

Processos Estocásticos

Paulo Lopes dos Santos
T-P Perdicoulis

14 de outubro de 2022

Conteúdo

1	Processos Estocásticos	1
1.1	Média e Covariância	3
1.2	Processos estocásticos estacionários	4
1.3	Processos ergódicos	8
1.4	Densidade espectral	9
1.5	Previsão linear em PE	12
1.6	Modelos de entrada-saída de processos estocásticos	16
1.7	Modelos de PE no espaço de estados	17
1.8	Espaços de vectores aleatórios	20
1.9	Projectões ortogonais em espaços de vectores aleatórios	21
1.10	Inovações	24

Resumo

Capítulo 1

Processos Estocásticos

No mundo real, existem muitas variáveis aleatórias (v.a.s) que evoluem no tempo segundo leis probabilísticas. Como exemplo dessas variáveis, podemos referir a evolução da temperatura num determinado local (ver Figura 1), o ruído térmico em circuitos

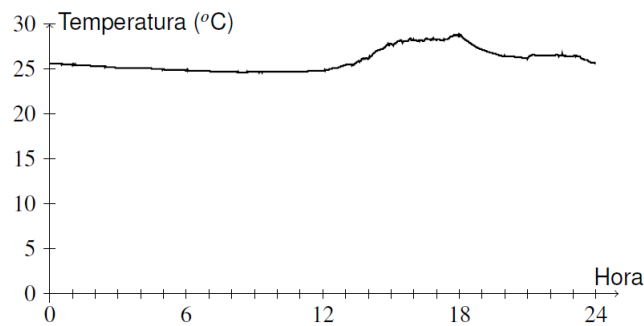


Figura 1.1: Evolução da temperatura num determinado local

eléctricos, o ruído de sensores (Figura 1, os sinais de radar, a cotação das acções na bolsa, o consumo de água numa cidade, a taxa de ocupação hoteleira de uma região, etc.. Chamamos *processo estocástico* (PE), ou série temporal, ao conjunto de todas as variações possíveis destas variáveis. Os PEs são, pois, famílias de funções (reais ou complexas) do tempo, ou seja, um PE é um conjunto de variáveis aleatórias (v.a), indexadas por um conjunto T . Sendo Ω o espaço amostral associado a essas v.a.s, podemos expressar um processo estocástico na forma

$$\{x(t, \omega), t \in T, \omega \in \Omega\}.$$

Se fixarmos $t = t_1$, teremos uma v.a. $x(t_1, \omega)$ no espaço amostral Ω . Se fixarmos o evento $\omega = \omega_1$ obteremos uma função do tempo $x(t, \omega_1)$ designada por *realização*.

Exemplo 1 Variável aleatória e realização dum PE:

Se $x(t, \omega)$ for a evolução da temperatura ao longo de um dia, então, se $t = 12h$, $x(12, \omega)$ é uma variável aleatória que descreve a temperatura ao meio-dia. Se $\omega = 15$

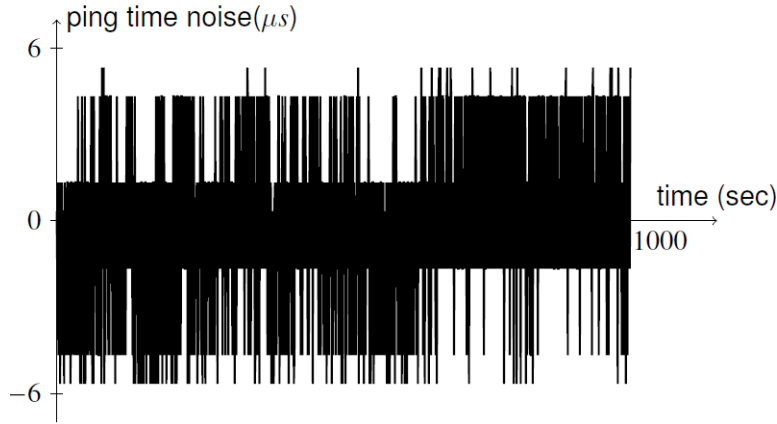


Figura 1.2: Ruído de um sensor ultrassônico

de outubro de 2021, $x(t, 15 \text{ de outubro de } 2021)$ é a função que mapeia a temperatura em função do tempo, isto é, é a realização de $x(t, \omega)$ no dia 15 de outubro de 2021. $\square$

Para simplificar a notação, representaremos o PE $x(t, \omega)$ apenas por $x(t)$ ou $\{x(t)\}$, deixando cair o evento na lista de argumentos como é usual na representação das variáveis aleatórias.

Consideremos o conjunto dos índices, T , como $\mathbb{R} =]-\infty, \infty[$ ou ainda um intervalo finito $]a, b[\in \mathbb{R}$. Dizemos, neste caso, que o processo estocástico é de parâmetro contínuo, ou simplesmente, *contínuo*. Por outro lado, dizemos que o PE é de parâmetro discreto (ou simplesmente, *discreto*) quando $T \equiv \mathbb{Z} = \{t = 0, \pm 1, \pm 2, \dots\}$ ou ainda $T \equiv \mathbb{Z} = \{t = t_1, t_1 + 1, \dots, t_1 + n\} \subset \mathbb{Z}$.

Exemplo 2 - PE de parâmetro contínuo:

Consideremos, de novo, a evolução da temperatura ao longo de um dia, já com a nova notação $x(t)$. Como varia continuamente no intervalo $t \in [0, 24[$, $x(t)$ é um PE de parâmetro contínuo. $\square$

Exemplo 3 - PE de parâmetro discreto:

Seja $x^*(t)$ um PE que resulta da amostragem de $x(t)$ de hora a hora, isto é, $t \in \{0, 1, \dots, 23\}$. Então $x^*(t)$ é PE de parâmetro discreto. $\square$

Tomemos os instantes $t_1, t_2, \dots, t_k$. Nestes instantes, o PE pode ser visto como

uma v.a. k -dimensional $X_{t_1:t_k} = \begin{pmatrix} x(t_1) \\ x(t_2) \\ \vdots \\ x(t_k) \end{pmatrix}$ cuja função de distribuição é definida por

$$P\{x(t_1) \leq a_1, x(t_2) \leq a_2, \dots, x(t_k) \leq a_k, \}$$

$$\int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \int_{-\infty}^{a_k} p_{t_1, t_2, \dots, t_k}(x_1, \dots, x_k) dx_1 dx_2, \dots, dx_k$$

onde $a_1, a_2, \dots, a_k \in \mathbb{R}$ e $p_{t_1, t_2, \dots, t_k}(x_1, \dots, x_k)$ é a *função de densidade de probabilidade conjunta* de $X_{t_1:t_k}$. Esta distribuição conjunta diz-se uma distribuição finita do PE nos instantes $t_1, t_2, \dots, t_k$. Tem-se, então, que a distribuição de um PE pode ser determinada por todas as suas distribuições finitas. Em particular, se todas as distribuições de x forem gaussianas, o PE diz-se *gaussiano*.

1.1 Média e Covariância

Acabamos de ver que os PE são conjuntos de v.a.. Dado estas últimas serem frequentemente caracterizadas pelos seus momentos de primeira e segunda ordem, é natural que o mesmo aconteça com os PE. Definimos então:

$$\begin{aligned} M(t_1, t_2, \dots, t_k) &= E\{x(t_1), x(t_2), \dots, x(t_k)\} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 \cdots x_k p_{t_1, t_2, \dots, t_k}(x_1, \dots, x_k) dx_1 dx_2 \cdots dx_k \end{aligned} \quad (1.1)$$

como *momento de ordem k* do PE $\{x(t)\}$, onde $E\{\cdot\}$ é a esperança matemática. No que se segue, como aliás referimos atrás, estamos sobretudo interessados nos momentos de primeira e segunda ordem, que são respectivamente designados por *função média* e *função (auto-)covariância*. A média é representada por:

$$\mu_x(t) = E\{x(t)\} = M_x(t) \quad (1.2)$$

e a covariância por:

$$A_{xx}(t, s) = E\{(x(t) - \mu_x(t))(x(s) - \mu_x(s))\} = M_x(t, s) - M_x(t)M_x(s). \quad (1.3)$$

Outra representação para a covariância é $cov\{x(t), x(s)\}$. Em particular:

$$\sigma_x^2(t) = cov\{x(t), x(t)\} = A_{xx}(t, t) \quad (1.4)$$

que designamos por *variância* de $\{x(t)\}$.

PE com variância finita dizem-se PE de segunda ordem.

Exemplo 4 - Função de densidade de probabilidade dum PE gaussiano:

Seja $\{x(t), t = 0, \pm 1, \pm 2, \dots\}$ um PE gaussiano. Sejam $\mu_x(t)$ e $\sigma(t, s)$ a média e a covariância, respectivamente. Então a função de densidade de probabilidade conjunta de $x(t_1), \dots, x(t_k)$ é expressa como :

$$p_{t_1, t_2, \dots, t_k}(x_1, \dots, x_k) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{i,j=1}^k \Sigma_{i,j}^{-1} (x_i - \mu_x(t_i)) (x_j - \mu_x(t_j)) \right\}$$

onde $\Sigma = (\sigma(t_i, t_j)) \in \mathbb{R}^{k \times k}$ é a matriz de covariâncias, e $\Sigma_{i,j}^{-1}$ designa o elemento (i, j) da matriz inversa Σ^{-1} . Se $x(t)$ for ruído branco temos então que Σ é uma matriz diagonal. $\square$

Consideremos agora $\{x(t), t = 0, \pm 1, \pm 2, \dots\}$ um vector de PE de dimensão n , i.e.,

$$x(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix} \in \mathbb{R}^n$$

onde $x_i(t), i = 1, \dots, n$ são PE escalares. Posto isto, definimos então a média vectorial (ou média de $x(t)$) como:

$$\mu_x(t) = Ex(t) = \begin{pmatrix} Ex_1(t) \\ \vdots \\ Ex_n(t) \end{pmatrix} = \begin{pmatrix} \mu_{x_1}(t) \\ \vdots \\ \mu_{x_n}(t) \end{pmatrix} \quad (1.5)$$

bem como a covariância vectorial (ou covariância de $x(t)$):

$$\begin{aligned} \Lambda_{xx}(t, s) &= E \left\{ (x(t) - \mu_x(t)) (x(s) - \mu_x(s))^T \right\} = \\ &= \begin{pmatrix} E\{\tilde{x}_1(t)\tilde{x}_1(s)\} & E\{\tilde{x}_1(t)\tilde{x}_2(s)\} & \cdots & E\{\tilde{x}_1(t)\tilde{x}_n(s)\} \\ E\{\tilde{x}_2(t)\tilde{x}_1(s)\} & E\{\tilde{x}_2(t)\tilde{x}_2(s)\} & \cdots & E\{\tilde{x}_2(t)\tilde{x}_n(s)\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{\tilde{x}_n(t)\tilde{x}_1(s)\} & E\{\tilde{x}_n(t)\tilde{x}_2(s)\} & \cdots & E\{\tilde{x}_n(t)\tilde{x}_n(s)\} \end{pmatrix} \end{aligned} \quad (1.6)$$

sendo $\tilde{x}(t) = x(t) - \mu_x(t)$. Os elementos da diagonal de $\Lambda_{xx}(t, s)$ são as funções de covariância de $\{x_i(t), t = 0, \pm 1, \pm 2, \dots\}$, sendo os restantes as funções de covariância cruzada entre $\{x_i(t)\}$ e $\{x_j(t)\}, i \neq j$.

1.2 Processos estocásticos estacionários

Um PE diz-se *estacionário* se não varia com o tempo. Isto significa que a função de densidade de probabilidade conjunta só depende da diferença temporal entre as amostras, sendo independente do intervalo de tempo entre os instantes em que são consideradas. Estatisticamente, isto equivale a afirmar que o futuro é igual ao passado, e pode ser expresso através da função de densidade de probabilidade conjunta:

$$p_{t_1, t_2, \dots, t_k}(x_1, x_2, \dots, x_k) = p_{t_1 + \tau, t_2 + \tau, \dots, t_k + \tau}(x_1, x_2, \dots, x_k), \quad \tau = 0, \pm 1, \pm 2, \dots \quad (1.7)$$

Se esta condição se verificar para qualquer valor de τ , dizemos ser o PE *fortemente estacionário*. Num PE fortemente estacionário, com momento de segunda ordem, a média é constante e a covariância depende da diferença entre 2 instantes, i.e.,

$$\mu_x(t) = E\{x(t)\} = \mu_x(0) = \mu_x \quad (1.8)$$

$$\Lambda_{xx}(t, s) = \Lambda_{xx}(t - s, 0) = \Lambda_{xx}(t - s) \quad (1.9)$$

Exemplo 5 - Ruído branco:

Um PE $\{v(t), t = 0, 1, \dots\}$ diz-se *ruído branco* se as v.a. $v(t)$ e $v(s)$ forem igualmente distribuídas e independentes entre si para qualquer $t \neq s$, i.e.,

$$p_{t,s}(v(t), v(s)) = p_t(v(t)) p_s(v(s)), \quad t \neq s$$

Como $v(t)$ e $v(s)$ são igualmente distribuídas, então $\mu_v(t) = E\{v(s)\} = \mu_v$,

$$M_v(t, t) = E\{v(t)^2\} = \mu_v^2 + \sigma_v^2$$

$$M_v(t, s) = E\{v(t)v(s)\} = E\{v(t)\}E\{v(s)\} = \mu_v^2$$

Temos ainda que

$$\begin{aligned} \Lambda_{xx}(t, s) &= E\{(v(t) - \mu_v)(v(s) - \mu_v)\} \\ &= E\{v(t)v(s) - \mu_v v(s) - \mu_v v(t) + \mu_v^2\} \\ &= E\{v(t)v(s)\} - \mu_v E\{v(t)\} - \mu_v E\{v(s)\} + \mu_v \mu_v \\ &= E\{v(t)v(s)\} - \mu_v^2 \\ &= \begin{cases} 0 & , \quad t \neq s \\ E\{v(t)^2\} - \mu_v^2 & , \quad t = s \end{cases} \end{aligned}$$

Sendo só dependente de $t - s$, Logo, o ruído branco é um PE estacionário.

O ruído branco é um sinal imprevisível, pois, o seu valor, num dado instante, é independente de todo o passado e futuro. A aleatoriedade do ruído branco é ilustrada nas figuras 1.3 e 1.4 que mostram realizações de ruído branco gaussiano e ruído branco binário, respetivamente □

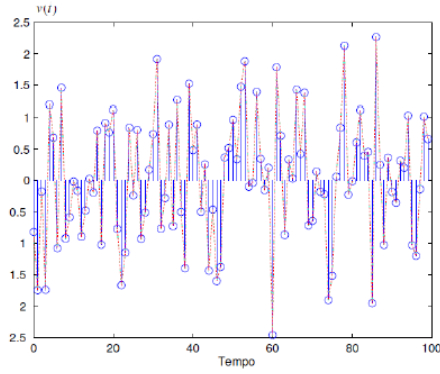


Figura 1.3: Realização de ruído branco gaussiano de média nula.

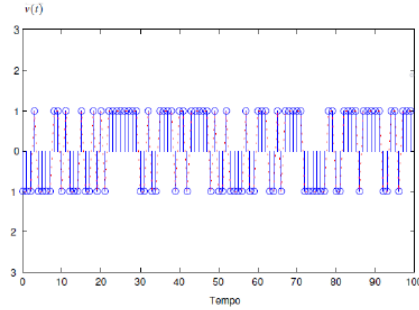


Figura 1.4: Realização de ruído branco binário de média nula.

No entanto, existem PEs com distribuições diferentes em instantes de tempo diferentes mas com momentos de primeira e segunda ordem estacionários¹. Dizemos que estes PEs são *fracamente estacionários*. É óbvio que os PEs fortemente estacionários também são fracamente estacionários, não sendo no entanto o recíproco verdadeiro.

Exemplo 6 - Caminho aleatório:

O ruído branco é usado, de forma adequada, para gerar diferentes processos com diferentes propriedades estocásticas, como por exemplo, o caminho aleatório $x(t)$, que é expresso como uma soma de ruídos brancos:

$$x(t) = v(1) + v(2) + \dots + v(t), \quad t = 1, 2, \dots \quad (1.10)$$

com $x(0) = 0$. Podemos ainda escrever (1.10) recursivamente como:

$$x(t) = x(t-1) + v(t), \quad \text{com } x(0) = 0 \quad (1.11)$$

Calculemos a sua média e variância.

Considerando $\{v(t)\}$ um PE gaussiano de média zero e variância σ_v^2 , temos que $E\{v(t)\} = 0$ e $E\{v(t)v(s)\} = \delta(t-s)\sigma_v^2$, em que $\delta(t)$ é o impulso de Dirac. Podemos então calcular a média do passeio aleatório como:

$$\mu_x(t) = E\{v(1) + v(2) + \dots + v(t)\} = 0.$$

No que diz respeito à covariância, considerando $t > s$, $x(t) - x(s) = \sum_{i=s+1}^t v(i)$ e

$x(s) = \sum_{i=1}^s v(i)$ são independentes, temos então:

$$\begin{aligned} \Lambda_{xx}(t, s) &= E\{x(t)x(s)\} = E\{[x(t) - x(s)]x(s) + x(s)^2\} = E\{x(s)^2\} = \\ &= \sum_{i=1}^s E\{v(i)^2\} = s\sigma_v^2 \end{aligned}$$

De igual forma, quando $t < s$, temos que $\Lambda_{xx}(t, s) = t\sigma_v^2$. Donde se conclui que a covariância de um caminho aleatório é $\Lambda_{xx}(t, s) = \min(t, s)\sigma_v^2$, que não é de forma alguma uma função de $t - s$; donde se conclui não ser o caminho aleatório um PE estacionário.

A Figura 1.5 mostra uma realização dum passeio aleatório. □

Sendo $\{x(t)\}$ um processo estacionário de segunda ordem, sua média é constante, igual a μ_x . Temos sem perda de generalidade que $x(t) - \mu_x$, é um PE de média nula pelo que iremos sempre considerar, daqui para a frente, que $E\{x(t)\} = 0$. Por outro lado, sendo a sua função de covariância unicamente dependente de $t - s$, iremos expressá-la na forma

$$\Lambda_{xx}(\tau) = E\{x(t+\tau)x(t)\}, \tau = 0, \pm 1, \pm 2, \dots \quad (1.12)$$

Lema 1. *A função covariância de um processo estacionário $\{x(t)\}$ tem as seguintes propriedades:*

¹Um momento de primeira ordem é estacionário se não variar com o tempo. Um momento de segunda ordem de segunda ordem é estacionário se só depender do espaçamento entre os instantes de tempo

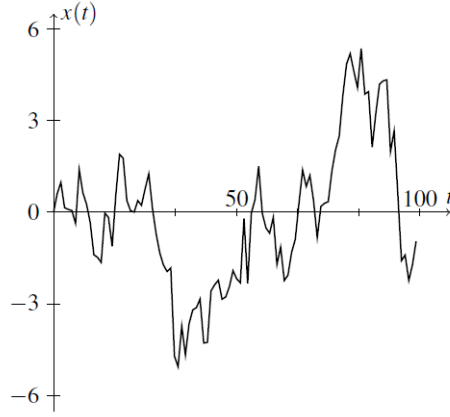


Figura 1.5: Realização dum passeio aleatório.

- (i) (Limite) $|A_{xx}(\tau)| \leq A_{xx}(0) = \sigma_x^2, \quad \tau = 0, \pm 1, \pm 2, \dots$
- (ii) (Simetria) $A_{xx}(\tau) = A_{xx}(-\tau)$.
- (iii) (Não negatividade) $\sum_{i,k=1}^n a_i a_k A_{xx}(\tau_i - \tau_k) \geq 0$

Estas propriedades permitem-nos testar se um processo é estacionário. Caso não se verifiquem temos a certeza de estar perante um processo não estacionário.

Exemplo 7 - Funções de covariância válidas e inválidas:

Das três sequências discretas representadas na Figura 1.6 só a de cima é que obedece às três propriedades do Lema 1 e, consequentemente, só esta é que pode ser a função de covariância dum processo estocástico estacionário. $\square$

Consideremos o PE conjunto $\{x(t), y(t), t = 0, \pm 1, \pm 2, \dots\}$ de média nula.

Os processos $\{x(t)\}$ e $\{y(t)\}$ dizem-se *juntamente estacionários* se, considerando o vector PE $w(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$, este for um PE estacionário. Dada a matriz de covariância de w :

$$\begin{aligned} A_{ww}(t + \tau, t) &= \begin{pmatrix} A_{xx}(t + \tau, t) & A_{xy}(t + \tau, t) \\ A_{yx}(t + \tau, t) & A_{yy}(t + \tau, t) \end{pmatrix} = E \left\{ \begin{pmatrix} x(t + \tau) \\ y(t + \tau) \end{pmatrix} (x^T(t) y^T(t)) \right\} \\ &= \text{cov} \{w(t + \tau)w(t)\}, \end{aligned} \quad (1.13)$$

a estacionaridade implica que as covariâncias de (1.13) sejam somente função da diferença τ .

A $A_{xy}(\tau) = E \{x(t + \tau)y(t)\}$ chama-se *covariância cruzada* entre x e y .

Lema 2. A covariância cruzada, $A_{xy}(\tau)$, tem as seguintes propriedades:

- (i) (Anti-simetria) $A_{xy}(\tau) = A_{yx}(-\tau)$
- (ii) (Limite) $|A_{xy}(\tau)| \leq A_{xx}(0)A_{yy}(0), \quad \tau = \pm 1, \pm 2, \dots$

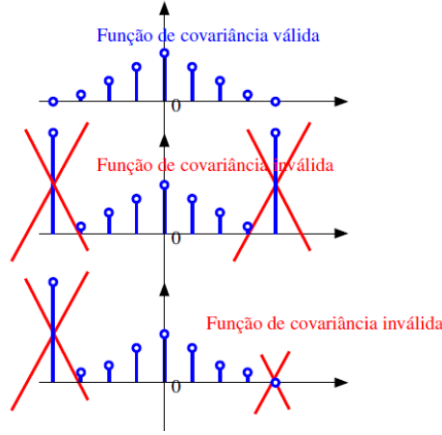


Figura 1.6: Sequências candidatas a função de covariância

1.3 Processos ergódicos

Um dos grandes desafios da análise de PEs é estimar os seus parâmetros estocásticos através da observação dos dados ao longo do tempo. Dado os parâmetros estarem relacionados com os valores esperados de alguma função de um PE, estudamos então o problema da estimação da média e covariância de um PE estacionário.

Suponhamos $\{x(t), t = 0, \pm 1, \pm 2, \dots\}$ um PE estacionário de segunda ordem. Definimos a média temporal deste PE como

$$m_x(N) = \frac{1}{2N+1} \sum_{t=-N}^N x(t), \quad (1.14)$$

e correlação temporal como

$$r_{xx}(\tau, N) = \frac{1}{2N+1} \sum_{t=-N}^N x(t+\tau)x(t), \quad \tau = 0, \pm 1, \pm 2, \dots \quad (1.15)$$

Se $\{x(t), t = 0, \pm 1, \pm 2, \dots\}$ tiver média nula, esta quantidade é também designada *função de (auto-)covariância*. As funções média (1.8) e covariância (1.9) são definidas como médias de um conjunto, i.e, como valores esperados. Mas (1.15) e (1.14) são definidas como médias temporais de uma realização.

$$x = (\dots, x(-1), x(0), x(1), \dots)$$

Aquando da análise de dados, nós lidamos antes com uma realização, gerada por uma experiência, e não com um conjunto de dados. Posto isto, sob o ponto de vista prático, a definição de momentos em termos de média temporal é preferível à média definida por um conjunto de dados. Falta agora saber se a média temporal, $r_{xx}(\tau)$, é ou não igual à média do conjunto $A_{xx}(\tau)$.

Todo o PE cujos parâmetros estatísticos são estimados de forma precisa a partir de uma simples realização, diz-se *processo ergódico*. Ou seja, num processo ergódico

as médias temporais convergem para as esperanças matemáticas e são utilizadas como estimativas destas. No que se segue enunciamos teoremas ergódicos para a esperança e a covariância.

Definição 1. Diz-se que um processo estacionário tem média ergódica se

$$\lim_{N \rightarrow \infty} E \left\{ [m_x(N) - \mu_x]^2 \right\} = 0,$$

em que

$$\begin{aligned} \mu_x &= E \{x(t)\} \\ m_x(N) &= \frac{1}{2N+1} \sum_{t=-N}^N x(t), \end{aligned}$$

ou seja se $m_x(N)$ convergir em média quadrática para μ_x .

Definição 2. Um PE ergódico, $x(t)$, de média nula tem covariância ergódica se

$$E \{r_{xx}(\tau) - \Lambda_{xx}(\tau)\} = 0$$

em que

$$\begin{aligned} \lambda_{xx}(\tau) &= E \{x(t)x(t-\tau)\} \\ r_{xx}(\tau) &= \frac{1}{2N+1} \sum_{t=-N}^N E \{x(t+\tau)x(t)\}. \end{aligned}$$

Maior detalhe sobre as condições de ergodicidade da média e da covariância podem ser vistas em T. Katayama. A partir deste momento só serão considerados PE com média e covariância ergódicas.

1.4 Densidade espectral

Se a função de covariância dum processo estacionário $x(t)$ for absolutamente somável, isto é, se

$$\sum_{\tau=-\infty}^{\infty} |\Lambda_{xx}(\tau)| < \infty,$$

então terá uma transformada de Fourier. Essa transformada é designada como *densidade espectral* de $x(t)$.

Definição 3. Se $x(t)$ for um PE estacionário com função de covariância absolutamente somável, então

$$\Phi_{xx}(\omega) = \sum_{\tau=-\infty}^{\infty} \Lambda_{xx}(\tau) e^{-j\omega\tau}$$

é a sua densidade espectral.

Como $\Phi_{xx}(\omega)$ é a transformada de Fourier de $\Lambda_{xx}(\tau)$, $\Lambda_{xx}(\tau)$ pode ser determinado através da transformada inversa de Fourier:

$$\Lambda_{xx}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{xx}(\omega) e^{j\omega\tau} d\omega$$

Vemos, assim, que existe uma relação de um para um entre $\Lambda_{xx}(\tau)$ e $\Phi_{xx}(\omega)$, pelo que ambas as funções transportam a mesma informação. Então, para $\tau = 0$, temos:

$$\Lambda_{xx}(0) = \sigma_x^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{xx}(\omega) d\omega.$$

Considerando que $x(t)$ é ergódico podemos escrever

$$\Lambda_{xx}(0) = r_{xx}(0) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{t=-N}^N x(t)^2 = P_x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{xx}(\omega) d\omega,$$

em que P_x significa potência de $x(t)$. Vemos, deste modo, que Φ_{xx} representa a distribuição da potência de $x(t)$ pelo espectro $\omega \in]-\pi, \pi]$ e daí o nome de densidade espectral.

Lema 3. A densidade espectral $\Phi_{xx}(\omega)$ tem as seguintes propriedades:

- (i) $\Phi_{xx}(\omega) = \Phi_{xx}(-\omega)$ (simetria par)
- (ii) $\Phi_{xx}(\omega) \geq 0$ (não negativa)

Demonstração. A propriedade (i) resulta do facto de $\lambda_{xx}(\tau)$ ser uma sequência par e (ii) é consequência de que a variância de $x(t)$, $\sigma_x(t) = \int_{-\pi}^{\pi} \Phi_{xx}(\omega) d\omega$ ser não negativa e do Lema 4. Se Φ_{xx} fosse negativo numa banda de frequências $\omega_1 < \omega < \omega_2$, seria possível obter, através duma filtragem adequada², um sinal com variância negativa, o que é uma impossibilidade. $\square$

Exemplo 8 - Densidade espectral do ruído branco

Seja $e(t)$ ruído branco. Então, a sua sequência de covariância é

$$\lambda_{ee}(\tau) = \begin{cases} \sigma_e^2, & \tau = 0 \\ 0, & \tau \neq 0, \end{cases}$$

sendo a sua densidade espectral

$$\Phi_{ee}(\omega) = \sum_{\tau=-\infty}^{\infty} \lambda_{ee}(\tau) e^{-j\omega\tau} = \lambda_{ee}(0) = \sigma_e^2. \quad (1.16)$$

Logo, o ruído branco tem uma densidade espectral constante, podendo-se afirmar que, tal como a luz branca, tem um espectro plano (ver Figura 1.4). Daí o nome de ruído branco. $\square$

É muito frequente os PE propagarem-se através de sistemas dinâmicos. Veremos, em seguida, o que acontece a um processo quando se propaga através de um sistema linear invariante no tempo e estável.

Lema 4. Se a entrada de um sistema linear e invariante no tempo com função de transferência $G(z)$ estável for um PE estacionário de 2^a ordem e média nula $\{u(t)\}$, a saída $\{y(t)\}$ também é um PE estacionário de 2^a ordem e média nula com densidade espectral

$$\Phi_{yy}(\omega) = |G(e^{j\omega})|^2 \Phi_{uu}(\omega).$$

²Por exemplo, um filtro passa-banda ideal com banda de passagem $\omega_1 < \omega < \omega_2$.

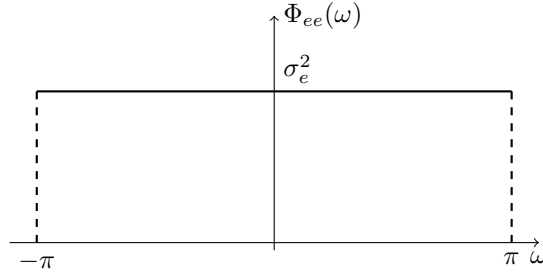


Figura 1.7: Densidade espectral do ruído branco.

Demonstração. Ver T. Katayama. $\square$

É comum modelar um PE estacionário como a saída dum sistema causal linear e invariante no tempo, estável e de fase mínima, excitado por ruído branco. Este procedimento é fundamentado pelo seguinte teorema:

Teorema 1 (Factorização da densidade espectral). *A função de densidade espectral $\Phi_{yy}(\omega)$ pode ser factorizada na forma*

$$\Phi_{yy}(\omega) = H(e^{j\omega}) H(e^{-j\omega}) \sigma^2$$

em que

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k}$$

é uma função de transferência estável com $h(0) = 1$ e com todos os zeros no interior do círculo unitário.

Demonstração. Ver T. Katayama. $\square$

Este teorema e o Lema 4 permitem modelar qualquer PE como a saída de um sistema excitado por ruído branco. Veremos alguns desses modelos na secção seguinte.

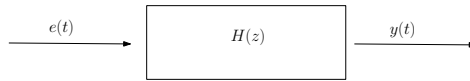


Figura 1.8: Processo estocástico estacionário, com $e(t)$ ruído branco.

Antes, porém, convém referir que o termo densidade espectral também é utilizado para designar a transformada z da sequência de autocorrelação, ou seja

$$\Phi_{xx}(z) = \sum_{\tau=-\infty}^{\infty} A_{xx}(\tau) z^{-\tau}.$$

1.5 Previsão linear em PE

Seja $\{y(t)\}$ um PE de média nula, estacionário e de segunda ordem. O nosso problema é prever $y(t+m)$, $m = 1, 2, \dots$ a partir de $y(t), y(t-1), \dots$. De acordo com o Teorema 1, podemos considerar que $y(t)$ é a saída dum sistema linear e invariante no tempo com função de transferência

$$H(z) = \sum_{k=0}^{\infty} h(k)z^{-k} \quad (1.17)$$

com $h(0) = 1$. Tem-se, então

$$y(t) = \sum_{k=0}^{\infty} h(k)e(t-k) = e(t) + h(1)e(t-1) + \dots \quad (1.18)$$

e

$$\begin{aligned} y(t+m) &= e(t+m) + h(1)e(t+m-1) + \dots + h(m-1)e(t+1) \\ &\quad + h(m)e(t) + h(m+1)e(t-m) + \dots \\ &= \sum_{k=0}^{m-1} h(s)e(t+m-k) + \sum_{k=0}^{\infty} h(m+k)e(t-k). \end{aligned} \quad (1.19)$$

Se definirmos o operador q tal que

$$qx(t) = x(t+1) \quad (1.20)$$

podemos reescrever as três últimas equações na forma

$$H(q) = \sum_{k=0}^{\infty} h(k)q^{-k} \quad (1.21)$$

$$y(t) = \sum_{k=0}^{\infty} h(k)q^{-k}e(t) = H(q)e(t) \quad (1.22)$$

$$\begin{aligned} y(t+m) &= H(q)q^m e(t) \\ &= \sum_{k=0}^{m-1} h(k)q^{m-k}e(t) + \sum_{k=0}^{\infty} h(m+k)q^{-k}e(t) \end{aligned} \quad (1.23)$$

Como $e(t)$ é ruído branco, a melhor previsão de $e(t+1), \dots, e(t+n)$ são os seus valores médios que são iguais a zero. Logo, $\hat{y}(t+1)$, a melhor previsão linear de $y(t+m)$ no instante t é

$$\hat{y}(t+m) = \sum_{k=0}^{\infty} h(m+k)q^{-k}e(t). \quad (1.24)$$

De (1.23),

$$\sum_{k=0}^{\infty} h(m+k)q^{-k}e(t) = H(q)q^m e(t) - \sum_{k=0}^{m-1} h(k)q^{m-k}e(t) \quad (1.25)$$

e, consequentemente,

$$\hat{y}(t+m) = \left[H(q)q^m - \sum_{s=0}^{m-1} h(s)q^{m-s} \right] e(t) \quad (1.26)$$

Como, segundo o Teorema 1, $H(q)$ tem todos os zeros no interior do círculo unitário, então é inversamente estável e, dada a sequência $\dots, y(t-1), y(t)$, a sequência $\dots, e(t-1), e(t)$ pode ser calculada através de

$$e(t) = H^{-1}(q)y(t) = \sum_{k=0}^{\infty} \bar{h}(k)q^{-k}y(t), \quad (1.27)$$

em que $\bar{h}(k)$, $k = 0, \dots, \infty$ é a resposta impulsional de $H(z)^{-1}$. Logo

$$\hat{y}(t+m) = \left[H(q)q^m - \sum_{s=0}^{m-1} h(s)q^{m-s} \right] H^{-1}(q)y(t). \quad (1.28)$$

De (1.23) e (1.24), o erro de previsão é

$$\begin{aligned} \hat{e}(t+m) &= y(t+m) - \hat{y}(t+m) \\ &= \sum_{k=0}^{m-1} h(k)q^{m-k}e(t) + \sum_{k=0}^{\infty} h(m+k)q^{-k}e(t) - \sum_{k=0}^{\infty} h(m+k)q^{-k}e(t) \\ &= \sum_{k=0}^{m-1} h(k)q^{m-k}e(t) \\ &= e(t+m) + h(1)e(t+m-1) + \dots + h(m-1)e(t+1) \end{aligned}$$

Na perspectiva da identificação de sistemas, a previsão mais interessante acontece para $m = 1$ que pode ser obtida através de (1.28) fazendo $m = 1$ e recordando que $h(0) = 1$:

$$\hat{y}(t+1) = [H(q)q - 1]e(t) = [H(q) - 1]qe(t) = [H(q) - 1]e(t+1)$$

como

$$e(t) = H^{-1}(q)y(t)$$

então

$$\begin{aligned} \hat{y}(t+1) &= [H(q) - 1]H^{-1}(q)y(t+1) \\ \Leftrightarrow \hat{y}(t) &= [H(q) - 1]H^{-1}(q)y(t) \end{aligned}$$

sendo o erro de previsão $y(t) - \hat{y}(t) = e(t)$.

Exemplo 9

O PE gerado por

$$y(t) = u(t) - 2u(t-1)$$

onde $u(t)$ é ruído branco com média nula e variância unitária, tem um espectro generalizado

$$\Phi_{yy}(q) = G(q)G(q^{-1})\sigma_u^2$$

onde

$$G(q) = \frac{Y(q)}{U(q)} = 1 - 2q^{-1} = \frac{q-2}{q}$$

Se suposermos que $u(t)$ é desconhecido, podemos tentar prever $y(t)$ no instante $t-1$ através de

$$\hat{y}(t) = (G(q) - 1)G^{-1}(q)y(t) = -\frac{2q^{-1}}{1-2q^{-1}}y(t) = -\frac{2}{q-2}y(t) = G_p(z)y(t).$$

No entanto este previsor não funciona, pois não é estável ($G_p(z)$ tem um zero fora do círculo unitário). A Figura 1.9 ilustra inutilidade deste previsor, onde se pode constatar que o erro de previsão cresce exponencialmente. Como podemos então obter o melhor

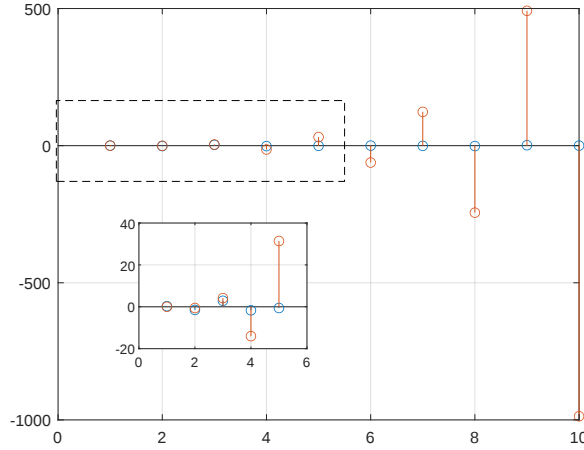


Figura 1.9: Realização de $y(t) = (1 - 2q^{-1})u(t)$ (azul) e previsão $\hat{y}(t) = -2q^{-1}/(1-2q-1)u(t)$ (vermelho). Pode-se ver na ampliação das seis primeiras amostras que, a partir de $t=4$, o erro de previsão cresce exponencialmente.

previsor linear de $y(t)$?

Para responder a esta pergunta, começamos por tentar factorizar o espectro na forma

$$\Phi_{yy}(q) = H(q)H(q^{-1})\sigma_e^2$$

em que $H(q)$ é uma função de transferência com todos os pólos e zeros no interior do círculo unitário e $h(0) = 1$. Os pólos e zeros de $H(q^{-1})$ são os inversos dos de $H(q)$ e, como tal, estão todos no exterior do círculo unitário. Uma vez obtida esta factorização, assumimos ser $y(t)$ gerado pelo modelo

$$y(t) = H(z)e(t)$$

a partir do qual podemos dimensionar o previsor:

$$\hat{y}(t) = (H(q) - 1)H^{-1}(q)y(t),$$

pois agora $H^{-1}(q)$ é estável. Notemos que a estabilidade de $H^{-1}(q)$ é fundamental para que possamos prever $e(t)$ a partir de $y(t)$. O processo que temos vindo a considerar tem um espectro

$$\Phi_{yy}(q) = G(q)G(q^{-1}) = (1 - 2q^{-1})(1 - 2q) = H(q)H(q^{-1})\sigma^2.$$

Como o zero de $G(q)$ está no exterior do círculo unitário, o de $H(q)$ tem que ser o seu inverso, isto é,

$$H(q) = 1 - 0.5q^{-1},$$

sendo

$$\begin{aligned}\Phi_{yy}(\omega) &= (1 - 0.5q^{-1})(1 - 0.5q)\sigma^2 = (1 - 2q^{-1})(1 - 2q) \\ &= (1 - 0.5q)(1 - 0.5q^{-1})(-2q^{-1})(2q) \\ &= (1 - 0.5q)(1 - 0.5q^{-1})4,\end{aligned}$$

pelo que $\sigma^2 = 4$. Temos então o seguinte modelo conceptual

$$y(t) = e(t) - 0.5e(t-1)$$

em que $e(t)$ é ruído branco de média nula e variância $\sigma^2 = 4$. O melhor previsor linear a um passo é, finalmente,

$$\hat{y}(t) = [H(q) - 1]H(q^{-1})y(t) = -\frac{0.5q^{-1}}{1 - 0.5q^{-1}}y(t) = H_p(q)y(t).$$

Como o erro de previsão é igual a $e(t)$, a sua variância é $\sigma^2 = 4$ sendo, por isso, maior do que a que, erradamente, estávamos à espera quando tentámos prever $y(t)$ através de $G_p(q)y(t)$. Notemos que a variância de $e(t)$ é igual à de $u(t)$, multiplicada por um factor igual ao quadrado do zero de $G(q)$ que reflectimos para o interior do círculo unitário quando formamos $H(q)$. Se tivéssemos reflectido mais do que um zero, o factor que multiplicaria σ_u^2 seria o quadrado do produto do módulo dos zeros reflectidos. A Figura 1.10 compara uma realização de $y(t)$ com a previsão $\hat{y}(t) = H_p(q)y(t)$ onde se pode constatar que $\hat{y}(t)$ consegue prever $y(t)$, embora com algum erro. A variância

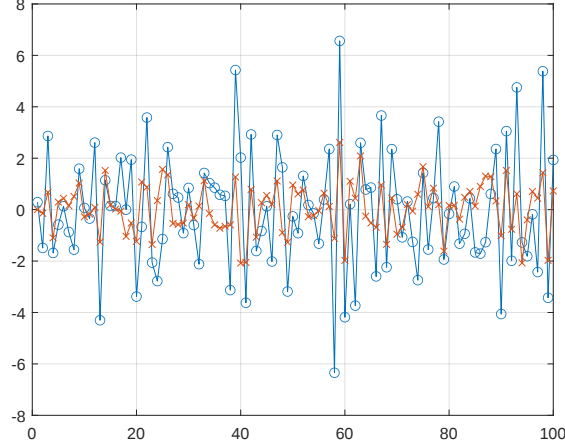


Figura 1.10: Realização de $y(t) = (1 - 2q^{-1})u(t)$ (azul) e previsão $\hat{y}(t) = -0.5q^{-1}/(1 - 0.5q^{-1})y(t)$ (vermelho).

de $y(t)$ é

$$E\{y^2(t)\} = E\{u^2(t) - 2u^2(t-1)\} = E\{u^2(t)\} + 4E\{u^2(t-1)\} = 1 + 4 = 5.$$

Logo, Prevendo $y(t)$ através de $E\{y(t)\} = 0$ obtemos um erro com variância 5 (desvio padrão) $\square$

1.6 Modelos de entrada-saída de processos estocásticos

O Teorem 1 permite-nos encarar um PE como a saída de um sistema causal, linear e invariante no tempo, estável e de fase mínima (sem zeros fora do círculo unitário). O Lema 4 mostra-nos ser a condição de estabilidade absolutamente essencial para a estacionaridade do processo. Veremos adiante que a condição de fase mínima é fundamental para se poder fazer a sua previsão. Neste contexto, um PE $\{y(t)\}$ pode ser descrito pelo modelo:

$$y(t) = \sum_{k=0}^{\infty} h(k)e(t-k) \quad \text{e} \quad h(0) = 1$$

em que os pólos e zeros de

$$H(z) = \sum_{k=0}^{\infty} h(k)z^{-k}$$

estão todos no interior do círculo unitário e $e(t)$ é ruído branco com média e variância

$$E\{e(t)^2\} = \sigma^2.$$

Se $h(k) = 0$ para $k > n$ e se designarmos $h(k)$ por c_k , o modelo será

$$y(t) = \sum_{k=0}^n c_k e(t-k)$$

Definindo z^{-1} como o operador atraso, ou seja $z^{-1}x(t) = x(t-1)$, podemos re-escrever esta equação na forma

$$y(t) = C(z)e(t)$$

em que

$$C(z) = 1 + c_1 z^{-1} + \dots + c_n z^{-n}$$

Como neste modelo há uma regressão da entrada $e(t)$, é conhecido como modelo de média model (MA, *moving average*).

Estes modelos são muito simples e permitem descrever qualquer PE. No entanto, para que tenham uma precisão aceitável, necessitam frequentemente de ter uma ordem elevada (n muito grande). Nestas condições é preferível descrever $y(t)$ por um modelo do tipo:

$$y(t) + a_1 y(t-1) + \dots + a_n y(t-n) = e(t)$$

que com

$$A(z) = 1 + a_1 z^{-1} + \dots + a_n z^{-n}$$

pode ser re-escrita como

$$A(z)y(t) = e(t).$$

Nestes modelos existe uma regressão em $y(t)$, pelo que são conhecidos como modelos auto-regressivos (AR, *auto-regressive*). A sua função de transferência é

$$H(z) = \frac{1}{A(z)}$$

e a sua ordem n costuma ser significativamente inferior à dos modelos MA. No entanto, se utilizarmos modelos com regressão na entrada, $e(t)$, e na saída, $y(t)$, conseguimos modelos e menor ordem. Estes modelos são do tipo:

$$y(t) + a_1 y(t-1) + \dots + a_n y(t-n) = e(t) + c_1 e(t-1) + \dots + c_n e(t-n)$$

e podem ser escritos na forma compacta

$$A(z)y(t) = C(z)e(t).$$

São conhecidos modelos auto-regressivos de média móvel (ARMA).

Os PE são conhecidos pelo nome dos modelos que os geram. Assim, os processos gerados por modelos ARMA são PE ARMA, os gerados por modelos AR são PE AR e os gerados por modelos MA são PE MA.

A invertibilidade de $H(z)$ requer que $\frac{1}{H(z)}$ seja uma função estável. Isto permite que se possa gerar ruído branco alimentando um sistema linear e invariante no tempo com função de transferência $\frac{1}{H(z)}$, com o processo $y(t)$. Por este motivo, $\frac{1}{H(z)}$ é

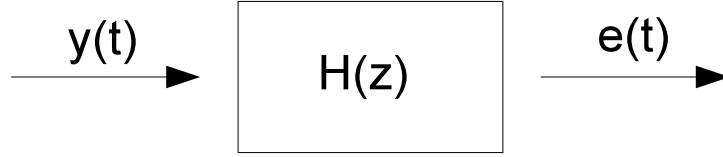


Figura 1.11: Processo estocástico estacionário, com $e(t)$ ruído branco.

também conhecido como filtro branqueador.

1.7 Modelos de PE no espaço de estados

Os PE estacionários de média nula também podem ser descritos por modelos do tipo:

$$\begin{aligned} x(t+1) &= Ax(t) + q(t) \\ y(t) &= Cx(t) + r(t) \end{aligned}$$

onde $x(t), q(t) \in \mathbb{R}^n, y(t), r(t) \in \mathbb{R}^\ell, A \in \mathbb{R}^{n \times n}, C \in \mathbb{R}^{\ell \times n}$. Neste modelo, $q(t), r(t)$ são sequências de modelo branco estacionárias, de média nula, independentes de $x(t)$, com função de covariância conjunta

$$E \left\{ \begin{pmatrix} q(t+\tau) \\ r(t+\tau) \end{pmatrix} \begin{pmatrix} q(t) \\ r(t) \end{pmatrix} \right\} = \begin{cases} \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} & , \quad \tau = 0 \\ 0_{(n+\ell) \times (n+\ell)} & , \quad \tau \neq 0 \end{cases}$$

A função covariância de $y(t)$ será

$$\begin{aligned}
\Lambda_{yy}(\tau) &= E \{y(\tau+t)y(t)\} \\
&= E \left\{ (Cx(t+\tau) + r(t+\tau)) (Cx(t) + r(t))^T \right\} \\
&= E \left\{ (Cx(t+\tau) + r(t+\tau)) (x^T(t)C^T + r^T(t))^T \right\} \\
&= E \{ Cx(t+\tau)x^T(t)C^T + Cx(t+\tau)r^T(t) + r(t+\tau)x^T(t)C^T + r(t+\tau)r^T(t) \} \\
&= CE \{x(t+\tau)x^T(t)\} C^T + CE \{x(t+\tau)r^T(t)\} + E \{r(t+\tau)x^T(t)\} C^T + E \{r(t+\tau)r^T(t)\} \\
&= C\Lambda_{xx}(\tau)C^T + C\Lambda_{xr}(\tau) + \Lambda_{rx}(\tau)C^T + \Lambda_{rr}(\tau)
\end{aligned}$$

Como para $\tau > 0$,

$$x(t+\tau) = A^\tau x(t) + \sum_{k=0}^{\tau-1} A^{\tau-1-k} q(t+\tau)$$

temos:

$$\begin{aligned}
\Lambda_{xr}(\tau) &= E \{x(t+\tau)r^T(t)\} \\
&= E \left\{ \left(A^\tau x(t) + \sum_{k=0}^{\tau-1} A^{\tau-1-k} q(t+\tau) \right) r^T(t) \right\} \\
&= A^\tau E \{x(t)r^T(t)\} + \sum_{k=0}^{\tau-1} A^{\tau-1-k} E \{q(t+\tau)r^T(t)\}
\end{aligned}$$

Uma vez que $r(t)$ é invariante de $x(t)$ então $E \{x(t)r^T(t)\} = 0_{n \times \ell}$. Por outro lado, da função de covariância conjunta, vemos que:

$$E \{q(t+k)r^T(t)\} = \begin{cases} 0_{n \times \ell} & , \quad k \neq 0 \\ S & , \quad k = 0 \end{cases}$$

o que nos permite concluir que $\Lambda_{xr}(\tau) = A^{\tau-1}S$. Quando $\tau \leq 0$, $x(t+\tau)$ não tem nada a ver com $r(t)$ sendo por isso independentes estas variáveis e, consequentemente, $\Phi_{xr} = 0_{n \times \ell}$. Por outro lado:

$$\begin{aligned}
\Lambda_{rx}(\tau) &= E \{r(t+\tau)x^T(t)\} \\
&= (E \{x(t)r^T(t+\tau)\})^T \\
&= \Lambda_{xr}(-\tau)
\end{aligned}$$

Sendo por isso

$$\Lambda_{rx}(\tau) = \begin{cases} S^T (A^{-\tau-1})^T & , \quad \tau < 0 \\ 0_{n \times \ell} & , \quad \tau \geq 0 \end{cases}$$

sendo $r(t)$ ruído branco com média nula e covariância R , a sua função de covariância é

$$\Lambda_{rr}(\tau) = \begin{cases} R & , \quad \tau = 0 \\ 0_{\ell \times \ell} & , \quad \tau \neq 0 \end{cases}$$

Finalmente designamos:

$$\Lambda_{xx}(\tau) = \Pi(\tau), \tau \neq 0$$

e simplesmente por $\Pi, \tau = 0$. Com tudo isto, podemos reescrever a função de covariância de $y(t)$

$$\Lambda_{yy}(\tau) = \begin{cases} C\Pi C^T + R & , \quad \tau = 0 \\ C\Pi C(\tau)^T + CA^{\tau-1}S & , \quad \tau > 0 \\ C\Pi(\tau)C^T + S^T (A^{-\tau-1})^T C^T & , \quad \tau < 0 \end{cases}$$

Π pode ser calculado através de:

$$\begin{aligned} \Pi &= E \{x(t+1)x^T(t+1)\} \\ &= E \{(Ax(t) + q(t))(Ax(t) + q(t))^T\} \\ &= E \{(Ax(t) + q(t))(x^T(t)A^T + q^T(t))\} \\ &= AE \{x(t)x^T(t)\} A^T + E \{q(t)x^T(t)\} A^T + AE \{x(t)q^T(t)\} + E \{q(t)q^T(t)\} \\ &= A\Pi A^T + \Lambda_{qx}(0)A^T + A\Lambda_{xq}(0) + R \end{aligned}$$

Como $q(t)$ é independente de $x(t)$, temos que $\Lambda_{qx}(0) = \Lambda_{xq}(0) = 0_{n \times n}$ e Π é a solução da equação de Lyapunov:

$$\Pi = A\Pi A^T + Q.$$

Para calcular $\Pi(\tau)$ fazemos:

$$\begin{aligned} \Pi(\tau) &= E \{x(t+\tau)x^T(t)\} \\ &= E \{(Ax(t+\tau-1) + q(t+\tau-1))x^T(t)\} \\ &= AE \{x(t+\tau-1)x^T(t)\} + E \{q(t+\tau-1)x^T(t)\} \\ &= A\Pi(\tau-1) + \Lambda_{qx}(\tau-1) \end{aligned}$$

Sendo $\tau > 0$, como $q(t), q(t+1), \dots$ é uma sequência independente de $x(t)$, temos: $\Lambda_{qx}(\tau-1) = 0_{n \times n}$. Sendo $\Pi(\tau) = A\Pi(\tau-1)$, recursivamente, tem-se então que:

$$\Pi(\tau) = A^\tau \Pi(\tau-1).$$

Como

$$\begin{aligned} \Pi(\tau) &= E \{x(t+\tau)x^T(t)\} \\ &= (E \{x(t)x^T(t+\tau)\})^T \\ &= \Pi(-\tau) \end{aligned}$$

$\Pi(\tau)$, para $\tau < 0$ será

$$\Pi(\tau) = \Pi(A^{-\tau})^T.$$

Substituindo $\Pi(\tau)$ em:

$$\Lambda_{yy}(\tau) = \begin{cases} C\Pi C^T + R & , \quad \tau = 0 \\ CA^\tau \Pi C^T + CA^{\tau-1}S & , \quad \tau > 0 \\ C\Pi(CA^{-\tau})^T + S^T (CA^{-\tau-1})^T & , \quad \tau < 0 \end{cases}$$

1.8 Espaços de vectores aleatórios

Antes de abordar o filtro de Kalman, vamos alargar o conceito de espaço gerado por v.a.s de média nula, escalares apresentado na Secção 1.5 a v.a. de dimensão n . Seja então uma v.a. n -dimensional, $x \in \mathbb{R}^n$, com momento de segunda ordem finito

$$E \left\{ \|x\|_2^2 \right\} = \sum_{i=1}^n E \left\{ x_i^2 \right\} < \infty$$

e média nula. Designaremos por H o conjunto de todos os vectores aleatórios n -dimensionais com média nula e momento de segunda ordem finito

$$H = \left\{ x \mid E \{x\} = 0_n \text{ e } E \left\{ \|x\|_2^2 \right\} < \infty \right\}$$

É muito fácil provar que H é um espaço linear. Se para $x, y \in H$, definirmos o produto interno como

$$(x, y)_H = E \{x^T y\},$$

induzimos em H a norma

$$\|x\|_2^2 = \sqrt{(x, x)_H} = \sqrt{E \left\{ \|x\|_2^2 \right\}}$$

que o completa, passando assim H a ser um espaço de Hilbert dos vectores aleatórios de média nula e momento de segunda ordem finito.

O produto interno permite também introduzir o conceito de perpendicular. Dizemos que $x, y \in H$ e $x \perp y$ se $(x, y)_H = 0$. Se $Y \subset H$ e $(x, y)_H = 0, \forall y \in Y$, dizemos que $x \perp Y$. Do Lemma ??, sabemos que existe um e um só $y_0 \in Y$ tal que

$$\|x - y_0\|_H \leq \|x - y\|_H \quad \forall y \in Y$$

Sendo então y_0 o mínimo de $\|x - y\|_H \quad \forall y \in Y$, temos que $x - y_0 \perp Y$ é a condição de optimalidade.

Sejam $y_1, y_2, \dots, y_N$ vectores aleatórios de dimensão p , com média nula e momento de segunda ordem finito. Definimos agora Y como sendo o espaço gerado por $y_1, y_2, \dots, y_N$, ou seja

$$Y = \left\{ \hat{x} = \sum_{i=1}^N A_i y_i \mid A_i \in \mathbb{R}^{n \times p} \right\} \quad (1.29)$$

onde todos os elementos $\hat{x} \in Y$ são vectores n -dimensionais com média nula e momento de segunda ordem finito. Se Y for completado com a norma $\|\cdot\|_H$, passa a ser um subespaço de H e, consequentemente, $Y \subset H$.

Lema 5. $\tilde{x} \in H$ é ortogonal ao subespaço Y definido em (1.29) se e só se

$$E \left\{ \tilde{x} y_i^T \right\} = 0_{n \times p}, \quad i = 1, \dots, n.$$

Demonstração. Se $E \{ \tilde{x} y_i^T \} = 0_{n \times p}$ então para qualquer $\hat{x} \in Y$

$$\begin{aligned}
(\tilde{x}^T, \hat{x})_H &= E \left\{ \tilde{x}^T \left(\sum_{i=1}^N A_i y_i \right) \right\} = \sum_{i=1}^N E \{ \tilde{x}^T (A_i y_i) \} \\
&= \sum_{i=1}^N E \{ \text{traço } \tilde{x} (A_i y_i)^T \} \\
&= \sum_{i=1}^N E \{ \text{traço } \tilde{x} y_i^T A_i^T \} \\
&= \sum_{i=1}^N \text{traço } E \{ \tilde{x} y_i^T A_i^T \} \\
&= \sum_{i=1}^N \text{traço } E \{ \tilde{x} y_i^T \} A_i^T = 0
\end{aligned}$$

Logo $\tilde{x} \perp \hat{x}$ e, consequentemente, $\tilde{x} \perp Y$.

Suponhamos agora que $\tilde{x} \perp Y$, logo

$$\begin{aligned}
(\tilde{x}, A_i y_i)_H &= E \{ \tilde{x}^T (A_i y_i) \} \\
&= E \{ \text{traço } \tilde{x} y_i^T A_i^T \} \\
&= \text{traço } E \{ \tilde{x} y_i^T A_i^T \} \\
&= \text{traço } E \{ \tilde{x} y_i^T \} A_i^T = 0, \quad i = 1, \dots, N
\end{aligned}$$

e como A_i pode ser qualquer matriz de ordem $n \times p$, podemos fazer $A_i = E \{ \tilde{x} y_i^T \}$ o que conduz a

$$\text{traço } E \{ \tilde{x} y_i^T \} A_i^T = \text{traço } \{ A_i A_i^T \} = \|A_i\|_F^2 = 0 \implies A_i = E \{ \tilde{x} y_i^T \} = 0_{n \times p}, \quad i = 1, \dots, n$$

□

1.9 Projecções ortogonais em espaços de vectores aleatórios

A definição de produto interno permitiu-nos alargar o conceito de ortogonalidade a H . Ora se existe o conceito de ortogonalidade também existe o de projecção ortogonal. Assim, $y_0 \in Y$ e $Y \subset H$ é a projecção ortogonal de $x \in H$ em Y se $x - y_0 \perp Y$. Assim, a melhor aproximação de x em Y é a projecção de x em Y : $\hat{E}(x/Y)$.

Conseguimos desta forma transformar um problema de optimização num problema geométrico de determinar projecções. Sendo Y um subespaço de H , este problema pode ser resolvido com recurso a técnicas de Álgebra Linear.

Lema 6. *Seja Y um subespaço de H definido por $y_1, y_2, \dots, y_N \in \mathbb{R}^p$ e $x \in \mathbb{R}^n$ um elemento de H . A projecção ortogonal de x em Y é*

$$\hat{E}(x/Y) = \Sigma_{xy} \Sigma_{yy}^{-1} \bar{y} \quad (1.30)$$

em que

$$\begin{aligned}\bar{y} &= \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^{Np} \\ \Sigma_{x\bar{y}} &= E \{ x \bar{y}^T \} \in \mathbb{R}^{n \times Np} \\ \Sigma_{\bar{y}\bar{y}} &= E \{ \bar{y} \bar{y}^T \} \in \mathbb{R}^{Np \times Np}\end{aligned}$$

Demonstração. A expressão (1.30) mostra-nos que $\hat{E}(x/Y) \in Y$, pois é uma combinação linear de $y_1, y_2, \dots, y_N$. Para provar que é a projecção ortogonal de x em Y , só temos que demonstrar

$$\tilde{x} = x - \hat{E}(x/Y) \perp y_i, \quad i = 1, \dots, n$$

Se definirmos:

$$\begin{aligned}E_1 &= (I_p \mid 0_{p \times p} \mid \dots \mid 0_{p \times p}) \in \mathbb{R}^{p \times Np} \\ E_2 &= (0_{p \times p} \mid I_p \mid \dots \mid 0_{p \times p}) \in \mathbb{R}^{p \times Np} \\ &\vdots \\ E_i &= (0_{p \times p} \mid 0_{p \times p} \mid \dots \mid I_p \mid \dots \mid 0_{p \times p}) \in \mathbb{R}^{p \times Np} \\ &\vdots \\ E_N &= (0_{p \times p} \mid 0_{p \times p} \mid \dots \mid I_p) \in \mathbb{R}^{p \times Np}\end{aligned}$$

então $y_i = E_i \bar{y}$, $i = 1, \dots, N$.

o que nos permite calcular:

$$E \{ \tilde{x} \bar{y}_i^T \} = E \left\{ \left(x - \hat{E}(x/Y) \right) (E_i \bar{y})^T \right\} \quad (1.31)$$

$$= E \left\{ \left(x - \hat{E}(x/Y) \right) \bar{y}^T E_i^T \right\} \quad (1.32)$$

$$= E \left\{ \left(x - \hat{E}(x/Y) \right) \bar{y}^T \right\} E_i^T \quad (1.33)$$

$$= E \left\{ x \bar{y}^T - \Sigma_{x\bar{y}} \Sigma_{\bar{y}\bar{y}}^{-1} \bar{y} \bar{y}^T \right\} E_i^T \quad (1.34)$$

$$= (E \{ x \bar{y}^T \} - \Sigma_{x\bar{y}} \Sigma_{\bar{y}\bar{y}}^{-1} E \{ \bar{y} \bar{y}^T \}) E_i^T \quad (1.35)$$

$$= (\Sigma_{x\bar{y}} - \Sigma_{x\bar{y}} \Sigma_{\bar{y}\bar{y}}^{-1} \Sigma_{\bar{y}\bar{y}}) E_i^T \quad (1.36)$$

$$= (\Sigma_{x\bar{y}} - \Sigma_{x\bar{y}}) E_i^T = 0_{n \times p}, \quad i = 1, \dots, N \quad (1.37)$$

Como $E \{ \tilde{x} \bar{y}_i^T \} = 0_{n \times p}$, então pelo Lemma 5, $\tilde{x}$ é ortogonal a Y e consequentemente $\hat{E}(x/Y)$ é a projecção ortogonal de x em Y $\square$

Sejam Y_1 e Y_2 dois subespaços de H . Dizemos que Y é a *soma directa* de Y_1 e Y_2 , e designa-se por $Y_1 \oplus Y_2$ se

$$Y_1 \cap Y_2 = \emptyset \quad \text{e} \quad Y_1 \cup Y_2 = Y.$$

Vejam os de seguida como podemos decompôr a projecção de $x \in H$ num subespaço $Y = Y_1 \oplus Y_2$ e $Y_1 \perp Y_2$.

Lema 7. Sejam Y, Y_1, Y_2 subespaços de H tal que $Y = Y_1 \oplus Y_2$ e $Y_1 \perp Y_2$. A projecção ortogonal de $x \in H$ em Y é

$$\hat{E}(x/Y) = \hat{E}(x/Y_1) + \hat{E}(x/Y_2)$$

Demonstração. Sejam $y_{11}, y_{12}, \dots, y_{1N}, y_{21}, y_{22}, \dots, y_{2N}$ um conjunto de vectores aleatórios linearmente independentes que gera Y e em que $y_{11}, y_{12}, \dots, y_{1N} \in Y_1$ e $y_{21}, y_{22}, \dots, y_{2N} \in Y_2$. Se definirmos

$$\bar{y}_1 = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1N} \end{pmatrix}, \quad \bar{y}_2 = \begin{pmatrix} y_{21} \\ y_{22} \\ \vdots \\ y_{2N} \end{pmatrix} \quad \text{e} \quad \bar{y} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix}$$

a projecção ortogonal $\hat{E}(x/Y)$ é dada por $\hat{E}(x/Y) = \Sigma_{x\bar{y}} \Sigma_{\bar{y}\bar{y}}^{-1} \bar{y}$ em que

$$\begin{aligned} \Sigma_{x\bar{y}} &= E\{\tilde{x}\bar{y}^T\} = E\{\tilde{x}(\bar{y}_1^T | \bar{y}_2^T)\} \\ &= E\{(\tilde{x}\bar{y}_1^T | \tilde{x}\bar{y}_2^T)\} \\ &= (E\{\tilde{x}\bar{y}_1^T\} | E\{\tilde{x}\bar{y}_2^T\}) \\ &= (\Sigma_{x\bar{y}_1} | \Sigma_{x\bar{y}_2}) \end{aligned}$$

e

$$\begin{aligned} \Sigma_{\bar{y}\bar{y}} &= E\{\bar{y}\bar{y}^T\} = E\left\{\begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} (\bar{y}_1^T | \bar{y}_2^T)\right\} \\ &= E\left\{\begin{pmatrix} \bar{y}_1\bar{y}_1^T & \bar{y}_1\bar{y}_2^T \\ \bar{y}_2\bar{y}_1^T & \bar{y}_2\bar{y}_2^T \end{pmatrix}\right\} \\ &= \begin{pmatrix} E\{\bar{y}_1\bar{y}_1^T\} & E\{\bar{y}_1\bar{y}_2^T\} \\ E\{\bar{y}_2\bar{y}_1^T\} & E\{\bar{y}_2\bar{y}_2^T\} \end{pmatrix} \end{aligned}$$

Como $Y_1 \perp Y_2$, temos que $0_{p_1 \times p_2} = E\{\bar{y}_1\bar{y}_2^T\} = E\{\bar{y}_2\bar{y}_1^T\} = 0_{p_2 \times p_1}$ e

$$\begin{aligned} \Sigma_{\bar{y}\bar{y}} &= \begin{pmatrix} \Sigma_{\bar{y}_1\bar{y}_1} & 0_{p_1 \times p_2} \\ 0_{p_2 \times p_1} & \Sigma_{\bar{y}_2\bar{y}_2} \end{pmatrix} \\ \implies \Sigma_{\bar{y}\bar{y}}^{-1} &= \begin{pmatrix} \Sigma_{\bar{y}_1\bar{y}_1}^{-1} & 0_{p_1 \times p_2} \\ 0_{p_2 \times p_1} & \Sigma_{\bar{y}_2\bar{y}_2}^{-1} \end{pmatrix} \end{aligned}$$

Substituindo agora $\Sigma_{x\bar{y}}$ e $\Sigma_{\bar{y}\bar{y}}$ pelos seu valores na expressão de $\hat{E}(x/Y)$, obtemos:

$$\begin{aligned} \hat{E}(x/Y) &= (\Sigma_{x\bar{y}_1} | \Sigma_{x\bar{y}_2}) \begin{pmatrix} \Sigma_{\bar{y}_1\bar{y}_1}^{-1} & 0_{p_1 \times p_2} \\ 0_{p_2 \times p_1} & \Sigma_{\bar{y}_2\bar{y}_2}^{-1} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \\ &= (\Sigma_{x\bar{y}_1} \Sigma_{\bar{y}_1\bar{y}_1}^{-1} | \Sigma_{x\bar{y}_2} \Sigma_{\bar{y}_2\bar{y}_2}^{-1}) \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \end{pmatrix} \\ &= \Sigma_{x\bar{y}_1} \Sigma_{\bar{y}_1\bar{y}_1}^{-1} \bar{y}_1 + \Sigma_{x\bar{y}_2} \Sigma_{\bar{y}_2\bar{y}_2}^{-1} \bar{y}_2 \\ &= \hat{E}(x/Y_1) + \hat{E}(x/Y_2) \end{aligned}$$

□

1.10 Inovações

Seja

$$Y = \left\{ \sum_{i=1}^N A_i y_i \mid A_i \in \mathbb{R}^{n \times p}, y_i \in \mathbb{R}^p \right\}$$

um subespaço de H . Existe um conjunto de vectores aleatórios $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N$ não correlacionados entre si, tal que

$$Y = \left\{ \sum_{i=1}^N A_i y_i = \sum_{i=1}^N \tilde{A}_i \tilde{y}_i \mid A_i, \tilde{A}_i \in \mathbb{R}^{n \times p}, y_i, \tilde{y}_i \in \mathbb{R}^p \right\}$$

Chamamos inovações aos vectores aleatórios $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N$.

Notemos que a dimensão das inovações não tem de ser a mesma de $\hat{x} \in H$, (as inovações têm dimensão p e $\hat{x}$ tem dimensão n).

Exemplo 10 Vamos calcular as inovações do subespaço Y_N gerado por as v.a.s $y_1, \dots, y_N$. Sejam Y_k os subespaços:

$$Y_k = \left\{ \sum_{i=1}^k A_i y_i \mid A_i \in \mathbb{R}^{n \times p}, y_i \in \mathbb{R}^p \right\} \quad k = 1, \dots, n.$$

Definimos

$$\begin{aligned} \tilde{y}_1 &= y_1 \\ \tilde{y}_2 &= y_2 - \hat{E}(y_2 / Y_1) = y_2 - B_{21} y_1 \\ &\vdots \\ \tilde{y}_N &= y_N - \hat{E}(y_N / Y_{N-1}) = y_N - \sum_{k=1}^N B_{Nk} y_k \end{aligned}$$

onde $\hat{E}(y_k / Y_{k-1})$, $k = 1, \dots, N$ é a projecção ortogonal de y_k no subespaço Y_{k-1} . Estas equações podem ser escritas na forma matricial

$$\begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_N \end{pmatrix} = \begin{pmatrix} I_p & 0_{p \times p} & \cdots & 0_{p \times p} \\ B_{21} & I_p & \cdots & 0_{p \times p} \\ \vdots & & \ddots & \vdots \\ B_{N1} & ? & \cdots & I_p \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

Dado que nesta expressão a matriz de ordem $Np \times Np$ é triangular inferior com 1s na diagonal principal, então é não singular. Vemos assim que y_k também pode ser expresso como uma combinação linear de $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N$.

Se isto acontece, então $x_k \in Y_k \implies \hat{x}_k = \sum_{i=1}^k A_{ki} y_k = \sum_{i=1}^k \tilde{A}_{ki} \tilde{y}_k$. Vamos agora demonstrar que $\tilde{y}_k, k = 1, \dots, N$ são vectores não correlacionados. Como $\tilde{y}_k = y_k - \hat{E}(y_k / Y_{k-1})$, então $\tilde{y}_k \perp Y_{k-1} \implies \tilde{y}_k \perp \tilde{y}_i, i = 1, \dots, k-1$ e $k = 2, \dots, N$. Concluimos daqui serem todos os vectores $\tilde{y}_k$ perpendiculares entre si. Logo $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N$ são as inovações do conjunto de vectores $y_1, y_2, \dots, y_N$. Notemos ainda que este método se assemelha ao método de Gram-Schmidt para a construção de bases ortogonais para espaços vectoriais. $\square$