# PTC-5720 Controle Estocástico

Oswaldo Luiz do Valle Costa

Aula 7 - 2020

*PTC-EPUSP*

## Filtragem

Em um problema de filtragem desejamos determinar o "melhor´´ estimador da variável de estado $x_k$ dada as observações até o presente instante $\{y_0, \ldots, y_k\}$. A motivação seria:

1. As variáveis $x_k$ representam quantidades físicas importantes, como pressão, temperatura, etc, e que necessitam ser calculadas com a maior precisão possível.
2. Uma classe natural de controle seria via realimentação de estado, nesse caso $u_k = u(k, x_k)$. Como $x_k$ não é conhecido, uma idéia seria trocá-lo pelo estimador $\widehat{x}_k$.

## Filtragem

Sejam $X$ e $Y$ variáveis aleatórias e para uma função $e(y)$ defina

$$H(e) = E((g(X) - e(Y))^2).$$

Deseja-se resolver o seguinte problema (problema de filtragem):

$$min_{e(.)} H(e) = min_{e(.)} E((g(X) - e(Y))^2)$$

Solução ótima $e^*$:

$$e^*(Y) = E(g(X)|Y).$$

## Filtragem

Defina o produto interno $< Z_1; Z_2 >= Cov(Z_1, Z_2)$. Mostre que

$$< g(X) - e^*(Y); e(Y) >= 0$$

(isto é, $g(X) - e^*(Y) \perp e(Y)$), e como isso mostre que $e^*(Y)$ é a solução ótima desejada.

## Filtragem

$$
\begin{aligned}
< g(X) - e^*(Y); e(Y) > &= E((g(X) - e^*(Y))e(Y)) \\
&= E(E((g(X) - e^*(Y))e(Y)|Y))
\end{aligned}
$$

Note que

$$
\begin{aligned}
E((g(X) - e^*(Y))e(Y)|Y) &= E((g(X) - e^*(Y))|Y)e(Y) \\
&= (E(g(X)|Y) - E(e^*(Y)|Y))e(Y) \\
&= (E(g(X)|Y) - e^*(Y))e(Y) \\
&= (E(g(X)|Y) - E(g(X)|Y))e(Y) = 0.
\end{aligned}
$$

## Filtragem

Logo

$$\begin{aligned}\|g(X)-e(Y)\|^2 &= < g(X)-e(Y); g(X)-e(Y) > \\ &= < g(X)-e^*(Y)+\tilde{e}(Y); g(X)-e^*(Y)+\tilde{e}(Y) >\end{aligned}$$

onde

$$\tilde{e}(Y) = e^*(Y) - e(Y)$$

## Filtragem

Portanto

$$\begin{aligned}\|g(X) - e(Y)\|^2 &= \|g(X) - e^*(Y)\|^2 + \|\tilde{e}(Y)\|^2 \\ &\quad + 2 < g(X) - e^*(Y); \tilde{e}(Y) > \\ &= \|g(X) - e^*(Y)\|^2 + \|\tilde{e}(Y)\|^2\end{aligned}$$

pela ortogonalidade $g(X) - e^*(Y) \perp \tilde{e}(Y)$, e como o primeiro termo não depende da função $e$, o mínimo é atingindo quando se faz $\tilde{e}(Y) = 0$, isto é

$$e(Y) = e^*(Y).$$

Problema: $E(g(X)|Y)$ é difícil em geral de ser calculado.

## Filtragem Linear

Considere o vetor $Y$ de observações

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

e a variável a ser estimada $Y_0$. Na filtragem linear deseja-se obter $\alpha_1, \ldots, \alpha_n$ e $\alpha_0$ de modo a minimizar

$$E((Y_0 - \widehat{Y}_0)^2)$$

onde

$$\widehat{Y}_0 = \alpha_1 Y_1 + \ldots + \alpha_n Y_n + \alpha_0 = \alpha' Y + \alpha_0$$

## Filtragem Linear

Fixando os valores $\alpha_1, \ldots, \alpha_n$ e chamando $U = Y_0 - (\alpha_1 Y_1 + \ldots + \alpha_n Y_n)$ segue que

$$\min_{\alpha_0} E((U - \alpha_0)^2)$$

é atingido para $\alpha_0^* = E(U)$, portanto

$$\begin{aligned} U - \alpha_0^* = Y_0 - (\alpha_1 Y_1 + \ldots + \alpha_n Y_n) - \\ (E(Y_0) - (\alpha_1 E(Y_1) + \ldots + \alpha_n E(Y_n)) \\ = Y_0^c - (\alpha_1 Y_1^c + \ldots + \alpha_n Y_n^c) \end{aligned}$$

onde $Y_i^c = Y_i - E(Y_i)$. Portanto podemos por simplicidade assumir que as variáveis tem média nula, e determinar $\alpha_1, \ldots, \alpha_n$ de modo a minimizar

$$E((Y_0 - (\alpha_1 Y_1 + \ldots + \alpha_n Y_n))^2) = E((Y_0 - \alpha' Y)^2).$$

## Filtragem Linear - Interpretação Geométrica

Vamos considerar o problema com $n = 1$. Portanto temos $(Y_0, Y_1)$ são variáveis aleatórias com média nula e conjuntamente distribuídas. Queremos achar $\alpha$ de modo a minimizar

$$\min_{\alpha} E((Y_0 - \alpha Y_1)^2).$$

Segue que

$$\begin{aligned} E((Y_0 - \alpha Y_1)^2) &= E(Y_0^2) - 2\alpha E(Y_0 Y_1) + \alpha^2 E(Y_1^2) \\ &= \sigma_0^2 - 2\alpha\rho\sigma_0\sigma_1 + \alpha^2\sigma_1^2 \end{aligned}$$

Derivando em $\alpha$ e igualando a zero obtemos

$$\alpha^* = \frac{E(Y_0 Y_1}{E(Y_1^2)} = \frac{\rho\sigma_0\sigma_1}{\sigma_1^2} = \frac{\rho\sigma_0}{\sigma_1}$$

# Filtragem Linear - Interpretação Geométrica

Temos que o melhor estimador linear é:

$$\widehat{Y}_0 = \frac{\rho\sigma_0}{\sigma_1} Y_1.$$

Note que o erro $\widetilde{Y}_0 = Y_0 - \widehat{Y}_0$ é ortogonal a $Y_1$. Realmente,

$$E((Y_0 - \widehat{Y}_0)Y_1) = E(Y_0 Y_1) - \frac{\rho\sigma_0}{\sigma_1} E(Y_1^2) = \rho\sigma_0\sigma_1 - \frac{\rho\sigma_0}{\sigma_1}\sigma_1^2$$
$$= \rho\sigma_0\sigma_1 - \rho\sigma_0\sigma_1 = 0$$

e o erro dado por

$$\widetilde{Y}_0 = Y_0 - \widehat{Y}_0 = Y_0 - \frac{\rho\sigma_0}{\sigma_1} Y_1 = \sigma_0\left(\frac{Y_0}{\sigma_0} - \rho\frac{Y_1}{\sigma_1}\right)$$

com variância dada por

$$E(\widetilde{Y}_0^2) = \sigma_0^2(1 - 2\rho^2 + \rho^2) = \sigma_0^2(1 - \rho^2).$$

## Filtragem Linear - Interpretação Geométrica

Deve-se destacar que o erro $\widehat{Y}_0$ é descorrelacionado com a variável $Y_1$. Também é fácil verificar que $\alpha^*$ é o único parâmetro com essa propriedade. Realmente,

$$E((Y_0 - \alpha Y_1)Y_1) = 0 \Leftrightarrow \alpha = \frac{E(Y_0 Y_1)}{E(Y_1^2)} = \frac{\rho\sigma_0\sigma_1}{\sigma_1^2} = \frac{\rho\sigma_0}{\sigma_1}$$

## Filtragem Linear - Interpretação Geométrica

A interpretação geométrica deste problema é a seguinte. Suponha que $v_0$ e $v_1$ são vetores com comprimento $\sigma_0$ e $\sigma_1$ respectivamente, com um ângulo $\theta$ entre eles tal que $cos(\theta) = \rho$. A projeção $\widehat{v}_0$ de $v_0$ em $v_1$ é tal que a diferença $\widetilde{v}_0 = v_0 - \widehat{v}_0$ é ortogonal a $v_1$. A projeção $\widehat{v}_0$ é dada por

$$\widehat{v}_0 = \frac{\|v_0\|}{\|v_1\|} cos(\theta) v_1 = \rho \frac{\sigma_0}{\sigma_1} v_1.$$

Note que o produto interno é dado por

$$< v_0; v_1 >= \|v_0\| \|v_1\| cos(\theta) = \sigma_0 \sigma_1 \rho = E(Y_0 Y_1) = cov(Y_0, Y_1)$$

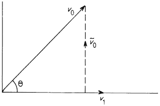


Fig. 3.1

given by (3.1.1) is the only one such that $(Y_0 - \alpha Y_1)$ and $Y_1$ are uncorrelated.

The geometric picture that goes along with this is as follows: Suppose $\mathbf{v}_0$, $\mathbf{v}_1$ are vectors in the plane which have lengths $\sigma_0$, $\sigma_1$ respectively and intersect at an angle $\theta$ such that $\cos\theta = \rho$ (see Fig. 3.1). The vector $\mathbf{v}_0$ can be expressed as the vector sum of its projection $\hat{\mathbf{v}}_0$ on to $\mathbf{v}_1$ and the difference $\tilde{\mathbf{v}}_0 = \mathbf{v}_0 - \hat{\mathbf{v}}_0$ which is orthogonal to $\mathbf{v}_1$. The projection $\hat{\mathbf{v}}_0$ is given by

$$\hat{\mathbf{v}}_0 = \sigma_0 \cos\theta\left(\frac{1}{\sigma_1}\mathbf{v}_1\right) = \rho\frac{\sigma_0}{\sigma_1}\mathbf{v}_1 \tag{3.1.4}$$

Comparing (3.1.2) and (3.1.4) we see that if the random variables $Y_0$, $Y_1$ are identified with the vectors $\mathbf{v}_0$, $\mathbf{v}_1$ respectively then the best linear estimate $\hat{Y}_0$ corresponds to the projection $\hat{\mathbf{v}}_0$ of $\mathbf{v}_0$ onto $\mathbf{v}_1$. The *inner* (or *dot*) *product* of the vectors $\mathbf{v}_0$ and $\mathbf{v}_1$ is

$$\mathbf{v}_0 \cdot \mathbf{v}_1 = \sigma_0\sigma_1\cos\theta = \sigma_0\sigma_1\rho = EY_0Y_1 = \operatorname{cov}(Y_0, Y_1).$$

Thus the vectors representing the random variables have lengths equal to the standard deviations of the random variables and inner product equal to the covariance. Notice in particular that if $\theta = 0$ or $\theta = \pi$ then the vectors are colinear and $\hat{\mathbf{v}}_0 = \pm\mathbf{v}_0 = \pm(\sigma_0/\sigma_1)\mathbf{v}_1$. Since $\rho = \cos\theta$ the equivalent condition on $\rho$ is that $\rho = \pm 1$. But we already saw in Chapter 1 that if $Y_0$, $Y_1$ have correlation coefficient $\pm 1$ then they are linearly related: $Y_0 = \pm(\sigma_0/\sigma_1)Y_1$. Thus 'linear estimation' can be done with zero error, as the geometric picture indicates.

In order to formalize the above discussion and generalize it to higher dimensions we need to review the geometrical properties of $\mathbb{R}^d$

## Filtragem Linear - Interpretação Geométrica

Note que em particular se $\theta = 0$ ou $\theta = \pi$ então os vetores $v_0$ e $v_1$ são co-lineares, e $\widehat{v}_0 = \pm v_0 = \pm\frac{\sigma_0}{\sigma_1}v_1$. Neste caso $\rho = \pm 1$ e como visto anteriormente, temos com probabilidade 1 que,

$$Y_0 = \pm\frac{\sigma_0}{\sigma_1}Y_1.$$

## Filtragem Linear - Exemplo

Considere o vetor $Y = \begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix}$ com média nula matrix de covariância dada por

$$Q = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 4 \end{pmatrix}$$

Qual é o melhor estimador linear de $Y_0$ com função de $Y_1$?
Resposta: $\sigma_0 = 1$, $\sigma_1 = 2$, $\rho\sigma_0\sigma_1 = 2\rho = 0.8 \Rightarrow \rho = 0.4$. Logo

$$\widehat{Y}_0 = \frac{\rho\sigma_0}{\sigma_1} Y_1 = 0.4\frac{1}{2} Y_1 = 0.2 Y_1.$$

Note que

$$\begin{aligned} E((Y_0 - \widehat{Y}_0)Y_1) &= E((Y_0 - 0.2Y_1)Y_1) = E(Y_0 Y_1) - 0.2E(Y_1^2) \\ &= 0.8 - 0.2 \times 4 = 0. \end{aligned}$$

## Filtragem Linear - Caso Geral

Sejam $X$ e $Y$ vetores aleatórios $p$ e $m$ dimensional, com média zero. Existe um único vetor aleatório $\widehat{X}$ (com equivalência com probabilidade 1) tal que

1. $\widehat{X} \in \mathcal{L}(Y)$
2. $X - \widehat{X} \perp \mathcal{L}(Y)$
3. $\widehat{X}$ é o mínimo estimador da média quadrática de $X$ dado $Y$, isto é, para qualquer $\beta$,

$$E((\beta'(X - \widehat{X})^2) = \min_{U \in \mathcal{L}(Y)} E((\beta' X - U)^2).$$

4. Se $cov(Y) > 0$ então $\widehat{X}$ é dado por:

$$\widehat{X} = E(XY')(cov(Y)^{-1})Y$$

## Filtragem Linear - Caso Geral

Seja $X = AY$ para alguma matrix $A$ com dimensão $p \times n$. Pela ortogonalidade, devemos ter que

$$E(\beta'(X - AY)\gamma' Y) = 0.$$

Logo, para quaisquer $\beta$, $\gamma$,

$$\beta'(E((X - AY)Y')\gamma = 0.$$

e portanto devemos ter

$$E((X - AY)Y') = 0 \Rightarrow AE(YY') = E(XY') \Rightarrow Acov(Y) = E(XY')$$
$$\Rightarrow A = E(XY')cov(Y)^{-1}$$

## Filtragem Linear - Média Não Nula

Nesse caso consideramos as varáveis centradas

$$X^c = X - E(X),$$

$$Y^c = Y - E(Y),$$

e temos que

$$\widehat{X} = E(X^c(Y^c)')cov(Y)^{-1}Y^c + E(X)$$

# Filtragem - Caso Normal

Suponha que $\begin{pmatrix} X \\ Y \end{pmatrix}$ seja conjuntamente normalmente distribuída. Então o melhor estimador afim de $X$ dado $Y$ coincide com $E(X|Y)$.
Considere o caso com média nula. Como $\widehat{X} = AY$ para alguma matriz $A$ então

$$\begin{pmatrix} X - \widehat{X} \\ Y \end{pmatrix} = \begin{pmatrix} I & -A \\ 0 & I \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

e $\begin{pmatrix} X - \widehat{X} \\ Y \end{pmatrix}$ é um vetor aleatório normal. Com $\widetilde{X} = X - \widehat{X}$ e $Y$ são descorrelacionados, segue que são independentes, e

$$\begin{aligned} E(X|Y) &= E(\widetilde{X} + \widehat{X}|Y) = E(\widetilde{X}|Y) + E(\widehat{X}|Y) \\ &= E(\widetilde{X}) + \widehat{X} = \widehat{X} \end{aligned}$$

## Filtragem - Caso Normal

Considere o caso com média não nula, com $\mu_X = E(X)$, $\mu_Y = E(Y)$. Considere a variável centrada $X^c = \mu_X$, $Y^c = Y - \mu_Y$.

$$\begin{aligned} E(X|Y) &= E(X^c + \mu_X|Y) = \mu_X + E(X^c|Y) \\ &= \mu_X + E(X^c|Y^c) = \mu_X + \widehat{X}^c = \widehat{X} \end{aligned}$$

1. $X|Y$ é normal com média $\widehat{X}$ e covariância $cov(X - \widehat{X})$, independente de $Y$.
2. isto porque $X = \widehat{X} + \widetilde{X}$ onde $\widehat{X}$ é função de $Y$ e $\widehat{X}$ é independente de $Y$. Logo

$$\begin{aligned} cov(X|Y) &= E(X - \widehat{X})(X - \widehat{X})'|Y) = E(\widetilde{X}\widetilde{X}'|Y) \\ &= E(\widetilde{X}\widetilde{X}') = cov(X - \widehat{X}). \end{aligned}$$

## Filtragem Linear

Caso Escalar:

$$A = E(XY')cov(Y)^{-1} = \rho \frac{\sigma_0 \sigma_1}{\sigma_1^2} = \rho \frac{\sigma_0}{\sigma_1}$$

Valor Numérico: $E(XY) = \rho\sigma_0\sigma_1 = 0,8$, $\sigma_1^2 = 4$, $\sigma_0 = 1$.

$$A = 0,8/4 = 0,2$$

## Exercício 1 - Filtragem Linear

Considere agora o vetor $\begin{pmatrix} X \\ Y_1 \\ Y_2 \end{pmatrix}$, com média nula e matriz de covariância dada por

$$cov(\begin{pmatrix} X \\ Y_1 \\ Y_2 \end{pmatrix}) = Q = \begin{pmatrix} 2 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ * & \frac{5}{2} & -\frac{1}{\sqrt{12}} \\ * & * & \frac{3}{2} \end{pmatrix}$$

a) Faça uma decomposição $Q = UU'$, onde $U$ é triangular superior, e obtenha o melhor estimador linear $\widehat{X}$ de $X$ dados $Y_1$, $Y_2$.

b) Refaça o item a) considerando a fórmula da projeção ortogonal.

c) Mostre que $X - \widehat{X} \perp Y_i$, $i = 1, 2$.

## Exercício 2 - Filtragem Linear

Seja $H$ uma matriz $r \times n$ e $z_k$, $k = 1, 2, \ldots$, uma sequência de ruído branco no sentido amplo, com média 0 e matriz de covariância $cov(z_k) = N > 0$. Suponha que

$$y_k = Hx + z_k$$

e também que $x$ e $z_k$ são descorrelacionadas para cada $k$. Portanto $y_k$ representa uma sequência de medidas de $x$ com erros de medida descorrelacionados $z_k$. Determine de forma "intuitiva" $\widehat{x}_k$, a projeção de $x$ no espaço $\mathcal{L}(y^k)$, onde

$$y^k = \begin{pmatrix} y_1 \\ \vdots \\ y_k \end{pmatrix}$$