

# Algoritmos para Processamento de Áudio, Imagem e Vídeo

## Notas de aula sobre processamento de sinais aleatórios

Estas notas de aula correspondem a uma parte da matéria do curso que não está no livro “Discrete Fourier Analysis and Wavelets: Applications to Signal and Image Processing” de S. Allen Broughton e Kurt M. Bryan, e estão baseadas nos capítulos 26 a 31 do livro *Signal processing: a mathematical approach* de Charles Byrne. O autor tem uma versão pre-print deste livro em seu site, no endereço <http://faculty.uml.edu/cbyrne/master.pdf>.

### Preâmbulo e Advertência: Aleatoriedade e Ruído

Ruído pode significar...

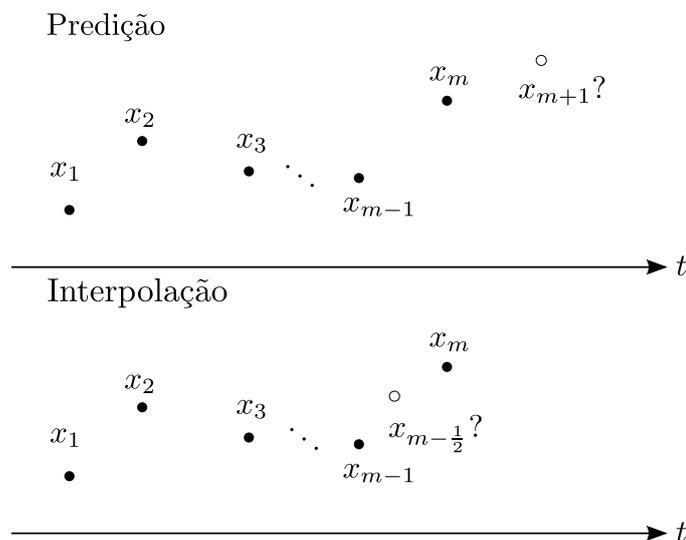
- ... algo que atrapalha a comunicação
- ... sons de máquinas, trovões, acidentes
- ... uma sonoridade complexa, sem alturas musicais definidas
- ... entre muitas outras coisas.

Aleatoriedade pode significar...

- ... que não compreendemos bem um processo (a ponto de prever seu resultado)
- ... que não conhecemos todas as variáveis relevantes do problema
- ... que um modelo completo de predição não é observável ou não seria computacionalmente tratável
- ... entre muitas outras coisas.

## 1 Predição e interpolação

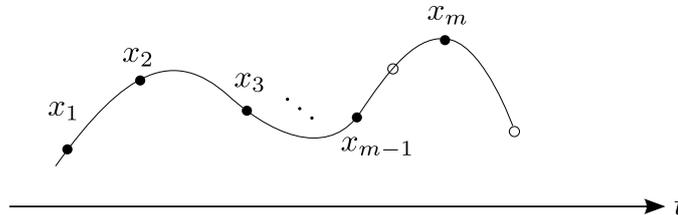
inferindo informação futura ou informação ausente



## 1.1 Modelos polinomiais

### Predição através de interpolação: construindo um modelo do sinal

$$f(t) = \sum_{k=0}^{m-1} a_k t^k$$



### Ajuste de modelos: estimando os coeficientes do polinômio

Método “difícil”: usar a base canônica (dos polinômios)

$$f(t) = \sum_{k=0}^{m-1} a_k t^k, \quad t = 1, \dots, m$$

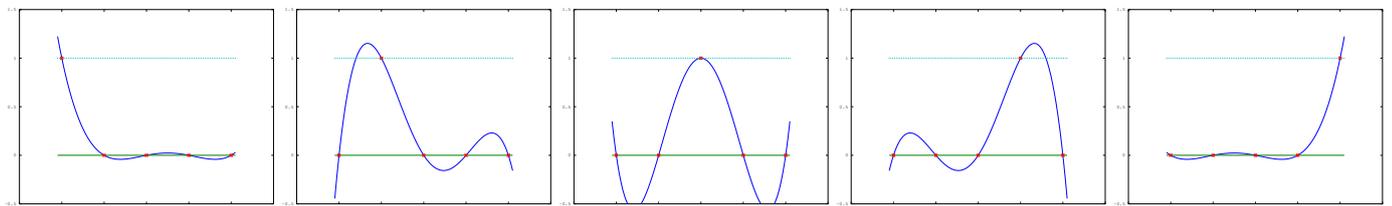
e resolver o sistema linear

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & 2^{m-1} \\ 1 & 3 & 9 & \cdots & 3^{m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & m & m^2 & \cdots & m^{m-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{m-1} \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{bmatrix}$$

### Ajuste de modelos: Método de Lagrange

Método fácil: método de Lagrange

$$f(t) = \sum_{k=1}^m x_k L_k(t) \quad \text{onde} \quad L_k(t) = \frac{\prod_{\substack{j=1, \dots, m \\ j \neq k}} (t - j)}{\prod_{\substack{j=1, \dots, m \\ j \neq k}} (k - j)}$$



Note que

$$L_k(t) = \frac{\prod_{\substack{j=1, \dots, m \\ j \neq k}} (t - j)}{\prod_{\substack{j=1, \dots, m \\ j \neq k}} (k - j)} = \begin{cases} 1, & \text{se } t = k \\ 0, & \text{se } t \in \{1, 2, \dots, m\}, t \neq k \end{cases}$$

Para a predição,  $\hat{x}_{m+1} = f(m+1) = \sum_{k=1}^m a_k x_k$ , com  $a_k = L_k(m+1)$ , ou ainda,

$$\hat{x}_n = \sum_{k=1}^m a_k x_{n-k}$$

com  $a_k = L_{m-k}(m+1)$  (filtro LIT/FIR, depende apenas de  $m$ ).

## Exemplo de Aplicação: Compactação de áudio e ADPCM

Estimadores:

1.  $\hat{x}_n = x_{n-1}$

2.  $\hat{x}_n - x_{n-1} = x_{n-1} - x_{n-2} \implies \hat{x}_n = 2x_{n-1} - x_{n-2}$

3.  $(\hat{x}_n - x_{n-1}) - (x_{n-1} - x_{n-2}) = (x_{n-1} - x_{n-2}) - (x_{n-2} - x_{n-3})$   
 $\implies \hat{x}_n = 3x_{n-1} - 3x_{n-2} + x_{n-3}$

4.  $(\hat{x}_n - x_{n-1}) - (x_{n-1} - x_{n-2}) - ((x_{n-1} - x_{n-2}) - (x_{n-2} - x_{n-3})) = (x_{n-1} - x_{n-2}) - (x_{n-2} - x_{n-3}) - ((x_{n-2} - x_{n-3}) - (x_{n-3} - x_{n-4}))$   
 $\implies \hat{x}_n = 4x_{n-1} - 6x_{n-2} + 4x_{n-3} - x_{n-4}$

ADPCM usa melhor estimador por bloco e guarda resíduo da predição.

## 2 Predição e otimização

### 2.1 Modelos lineares

#### Predição Linear: encontrando predições com erro mínimo

Ideia: procurar estimadores do sinal

$$x = (\dots, x_{n-1}, x_n, x_{n+1}, \dots)$$

da forma

$$\hat{x}_n = \sum_{k=1}^m a_k x_{n-k}$$

e que minimizem o erro

$$\|x - \hat{x}\|$$

em um horizonte finito  $n = K, \dots, L$ .

Ou seja, encontrar  $a = (a_1, a_2, a_3, \dots, a_m)^T \in \mathbb{R}^m$  que resolve

$$\min \|b - Ra\|$$

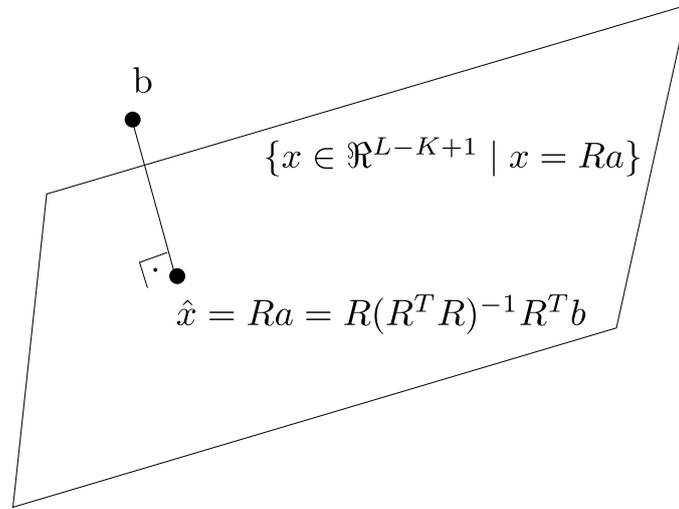
conhecendo-se  $b = (x_K, x_{K+1}, x_{K+2}, \dots, x_L)^T$  e

$$R = \begin{bmatrix} x_{K-1} & x_{K-2} & x_{K-3} & \cdots & x_{K-m} \\ x_K & x_{K-1} & x_{K-2} & \cdots & x_{K-m+1} \\ x_{K+1} & x_K & x_{K-1} & \cdots & x_{K-m+2} \\ x_{K+2} & x_{K+1} & x_K & \cdots & x_{K-m+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{L-1} & x_{L-2} & x_{L-3} & \cdots & x_{L-m} \end{bmatrix} \quad (\text{Toeplitz})$$

## Solução LPC (Linear Predictive Coding): método dos mínimos quadrados

$\min \|b - Ra\| \equiv \min f(a) = \frac{1}{2} \|b - Ra\|^2$ , cuja solução precisa satisfazer

$$\nabla f(\hat{a}) = 0 \iff -R^T(b - R\hat{a}) = 0 \iff \hat{a} = (R^T R)^{-1} R^T b$$



## Predição linear estocástica: a mesma coisa, só que diferente

O método de predição linear pode ser interpretado em termos de informação estatística do sinal. Na solução  $\hat{a} = (R^T R)^{-1} R^T b$  a matriz  $R^T R$  possui entradas

$$(R^T R)_{ij} = \sum_{l=0}^{L-K} x_{K-i+l} x_{K-j+l} \quad r_x(j-i)$$

que estão associadas à *auto-correlação* de  $x$  com atraso  $j-i$ , enquanto o vetor  $R^T b$  possui entradas

$$(R^T b)_j = \sum_{l=0}^{L-K} x_{K-j+l} x_{K+l} \quad r_x(j)$$

associadas à *auto-correlação* de  $x$  com atraso  $j$ .

Os coeficientes LPC podem ser obtidos alternativamente considerando a equação de filtro recursiva

$$y(n) = x(n) - \sum_{i=1}^M b_i y(n-i)$$

cuja função de transferência  $H(z) = \frac{1}{1 + \sum_{i=1}^M b_i z^{-i}}$  satisfaz o

Teorema de Wiener-Kinchin: 
$$\frac{1}{\left(1 + \sum_{i=1}^M b_i z^{-i}\right)} \approx \sum_{d=-M}^M r_x(d) z^d$$

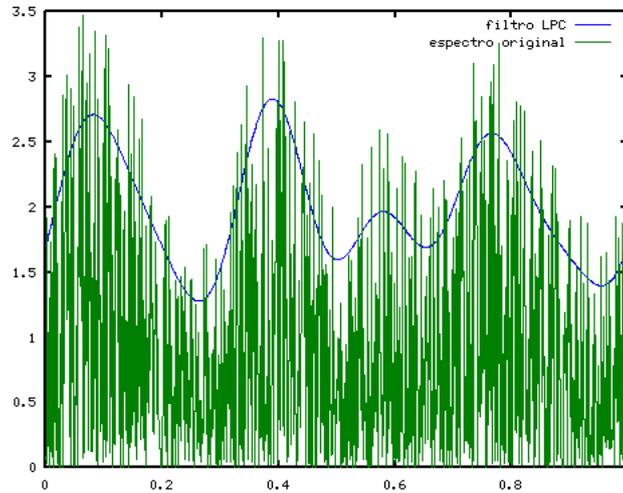
O teorema de Wiener-Kinchin fornece um sistema linear do tipo

$$\begin{bmatrix} r_x(0) & r_x(1) & r_x(2) & \cdots & r_x(M) \\ r_x(1) & r_x(0) & r_x(1) & \cdots & r_x(M-1) \\ r_x(2) & r_x(1) & r_x(0) & \cdots & r_x(M-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x(M) & r_x(M-1) & r_x(M-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(caso particular das *Equações de Wiener-Hopf*)

## Exemplo de Aplicação: Estimação de envoltória espectral e *morphing*

Coefficientes LPC definem um filtro recursivo (computacionalmente eficiente) com resposta em frequência próxima da do sinal original  $x$ :



Usar este filtro em um segundo sinal  $y$  permite a impressão de características espectrais de  $x$  em  $y$ . No caso de áudio está impressão empresta características timbrísticas de um som ao outro, sendo denominada *morphing* (em imagens esse termo se refere a outro tipo de processamento).

## 3 Processos aleatórios

### 3.1 Definições básicas

#### Processos aleatórios discretos: sequências de variáveis aleatórias

Considere a sequência  $(\dots, X_{-2}, X_{-1}, X_0, X_1, X_2, \dots)$  onde  $X_n$  é uma variável aleatória real com função de densidade de probabilidade  $f_n(x)$ , de tal modo que  $\text{Prob}[X_n \leq x] = \int_{-\infty}^x f_n(x) dx$ .

A *esperança* de  $X_n$  é definida como  $E(X_n) = \int x f_n(x) dx$ . Vamos supor deste ponto em diante que  $E(X_n) = 0, \forall n$ .

A *função de auto-correlação* é definida como

$$r_x(m, n) = E(X_m X_n),$$

e a *variância* de  $X_n$  é definida como  $\text{Var}(X_n) = E(X_n^2) - E(X_n)^2 = E(X_n^2) = r_x(n, n)$ .

O processo  $\{X_n\}$  é dito *estacionário em sentido amplo* quando valem as propriedades

1.  $E(X_n)$  independe de  $n$ ;
2.  $r_x(m, n)$  só depende de  $|m - n| \implies \text{Var}(X_n)$  independe de  $n$ ;

## Processos aleatórios discretos: matriz de autocorrelação de processos estacionários

Se  $\{X_n\}$  é estacionário em sentido amplo, denotamos por  $r_x(k) = E(X_{n+k}X_n)$ , e podemos construir a *matriz de auto-correlação* de ordem  $N \times N$  com as entradas

$$R_{mn} = r_x(m - n) = E(X_m X_n),$$

sendo que esta matriz pode ser representada de forma compacta como

$$R = E(XX^T),$$

onde  $X = (X_1, X_2, \dots, X_N)^T$ .

Evidentemente esta matriz é aleatória, e possui propriedades matemáticas relevantes, como por exemplo o fato de ser *simétrica* e *positiva semi-definida*, já que para qualquer  $y \in \mathbb{R}^N$  temos

$$y^T R y = y^T E(XX^T)y = E(y^T XX^T y) = E(\|X^T y\|^2) \geq 0.$$

## Processos aleatórios discretos: estimando a matriz de autocorrelação em processos realizados

Conhecendo a realização de uma parte do processo  $\{X_n\}$ , por exemplo,  $X_k = x_k, X_{k+1} = x_{k+1}, \dots, X_l = x_l$  (com  $l - k + 1 \gg N$ ) podemos estimar a matriz de covariância de ordem  $N$  como

$$\hat{R} = \frac{1}{l - k - N + 2} \sum_{i=0}^{l-k-N+1} Y_i Y_i^T,$$

onde

$$Y_i = \begin{bmatrix} x_{k+i} \\ x_{k+i+1} \\ \vdots \\ x_{k+i+N-1} \end{bmatrix}.$$

## 4 Predição estocástica

### 4.1 Modelos lineares

#### Predição estocástica: estimação linear $\hat{X}_n$ de $X_n$

Suponha que queiramos estimar  $X_n$  a partir de  $N$  termos anteriores através da expressão

$$\hat{X}_n = \sum_{k=1}^N a_k X_{n-k}.$$

Multiplicando os dois lados por  $X_{n-m}$  (para  $m = 1, \dots, N$ ) e tomando a esperança, teremos

$$E(\hat{X}_n X_{n-m}) = \sum_{k=1}^N a_k E(X_{n-k} X_{n-m}), \quad m = 1, \dots, N.$$

Supondo o processo estacionário em sentido amplo, teremos

$$E(\hat{X}_n X_{n-m}) = \sum_{k=1}^N a_k R_{mk}, \quad m = 1, \dots, N,$$

que corresponde a um sistema linear da forma  $Ra = b$ , onde  $R$  é a matriz de autocorrelação e  $b_m = E(\hat{X}_n X_{n-m})$ .

Vemos que o método LPC é aparentado ao método acima, porém usa estimativas finitas da matriz de covariância e sujeita-se à hipótese adicional  $E(\hat{X}_n X_{n-m}) = E(X_n X_{n-m})$ .

## 5 Filtragem estocástica

### 5.1 Hipóteses

#### Filtragem estocástica: separando sinal de ruído

Hipóteses:

1. o sinal  $x \in \mathbb{R}^N$  é desconhecido, porém possui média zero e matriz de covariância  $E(xx^T) = \sigma^2 I$  com  $\sigma$  conhecido;
2. o sinal passa por um processamento linear conhecido  $H \in \mathbb{R}^{M \times N}$ , e sofre adição de um ruído aleatório  $v \in \mathbb{R}^M$ , antes de ser observado como  $z = (Hx) + v = s + v \in \mathbb{R}^M$  ( $s \in \mathbb{R}^M$  é chamado de *componente do sinal*);
3.  $v$  possui média 0, matriz de covariância conhecida  $Q = E(vv^T)$ , e é independente de  $x$ ;
4. o estimador procurado para a componente do sinal  $s = Hx$  é linear nas observações  $z$ , ou seja,  $\hat{s} = B^T z$  para alguma matriz  $B \in \mathbb{R}^{M \times M}$  (a ser determinada).

#### Exemplo de Aplicação: Registro de sinais em ambientes ruidosos

Uma sala pode ser modelada como um sistema linear: dado um sinal  $x$  produzido por uma fonte sonora em uma sala, o que escutamos na realidade é  $Hx$ , onde  $H$  é a matriz circulante da resposta impulsiva de um modelo FIR.

Considerando  $v$  a soma dos ruídos (de ar-condicionado, lâmpadas, pássaros, aviões, etc), o sinal acústico registrado em um ponto da sala pode ser modelado como  $z = Hx + v$ , como no cenário acima.

O problema da filtragem estocástica consiste em obter o sinal “limpo” (por exemplo, o discurso de uma pessoa) a partir do registro “sujo” obtido por um microfone em uma sala ruidosa.

O mesmo problema poderia ser formulado a partir da recuperação de uma imagem projetada em uma sala, a partir de um registro em vídeo desta projeção, e sujeito à interferência de insetos, fumaça, etc.

### 5.2 Filtro de Wiener

#### Filtro de Wiener: o caso mais simples (impondo estrutura forte)

Suponha que a componente do sinal  $s$  é da forma  $s = au$ , onde  $a$  é um escalar desconhecido e  $u$  é um vetor *conhecido*.

Suponha ainda que o ruído também tem a forma  $v = bw$ , onde  $b$  é um escalar desconhecido e  $w$  é um vetor *conhecido*.

Se o número de observações for  $N \geq 2$ , seremos capazes de determinar  $a$  e  $b$  de forma unívoca a partir do conhecimento de  $z$ , já que  $z = s + v = au + bw$  e portanto  $a$  e  $b$  satisfazem

$$\begin{cases} z^T u &= a(u^T u) + b(w^T u) \\ z^T w &= a(u^T w) + b(w^T w) \end{cases}$$

O sistema acima terá solução única sempre que  $u$  e  $w$  forem linearmente independentes (do contrário sinal e ruído possuem precisamente a mesma estrutura e não faz sentido tentar separá-los).

### Um caso mais geral: impondo estrutura flexível

Considere que a componente do sinal possui a forma

$$s = \sum_{n=1}^K a_n u^n,$$

onde  $a_n$  são escalares desconhecidos e  $u^n$  são vetores conhecidos (por exemplo, componentes senoidais de frequências diversas).

Considere também que o ruído possui a forma

$$v = \sum_{m=1}^L b_m w^m,$$

onde  $b_m$  são escalares desconhecidos e  $w^m$  são vetores conhecidos.

Se  $N \geq K + L$  o problema teria solução única, mas na prática o mais comum é termos  $K, L \approx N$  (e.g. Fourier).

### Um caso mais geral: formulação matricial

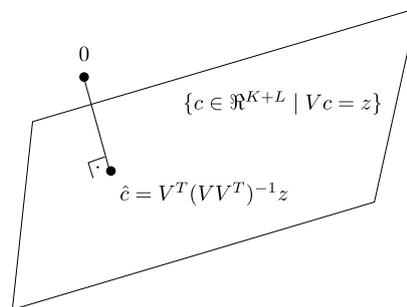
Sejam  $U \in \mathbb{R}^{N \times K}$  e  $W \in \mathbb{R}^{N \times L}$  as matrizes

$$U = \left[ \begin{array}{c|c|c|c} u^1 & u^2 & \dots & u^K \end{array} \right] \quad \text{e} \quad W = \left[ \begin{array}{c|c|c|c} w^1 & w^2 & \dots & w^L \end{array} \right]$$

Seja ainda  $V = [U \ W] \in \mathbb{R}^{N \times (K+L)}$  e  $c = (a_1, \dots, a_K, b_1, \dots, b_L)^T$ .

Queremos encontrar uma solução do sistema  $z = Ua + Wb = Vc$ , que possui muito mais variáveis do que restrições. Uma ideia é buscar uma solução de norma mínima:

$$\begin{cases} \min & \frac{1}{2} \|c\|^2 \\ \text{s.a} & Vc = z. \end{cases}$$



### Um caso mais geral: solução do sistema $z = Vc$

Este problema de otimização admite como solução o estimador

$$\hat{c} = V^T(VV^T)^{-1}z,$$

de onde obtemos o estimador da componente do sinal

$$\hat{s} = U\hat{c}_{[1,\dots,K]} = UU^T(VV^T)^{-1}z.$$

Observe que  $VV^T = (UU^T + WW^T)$  envolve a *matriz de correlação do sinal*  $UU^T$  e a *matriz de correlação do ruído*  $WW^T$ , que neste caso correspondem a estimadores a partir de espaços amostrais finitos (os conjuntos  $\{u^1, u^2, \dots, u^K\}$  e  $\{w^1, w^2, \dots, w^L\}$ ).

### A formulação estocástica: flexibilizando a estrutura de $s$ e $v$

Considere agora que os vetores de coeficientes

$$a = (a_1, a_2, \dots, a_K)^T \text{ e } b = (b_1, b_2, \dots, b_L)^T$$

são vetores aleatórios independentes, com distribuição uniforme e matrizes de covariância

$$E(aa^T) = I \text{ e } E(bb^T) = I,$$

de tal modo que  $E(ss^T) = E(Uaa^T U^T) = UU^T$  e  $E(vv^T) = E(Wbb^T W^T) = WW^T$ . Então

$$\begin{aligned} UU^T &= E(ss^T) \\ &= E(Hx(Hx)^T) \\ &= E(Hxx^T H^T) \\ &= HE(xx^T)H^T \\ &= \sigma^2 HH^T \end{aligned}$$

e

$$WW^T = E(vv^T) = Q.$$

### O filtro de Wiener vetorial: reaproveitando a solução do caso anterior

Lembrando que anteriormente obtivemos

$$\hat{s} = UU^T(UU^T + WW^T)^{-1}z$$

e adotando a informação do modelo estatístico

$$UU^T = \sigma^2 HH^T \text{ e } WW^T = Q,$$

temos a solução do filtro de Wiener estocástico, que é dado por

$$\begin{aligned} \hat{s} &= UU^T(UU^T + WW^T)^{-1}z \\ &= \sigma^2 HH^T(\sigma^2 HH^T + Q)^{-1}z \\ &= HH^T(HH^T + \sigma^{-2}Q)^{-1}z \end{aligned}$$

Lembrando ainda que  $B$  era o estimador tal que

$$\hat{s} = B^T z \text{ e } s = Hx,$$

segue que

$$B^T = HH^T(HH^T + \sigma^{-2}Q)^{-1}$$

e portanto

$$\boxed{\hat{x}_{\text{VWF}} = H^T(HH^T + \sigma^{-2}Q)^{-1}z}$$

## 5.3 Filtro de Kalman

### Filtro de Kalman: contexto de aplicação

Sistema dinâmico:

$$x^k = A^{k-1}x^{k-1} + m^{k-1}$$

onde

1.  $x^k \in \mathbb{R}^N$  é um vetor de estado do sistema no instante  $k$ ;
2.  $A^k \in \mathbb{R}^{N \times N}$  representa as transformações do sistema (em geral de acordo com algum modelo físico);
3.  $m^k$  representa o erro do modelo físico, que satisfaz  $E(m^k) = 0 \forall k$ , e a matriz de covariância  $E(m^k(m^k)^T) = M^k$  é conhecida;
4. no instante  $k$  obtemos a medida/observação  $z^k = H^k x^k + v^k$ , onde  $H^k$  é conhecida, e  $v^k$  possui média zero e variância conhecida  $E(v^k(v^k)^T) = Q^k$ .

### Filtro de Kalman: uso de estimadores a priori

Considere conhecido um estimador  $\hat{x}^{k-1}$  do sistema no instante  $k-1$ , e suponha que este estimador é não-enviesado, ou seja, que

$$E(\hat{x}^{k-1}) = E(x^{k-1}).$$

De acordo com o modelo físico podemos estimar a evolução do sistema como

$$y^k = A^{k-1}\hat{x}^{k-1}.$$

Este estimador  $y^k$  também é não-enviesado, já que

$$\begin{aligned} E(y^k - x^k) &= E(A^{k-1}\hat{x}^{k-1} - x^k) \\ &= A^{k-1}E(\hat{x}^{k-1}) - E(x^k) \\ &= A^{k-1}E(x^{k-1}) - E(x^k) \\ &= E(A^{k-1}x^{k-1} - x^k) \\ &= E(m^k) \\ &= 0 \end{aligned}$$

### Filtro de Kalman: estimadores a posteriori

Consideraremos estimadores do sinal  $\hat{x}^k$  lineares nas observações  $z^k$  e nos estimadores  $y^k$  obtidos a priori:

$$\hat{x}^k = (C^k)^T z^k + (D^k)^T y^k,$$

onde as matrizes  $C^k, D^k \in \mathbb{R}^{N \times N}$  devem ser estimadas.

A ideia é resolver o sistema de equações lineares (estocásticas)

$$\begin{cases} z^k = H^k x^k + v^k \\ y^k = x^k + w^k \end{cases}$$

usando a informação das matrizes de covariância dos resíduos  $v^k$  (ruído da medição) e  $w^k$  (erro da predição a priori).

## Filtro de Kalman: solução do sistema

Fazendo muitas (...) contas o estimador de Kalman é dado pela expressão

$$\hat{x}^k = y^k + G^k(z^k - H^k y^k), \quad (\text{preditor-corretor})$$

onde

- $G^k = R^k(H^k)^T (Q^k + H^k R^k (H^k)^T)^{-1}$ ,
- $Q^k$  é a matriz de correlação de  $w^k$ ,
- $R^k$  é a matriz de covariância de  $w^k = y^k - x^k$ , dada por  $R^k = M^{k-1} + A^{k-1} P^{k-1} (A^{k-1})^T$

A covariância de  $\hat{x}^k - x^k$  será dada por  $P^k = (I - G^k H^k) R^k$  e será usada no passo  $k + 1$  para determinar  $R^{k+1}$ .