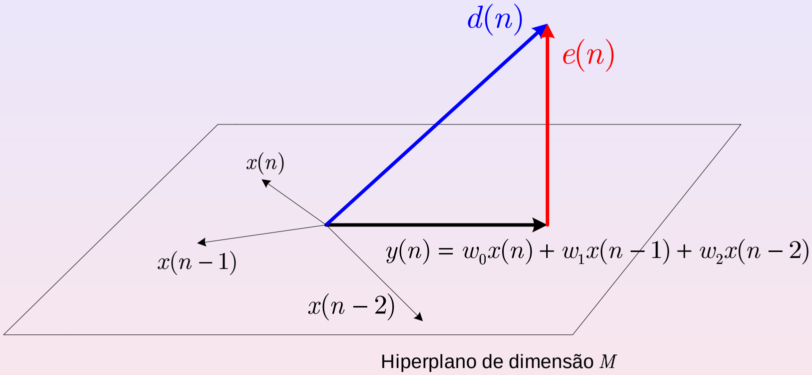
Lembrando que $e(n) = \left[d(n) - \sum_{i=0}^{M-1} w_i x(n-i)\right]$, então temos,

$$\mathbb{E}\left\{2 \cdot \frac{\partial e(n)}{\partial w_i} \cdot e(n)\right\} = 0$$

$$\boxed{\mathbb{E}\left\{x(n-i) \cdot e(n)\right\} = 0 \;\forall i} \quad (79)$$

⚠ Ou seja, $x(n-i)$ e $e(n)$ são **ortogonais**!

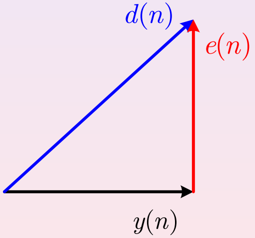
Princípio da ortogonalidade - cont.



Critério de minimização do erro quadrático médio equivale a um critério de ortogonalização!

Desta forma, estamos interessados em $d(n)$ colinear a $y(n)$ pois nesta condição

$$\exists\ \mathbf{w} \mid e(n) = 0 \Rightarrow \mathbb{E}\left\{e^2(n)\right\} = 0$$



Minimizar erro quadrático médio

$\Updownarrow$

Tornar erro ortogonal à saída do filtro

Meta: observar como se comporta o filtro quando está em torno da solução ótima.

Tomando-se a curva fornecida pelo erro médio quadrado temos a seguinte expressão:

$$\varepsilon = \mathbb{E}\left\{e^2(n)\right\} = \sigma_d^2 - 2\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \mathbf{w}^T\mathbf{R}_\mathbf{x}\mathbf{w}$$

Definindo $\Delta\mathbf{w} = \mathbf{w} - \mathbf{w}_{\text{opt}}$, e substituindo em (71) temos

$$\begin{aligned}
\varepsilon &= \sigma_d^2 - 2(\Delta\mathbf{w} + \mathbf{w}_{\text{opt}})^T\mathbf{p}_{\mathbf{x}d} + (\Delta\mathbf{w} + \mathbf{w}_{\text{opt}})^T\mathbf{R}_\mathbf{x}(\Delta\mathbf{w} + \mathbf{w}_{\text{opt}}) \\
&= \underbrace{\sigma_d^2 - 2\mathbf{w}_{\text{opt}}^T\mathbf{p}_{\mathbf{x}d} + \mathbf{w}_{\text{opt}}^T\mathbf{R}_\mathbf{x}\mathbf{w}_{\text{opt}}}_{\varepsilon_{\text{min}}} - 2\Delta\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \Delta\mathbf{w}^T\mathbf{R}_\mathbf{x}\Delta\mathbf{w} \\
&\quad + \Delta\mathbf{w}^T\mathbf{R}_\mathbf{x}\mathbf{w}_{\text{opt}} + \mathbf{w}_{\text{opt}}\mathbf{R}_\mathbf{x}\Delta\mathbf{w}
\end{aligned} \tag{80}$$

continuando...

$$
\begin{aligned}
\varepsilon &= \varepsilon_{\text{min}} - 2\Delta\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \Delta\mathbf{w}^T\mathbf{R}_{\mathbf{x}}\Delta\mathbf{w} + \Delta\mathbf{w}^T\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{p}_{\mathbf{x}d} \\
&\qquad + \mathbf{p}_{\mathbf{x}d}^T\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{R}_{\mathbf{x}}\Delta\mathbf{w} \\
&= \varepsilon_{\text{min}} - 2\Delta\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \Delta\mathbf{w}^T\mathbf{R}_{\mathbf{x}}\Delta\mathbf{w} + \Delta\mathbf{w}^T\mathbf{p}_{\mathbf{x}d} + \mathbf{p}_{\mathbf{x}d}^T\Delta\mathbf{w} \\
&= \varepsilon_{\text{min}} + \Delta\mathbf{w}^T\mathbf{R}_{\mathbf{x}}\Delta\mathbf{w}
\end{aligned}
\tag{81}
$$

Comportamento da curva MSE (*mean square error*) - cont.

Podemos ainda diagonalizar a matriz $\mathbf{R_x} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ em que $\mathbf{Q}$ é a matriz (ortogonal) dos autovetores de $\mathbf{R_x}$ dada por

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_M \end{bmatrix}$$

e

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_M \end{bmatrix}$$

daí, pode-se escrever

$$\varepsilon = \varepsilon_{\text{min}} + \Delta\mathbf{w}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\Delta\mathbf{w} \tag{82}$$

Define-se ainda o vetor $\mathbf{v}$ de parâmetros v_i tal que

$$\mathbf{v} = \mathbf{Q}^T \Delta \mathbf{w} \tag{83}$$

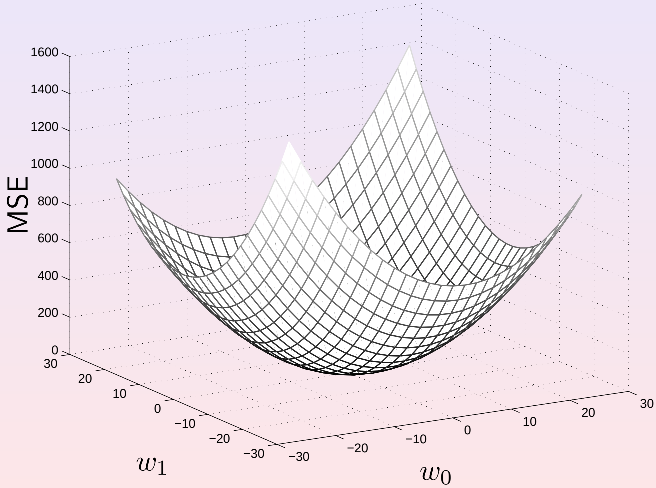
o que leva a

$$\varepsilon = \varepsilon_{\text{min}} + \mathbf{v}^T \mathbf{\Lambda} \mathbf{v} \tag{84}$$

Vantagem: $\mathbf{\Lambda}$ é uma matriz diagonal ao passo que $\mathbf{R}_\mathbf{x}$ não.

Comportamento da curva MSE (*mean square error*) - cont.

Filtro com dois coeficientes

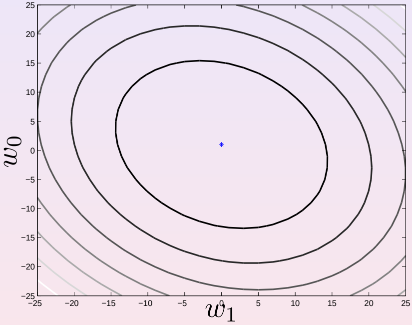


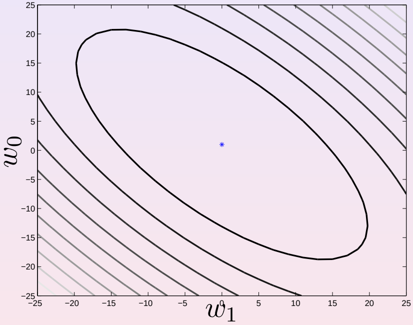
$$\varepsilon(v_0, v_1) = \varepsilon_{\mathsf{min}} + \lambda_1 v_1^2 + \lambda_2 v_2^2 \tag{85}$$

Lembrando que: $\mathbb{C}\left(\mathbf{R_x}\right) = \frac{\lambda_{\max}}{\lambda_{\min}}$ (número de condicionamento)

1. $\mathbb{C}\left(\mathbf{R_x}\right) \approx 1$ $\Leftrightarrow$ curvas MSE mais circulares $\Leftrightarrow$ $x(n)$ tem espectro mais plano
2. $\mathbb{C}\left(\mathbf{R_x}\right) \gg 1$ $\Leftrightarrow$ curvas MSE mais "elípticas" $\Leftrightarrow$ $x(n)$ tem espectro com picos

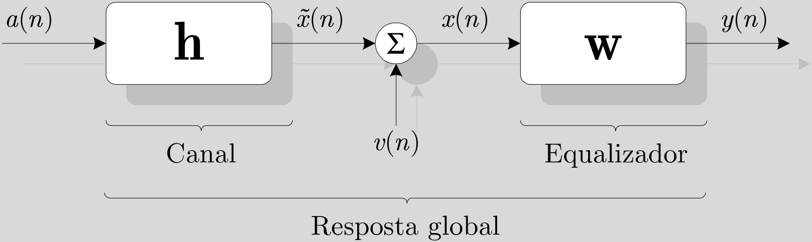
Curvas de nível em função do número de condicionamento





$$\mathbb{C}\left(\mathbf{R_x}\right) = 1.5 \qquad \mathbb{C}\left(\mathbf{R_x}\right) = 5.6667$$

Exemplo


$a(n)$
$\mathbf{h}$
$\tilde{x}(n)$
Σ
$x(n)$
$\mathbf{w}$
$y(n)$
Canal
$v(n)$
Equalizador
Resposta global

Exemplo - cont.

Buscar o filtro linear ótimo, no sentido da minimização do erro quadrático médio, que inverta o seguinte canal:

$$H(z) = 1 + 0.7z^{-1}$$

Assume-se que o alfabeto de transmissão é BPSK, ou seja, $a(n) \in \{-1, +1\}$ com igual probabilidade e que o sinal desejado será o do instante atual, ou seja, $d(n) = a(n)$.

Para este problema temos que calcular $\mathbf{R}_{\tilde{\mathbf{x}}}$ e $\mathbf{p}_{\mathbf{x}d}$ dadas para o problema em questão por

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \begin{bmatrix} 1.49 & 0.7 \\ 0.7 & 1.49 \end{bmatrix} \quad \text{e} \quad \mathbf{p}_{\mathbf{x}d} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{86}$$

Exemplo - cont.

Supondo ainda um ruído aditivo branco na entrada do filtro (receptor) de SNR = 20 dB, teremos então alguma perturbação no sinal recebido e em conseqüência na sua correlação. Assim, teremos $\mathbf{R_x} = (\mathbf{R}_{\tilde{\mathbf{x}}} + \sigma_v \mathbf{I})$.

De posse de tais quantidades, podemos então calcular

$$\mathbf{w}_{\text{opt}} = \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{p}_{\mathbf{x}d} = \begin{bmatrix} 0.8579 \\ -0.4058 \end{bmatrix} \quad (87)$$

E podemos ainda calcular a resposta global, fruto da convolução do canal com o equalizador:

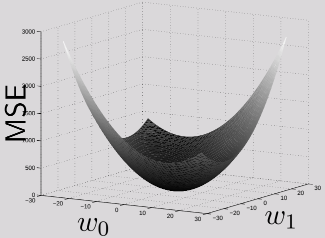
$$\mathbf{g} = \begin{bmatrix} 0.8579 & 0.1948 & -0.2840 \end{bmatrix}^T$$

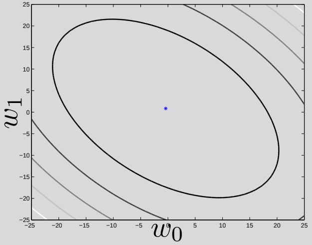
Aplicação - equalização de canais

Exemplo - cont.

Ainda temos,

$$\varepsilon_{\text{min}} = 0.1421$$





$$\mathcal{C}\left(\mathbf{R_x}\right) = 2.7890$$

O problema de predição

Definição

Predição: estimar uma amostra $x(n)$ a partir de um conjunto de valores conhecidos deste sinal.

É, em essência, um processo de filtragem no qual o sinal desejado é uma nova amostra da seqüência de entrada do filtro.

Linear: $\widehat{x}(n)$ é uma combinação linear dos valores passados.

Forward ou **backward**:

1. *Forward*: prever uma amostra futura a partir de uma coleção de amostras passadas
2. *Backward*: prever uma amostra no passado (desconhecida) a partir de um conjunto de amostras, inclusive presente

O problema de predição - cont.

De passo k

1. *Forward:*

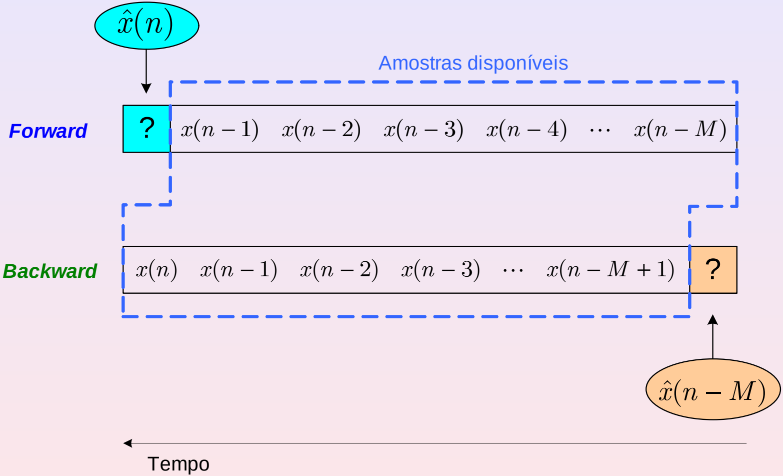
$$\widehat{x}(n) = \sum_{i=k}^{M+k-1} w_{f,i} \cdot x(n-i) \tag{88}$$

2. *Backward:*

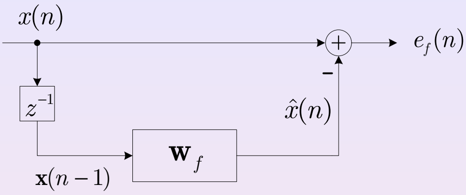
$$\widehat{x}(n-M-k+1) = \sum_{i=1}^{M-1} w_{b,i} \cdot x(n-i+1) \tag{89}$$

Inicialmente, nos deteremos nos processos de predição de passo unitário.

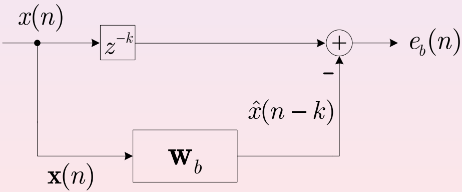
O problema de predição - cont.



O problema de predição - cont.



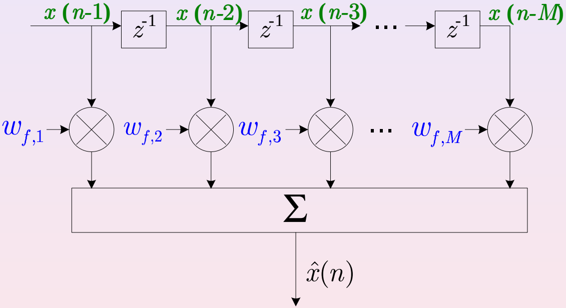
(a) Forward



(b) Backward

Estrutura do filtro

Para o preditor, usamos um filtro de linha de atraso (filtro FIR)



Modelagem

Para o caso de predição de um passo *forward* temos então

$$\widehat{x}(n) = \mathbf{w}_f^T \cdot \mathbf{x}(n-1) \tag{90}$$

em que $\mathbf{x}(n-1) = \begin{bmatrix} x(n-1) & \cdots & x(n-M) \end{bmatrix}^T$ e $\mathbf{w}_f = \begin{bmatrix} w_{f,1} & \cdots & w_{f,M} \end{bmatrix}^T$

E também

$$e_f(n) = x(n) - \widehat{x}(n) \tag{91}$$

Meta: tornar $\widehat{x}(n)$ o mais similar possível de $x(n)$.

Critério

Minimização do erro de predição quadrático!

$$\min_{\mathbf{w}_f} \mathbb{E}\left\{e_f^2(n)\right\} \tag{92}$$

Filtragem ótima $\times$ predição

Analogia

O problema é um caso particular da filtragem de Wiener em que

1. $d(n) \leftrightarrow x(n)$,
2. $\mathbf{x}(n) \leftrightarrow \mathbf{x}(n-1)$,
3. $w_i \; (i = 0, \ldots, M-1) \; \leftrightarrow w_{f,i} \; (i = 1, \ldots, M-1)$ e
4. $y(n) \leftrightarrow \widehat{x}(n)$

Hipóteses

1. $x(n)$ é um sinal estacionário no sentido amplo, que implica

$$r_x(i,j) = \mathbb{E}\left\{x(n-i)x(n-j)\right\} = r(i-j) \qquad (93)$$

2. $x(n)$ tem média nula e variância σ_x^2

Filtro preditor ótimo

Sabendo que

$$\begin{aligned} e_f(n) &= x(n) - \widehat{x}(n) \\ &= x(n) - \mathbf{w}_f^T \cdot \mathbf{x}(n-1) \\ e_f^2(n) &= x^2(n) - 2\mathbf{w}_f^T \cdot \mathbf{x}(n-1)x(n) + \mathbf{w}_f^T \cdot \mathbf{x}(n-1)\mathbf{x}(n-1)^T\mathbf{w}_f \end{aligned} \tag{94}$$

Temos então

$$\begin{aligned} \mathbb{E}\left\{e_f^2(n)\right\} &= \mathbb{E}\left\{x^2(n)\right\} - 2\mathbf{w}_f^T\mathbb{E}\left\{\mathbf{x}(n-1)x(n)\right\} \\ &\quad + \mathbf{w}_f^T\mathbb{E}\left\{\mathbf{x}(n-1)\mathbf{x}(n-1)^T\right\}\mathbf{w}_f \\ &= r_x(0) - 2\mathbf{w}_f^T\mathbf{r}_{x,f} + \mathbf{w}_f^T\mathbf{R}_\mathbf{x}\mathbf{w}_f \end{aligned} \tag{95}$$

em que $\mathbf{R}_\mathbf{x}$ e $\mathbf{r}_{x,f}$ são a matriz de autocorrelação de $\mathbf{x}(n-1)$ e o vetor de correlação entre $\mathbf{x}(n-1)$ e $x(n)$, respectivamente.

Filtro preditor ótimo - cont.

Para achar o valor ótimo, temos

$$\frac{\partial \mathbb{E}\left\{e_f^2(n)\right\}}{\partial \mathbf{w}_f} = 0 \tag{96}$$

$$-2\mathbf{r}_{x,f} + 2\mathbf{R}_\mathbf{x}\mathbf{w}_f = 0$$

O que nos fornece

Preditor *forward* ótimo

$$\mathbf{w}_{f,\text{opt}} = \mathbf{R}_\mathbf{x}^{-1} \cdot \mathbf{r}_{x,f} \tag{97}$$

em que

$$\mathbf{R}_\mathbf{x} = \begin{bmatrix} r_x(0) & \cdots & r_x(M-1) \\ \vdots & \ddots & \vdots \\ r_x(M-1) & \cdots & r_x(0) \end{bmatrix} \quad \text{e} \quad \mathbf{r}_{x,f} = \begin{bmatrix} r_x(1) \\ \vdots \\ r_x(M) \end{bmatrix}$$

Filtro preditor ótimo - cont.

- Se $x(n)$ é branco $\Rightarrow r_x(i) = 0$
- $\mathbf{w}_f$ é ótimo quando $\mathbb{E}\left\{e_f^2(n)\right\}$ é mínimo

Ou seja, podemos fazer

$$\frac{\partial \mathbb{E}\left\{e_f^2(n)\right\}}{\partial \mathbf{w}_f} = 0 \quad \Rightarrow \mathbb{E}\left\{\frac{\partial e_f^2(n)}{\partial \mathbf{w}_f}\right\} = 0$$
$$\mathbb{E}\left\{2e_f(n)\frac{\partial e_f(n)}{\partial \mathbf{w}_f}\right\} = 0 \tag{98}$$

Filtro preditor ótimo - cont.

Lembrando que $e_f(n) = x(n) - \sum_{i=i}^{M} w_{f,i} \cdot x(n-i)$, temos então

$$\boxed{\mathbb{E}\left\{e_f(n)x(n-i)\right\} = 0 \quad \forall\, i = 1, \ldots, M} \tag{99}$$

Mas sabe-se ainda que $e_f(n-k) = x(n-k) - \mathbf{w}_f^T\mathbf{x}(n-k-1)$, e substituindo-se em (99) temos

$$\mathbb{E}\left\{e_f(n) \cdot \left[e_f(n-i) + \mathbf{w}_f^T\mathbf{x}(n-i-1)\right]\right\} = 0$$

$$\mathbb{E}\left\{e_f(n)e_f(n-i)\right\} + \mathbf{w}_f^T \cancelto{\mathbf{0}}{\mathbb{E}\left\{\mathbf{x}(n-i-1)\right\}} = 0 \tag{100}$$

$$\boxed{\mathbb{E}\left\{e_f(n)e_f(n-i)\right\} = 0 \quad \forall\, i = 1, \ldots, M}$$

Filtro de branqueamento!

Filtro preditor ótimo - cont.

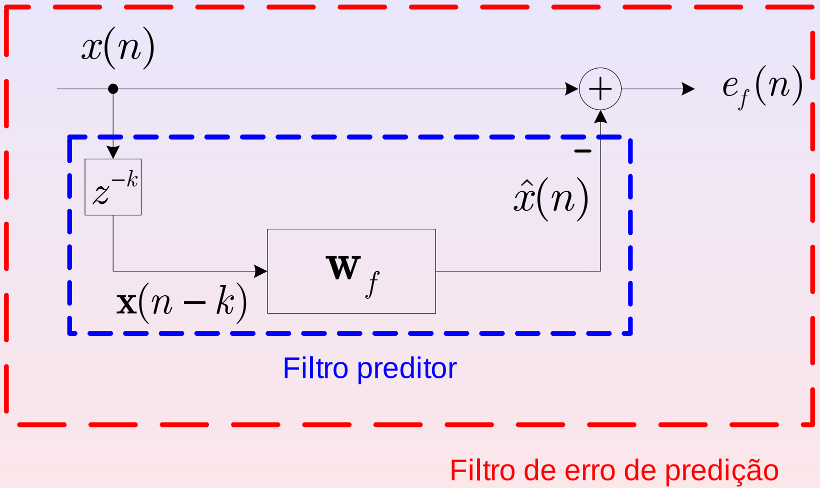
Conseqüência

Preditor de ordem infinita $\leftrightarrow$ Ruído branco no erro de saída

Meta do filtro

O preditor busca então fornecer um sinal $e_f(n)$ ortogonal entre si (branco): filtro de erro de predição é um "filtro branqueador"

$$M \to \infty \quad \Rightarrow \begin{cases} \mathbb{E}\{e_f(n)e_f(n-i)\} = 0 \quad \forall i \geq 1 \\ e(n) = 0 \\ e(n) = \text{ruído branco} \end{cases}$$


$x(n)$
$e_f(n)$
z^{-k}
$\hat{x}(n)$
$-$
$\mathbf{w}_f$
$\mathbf{x}(n-k)$
Filtro preditor
Filtro de erro de predição

Filtro preditor *backward*

De forma análoga ao caso *forward*, podemos derivar a equação do filtro de predição *backward*. Definindo

$$e_b(n) = x(n-M) - \mathbf{w}_b^T \mathbf{x}(n) \tag{101}$$

em que $\mathbf{x}(n) = \begin{bmatrix} x(n) & \cdots & x(n-M+1) \end{bmatrix}^T$

Teremos então

$$\begin{aligned} e_b^2(n) &= x^2(n-M) - 2\mathbf{w}_b^T \mathbf{x}(n) x(n-M) + \mathbf{w}_b^T \mathbf{x}(n)\mathbf{x}(n)^T \mathbf{w}_b \\ \Rightarrow \mathbb{E}\left\{e_b^2(n)\right\} &= r_x(0) - 2\mathbf{w}_b^T \mathbf{r}_{x,b} + \mathbf{w}_b^T \mathbf{R}_{\mathbf{x}} \mathbf{w}_b \end{aligned} \tag{102}$$

Filtro preditor *backward* - cont.

Assim, fazendo $\frac{\partial \mathbb{E}\left\{e_b^2(n)\right\}}{\partial \mathbf{w}_b}$, teremos

$$\boxed{\mathbf{w}_{b,\text{opt}} = \mathbf{R}_\mathbf{x}^{-1}\mathbf{r}_{x,b}} \tag{103}$$

em que

$$\mathbf{R}_\mathbf{x} = \begin{bmatrix} r_x(0) & \cdots & r_x(M-1) \\ \vdots & \ddots & \vdots \\ r_x(M-1) & \cdots & r_x(0) \end{bmatrix} \quad \text{e} \quad \mathbf{r}_{x,b} = \begin{bmatrix} r_x(M) \\ \vdots \\ r_x(1) \end{bmatrix}$$

Com isso, pode-se visualizar que o preditor ótimo *backward* tem a mesma estrutura do preditor forward, a menos da organização das correlações nos vetores $\mathbf{r}_{x,b}$ e $\mathbf{r}_{x,f}$.

Filtro preditor *backward* - cont.

Se escrevermos as equações dos filtros ótimos, *forward* e *backward*, na sua forma direta, ou seja

$$\mathbf{R_x} \mathbf{w}_f = \mathbf{r}_{x,f} \tag{104a}$$

$$\mathbf{R_x} \mathbf{w}_b = \mathbf{r}_{x,b} \tag{104b}$$

e notarmos que

$$\mathbf{r}_{x,f} = (\mathbf{r}_{x,b})^R \tag{105}$$

em que $(\cdot)^R$ é a operação de reversão no tempo, podemos deduzir que

$$\boxed{\mathbf{w}_f = (\mathbf{w}_b)^R} \tag{106}$$

Ou seja, é possível encontrar facilmente o preditor *backward* à partir de sua versão *forward* e vice-versa.

Resumo

- Cálculo do filtro ótimo linear depende apenas das estatísticas de segunda ordem do sinal
- Complexidade relativa ao cálculo das estatísticas e inversão da matriz
- Superfície de erro de forma quadrática
- Número de condicionamento da matriz determina comportamento das curvas de nível do erro médio quadrático
- Preditores direto e reverso ótimos têm a mesma estrutura da filtragem de Wiener
- Possibilidade de encontrar o preditor direto a partir do reverso e vice-versa