Escola Politécnica da Universidade de São Paulo

PTC3451 – Processamento Estatístico e Adaptativo de Sinais

Prof.ª Maria D. Miranda – EPUSP, 18/12/2021

Autor: **Vinicius Bueno de Moraes** | NUSP: **10256432**

Segunda Prova – Parte Computacional

**script.m completo utilizado para as resoluções segue em anexo aqui.*

Instruções:

- A parte computacional da segunda prova de PTC3451 será disponibilizada no dia 16 de dezembro às 15h e deverá ser entregue até o dia 18 de dezembro às 23h59.
- As questões devem ser resolvidas individualmente. Caso sejam identificados indícios de cópia ou de propagação de erros em alguma questão, a ela será atribuída nota zero em todas as provas relacionadas.

Questão 1 (3,0 pontos)

Um sinal de interesse é observado, ao longo do tempo, por meio de um sensor que introduz ruído. Considere o modelo em tempo discreto

$$x(n) = s(n) + \eta(n),$$

em que $x(n)$ é o sinal observado, $s(n)$ é o sinal de interesse e $\eta(n)$ é o ruído de medida. Supõe-se que os processos $s(n)$ e $\eta(n)$ possuem médias nulas, são estacionários e independentes entre si. O processo $\eta(n)$ é branco com variância $\sigma_\eta^2 = 1$.

A fim de levar em conta a diversidade temporal dos sinais envolvidos, definem-se os vetores

$$\boldsymbol{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ x(n-2) \end{bmatrix}, \qquad \boldsymbol{s}(n) = \begin{bmatrix} s(n) \\ s(n-1) \\ s(n-2) \end{bmatrix}, \qquad \boldsymbol{\eta}(n) = \begin{bmatrix} \eta(n) \\ \eta(n-1) \\ \eta(n-2) \end{bmatrix}$$

que satisfazem a equação vetorial

$$\boldsymbol{x}(n) = \boldsymbol{s}(n) + \boldsymbol{\eta}(n).$$

Sabe-se que a matriz de autocorrelação do sinal de interesse é dada por

$$\boldsymbol{R}_S = \mathbb{E}[\boldsymbol{s}(n)\, \boldsymbol{s}^T(n)] = \begin{bmatrix} r_s(0) & r_s(1) & r_s(2) \\ r_s(1) & r_s(0) & r_s(1) \\ r_s(2) & r_s(1) & r_s(0) \end{bmatrix},$$

em que $r_s(0) = 20{,}00,\ r_s(1) = 19{,}39\ e\ r_s(2) = 17{,}59$.

Considerando-se esse modelo vetorial, pedem-se:

(a) Determine a matriz de autocorrelação do sinal observado, isto é, $\boldsymbol{R}_x = \mathbb{E}[\boldsymbol{x}(n)\,\boldsymbol{x}^T(n)]$.

Como os processos $s(n)$ e $\eta(n)$ são não correlacionados, tem-se:

$$\boldsymbol{R}_x = \boldsymbol{R}_s + \boldsymbol{R}_\eta = \begin{bmatrix} 20{,}00 & 19{,}39 & 17{,}59 \\ 19{,}39 & 20{,}00 & 19{,}39 \\ 17{,}59 & 19{,}39 & 20{,}00 \end{bmatrix} + \sigma_\eta^2 \boldsymbol{I} \Rightarrow \boldsymbol{R}_x = \begin{bmatrix} 21{,}00 & 19{,}39 & 17{,}59 \\ 19{,}39 & 21{,}00 & 19{,}39 \\ 17{,}59 & 19{,}39 & 21{,}00 \end{bmatrix}$$

(b) Com base na resposta ao item (a), determine os vetores-coluna $\boldsymbol{q}_1, \boldsymbol{q}_2\ e\ \boldsymbol{q}_3 \in \mathbb{R}^3$ correspondentes às componentes principais do processo $x(n)$.
DICA: Você pode usar o MATLAB com o comando $[\boldsymbol{Q}, \boldsymbol{D}] = eig(\boldsymbol{R}_x)$. Lembre que as colunas $\boldsymbol{q}_i$ da matriz $\boldsymbol{Q}$ são os autovetores associados aos autovalores λ_i da matriz $\boldsymbol{D}$.

Usando a função do MATLAB citada na dica anterior, chegou-se em:

$$\boldsymbol{q}_1 = \begin{bmatrix} 0{,}4167 \\ -0{,}8079 \\ 0{,}4167 \end{bmatrix},\ \boldsymbol{q}_2 = \begin{bmatrix} 0{,}7071 \\ 0{,}0000 \\ -0{,}7071 \end{bmatrix},\ \boldsymbol{q}_3 = \begin{bmatrix} 0{,}5713 \\ 0{,}5893 \\ 0{,}5713 \end{bmatrix}.$$

(c) Sejam as projeções da observação $x(n)$, sobre suas componentes principais, dadas por

$y_1(n) = \boldsymbol{q}_1^T \boldsymbol{x}(n),\ y_2(n) = \boldsymbol{q}_2^T \boldsymbol{x}(n),\ y_3(n) = \boldsymbol{q}_3^T \boldsymbol{x}(n).$

i. Determine a variância explicada de cada uma das componentes principais.

De posse da dica apresentada no item anterior (b) e a definição de ***proporção de variância explicada*** sendo $\frac{\lambda_j}{\sum_{i=1}^{K} \lambda_i} * 100$ obteve-se para cada uma das componentes principais com o uso do MATLAB:

$$Variância\ 1 = 0{,}9975 \text{ e } Proporção\ da\ Variância\ Explicada\ 1 = 1{,}5833\%,$$
$$Variância\ 2 = 3{,}4100 \text{ e } Proporção\ da\ Variância\ Explicada\ 2 = 5{,}4127\%,$$
$$Variância\ 3 = 58{,}5925 \text{ e } Proporção\ da\ Variância\ Explicada\ 3 = 93{,}0040\%.$$

ii. Quais das componentes principais asseguram a maior relação sinal-ruído? Justifique adequadamente sua resposta.
DICA: Seja $\boldsymbol{A}$ uma matriz diagonalizável tal que $\boldsymbol{A} = \boldsymbol{Q\Lambda Q}^{-1}$ e $\boldsymbol{\Lambda}$ é uma matriz diagonal. Se $\boldsymbol{B} = \boldsymbol{A} + \delta\boldsymbol{I}$, então $\boldsymbol{B} = \boldsymbol{Q}(\boldsymbol{A} + \delta\boldsymbol{I})\boldsymbol{Q}^{-1}$ para todo δ real.

De posse da DICA anterior e da relação

$$R_x = Q\left(\sigma_s^2 \Lambda + \sigma_\eta^2 I\right) Q^T,$$

onde

$$\sigma_s^2 \Lambda + \sigma_\eta^2 I = \Lambda',$$

em que

$$\Lambda' = \begin{bmatrix} 0{,}9975 & 0 & 0 \\ 0 & 3{,}4100 & 0 \\ 0 & 0 & 58{,}5925 \end{bmatrix} \text{ e } \sigma_\eta^2 I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

chega-se diretamente, que a relação entre sinal (σ_s^2) e ruído (σ_η^2) para cada uma das componentes é:

$$SNR_{Y1} = \frac{0{,}9975-1}{1} \cong 0[dB];$$
$$SNR_{Y2} = \frac{3{,}4100-1}{1} \cong 2{,}41 \; ou \; 3{,}82[dB];$$
$$SNR_{Y3} = \frac{58{,}5925-1}{1} \cong 57{,}5925 \; ou \; 17{,}60[dB].$$

Desta forma, $SNR_{Y3} > SNR_{Y2} > SNR_{Y1}$ e deve-se escolher a projeção $y_3(n)$, tendo está sinal de interesse mais realçado frente ao ruído.

(d) *Parte computacional*. Na página da disciplina, é fornecido o arquivo **dadosAP2.mat** com três observações ruidosas de dados

$$\boldsymbol{x1} = \mathrm{x}(1\!:\mathrm{K}-2);$$
$$\boldsymbol{x2} = \mathrm{x}(2\!:\mathrm{K}-1);$$
$$\boldsymbol{x3} = \mathrm{x}(3\!:\mathrm{K});$$

Defina a matriz de dados

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x1} \\ \boldsymbol{x2} \\ \boldsymbol{x3} \end{bmatrix}$$

que possui 3 linhas e 8758 colunas. Utilize o PCA (*Principal Component Analysis*) com redução de dimensionalidade para realçar o sinal de interesse. Siga as instruções dadas a seguir:

i. Estime a matriz de covariância $\boldsymbol{R}_x$ a partir das observações fornecidas. Compare o resultado numérico com a resposta do item (a).

**script.m completo utilizado para as resoluções segue em anexo aqui.*

A partir dos seguintes comandos no MATLAB:

```
load dadosAP2.mat                  %Carrega os dados fornecidos (x1, x2 e x3)
X = [x1; x2; x3]                   %Composição da Matriz X
RxEstimado = X*X'/size(X, 2)  %Estimativa da Matriz de Covariância
```

Se chegou à estimação $\boldsymbol{R}_x$ através das observações fornecidas igual a:

$$\boldsymbol{R}_{x(estimado)} = \begin{bmatrix} 20{,}98 & 19{,}39 & 17{,}55 \\ 19{,}39 & 20{,}98 & 19{,}39 \\ 17{,}55 & 19{,}39 & 20{,}98 \end{bmatrix} \cong \boldsymbol{R}_{x(teórico)} = \begin{bmatrix} 21{,}00 & 19{,}39 & 17{,}59 \\ 19{,}39 & 21{,}00 & 19{,}39 \\ 17{,}59 & 19{,}39 & 21{,}00 \end{bmatrix}$$

Os resultados são verossímeis, a menos de diferenças resultantes da implementação experimental.

ii. Faça o PCA com base na decomposição em autovalores e autovetores dessa matriz. Compare os resultados numéricos obtidos com as respostas teóricas dos itens (b) e (c).

Através do comando no MATLAB

[QEstimado, DEstimado] = eig(RxEstimado)

obteve-se $\boldsymbol{q}_{1(estimado)}, \boldsymbol{q}_{2(estimado)}\ e\ \boldsymbol{q}_{3(estimado)}\ \in\ \mathbb{R}^3$ – em comparação com o item (b), tem-se:

$$\boldsymbol{q}_{1(estimado)} = \begin{bmatrix} 0{,}4168 \\ -0{,}8077 \\ 0{,}4169 \end{bmatrix} \cong \boldsymbol{q}_{1(teórico)} = \begin{bmatrix} 0{,}4167 \\ -0{,}8079 \\ 0{,}4167 \end{bmatrix},$$

$$\boldsymbol{q}_{2(estimado)} = \begin{bmatrix} 0{,}7071 \\ -0{,}0001 \\ -0{,}7071 \end{bmatrix} \cong \boldsymbol{q}_{2(teórico)} = \begin{bmatrix} 0{,}7071 \\ 0{,}0000 \\ -0{,}7071 \end{bmatrix},$$

$$\boldsymbol{q}_{3(estimado)} = \begin{bmatrix} 0{,}5712 \\ 0{,}5896 \\ 0{,}5711 \end{bmatrix} \cong \boldsymbol{q}_{3(teórico)} = \begin{bmatrix} 0{,}5713 \\ 0{,}5893 \\ 0{,}5713 \end{bmatrix},$$

Para a Variância Explicada Estimada, obteve-se – em comparação com o item (c):

$$Variância\ Explicada\ 1_{(estimado)} = 1{,}5833\% \text{ e } Variância\ Explicada\ 1_{(teórico)} = 1{,}5377\%,$$
$$Variância\ Explicada\ 2_{(estimado)} = 5{,}4127\% \text{ e } Variância\ Explicada\ 2_{(teórico)} = 5{,}4452\%,$$
$$Variância\ Explicada\ 3_{(estimado)} = 93{,}0040\% \text{ e } Variância\ Explicada\ 3_{(teórico)} = 93{,}0172\%.$$

Os resultados são verossímeis, a menos de diferenças de resultantes da implementação experimental.

iii. Para auxiliar na interpretação do PCA, faça gráficos de $x_1(n)$, $x_2(n)$, $x_3(n)$, $y_1(n)$, $y_2(n)$ e $y_3(n)$ ao longo do tempo.

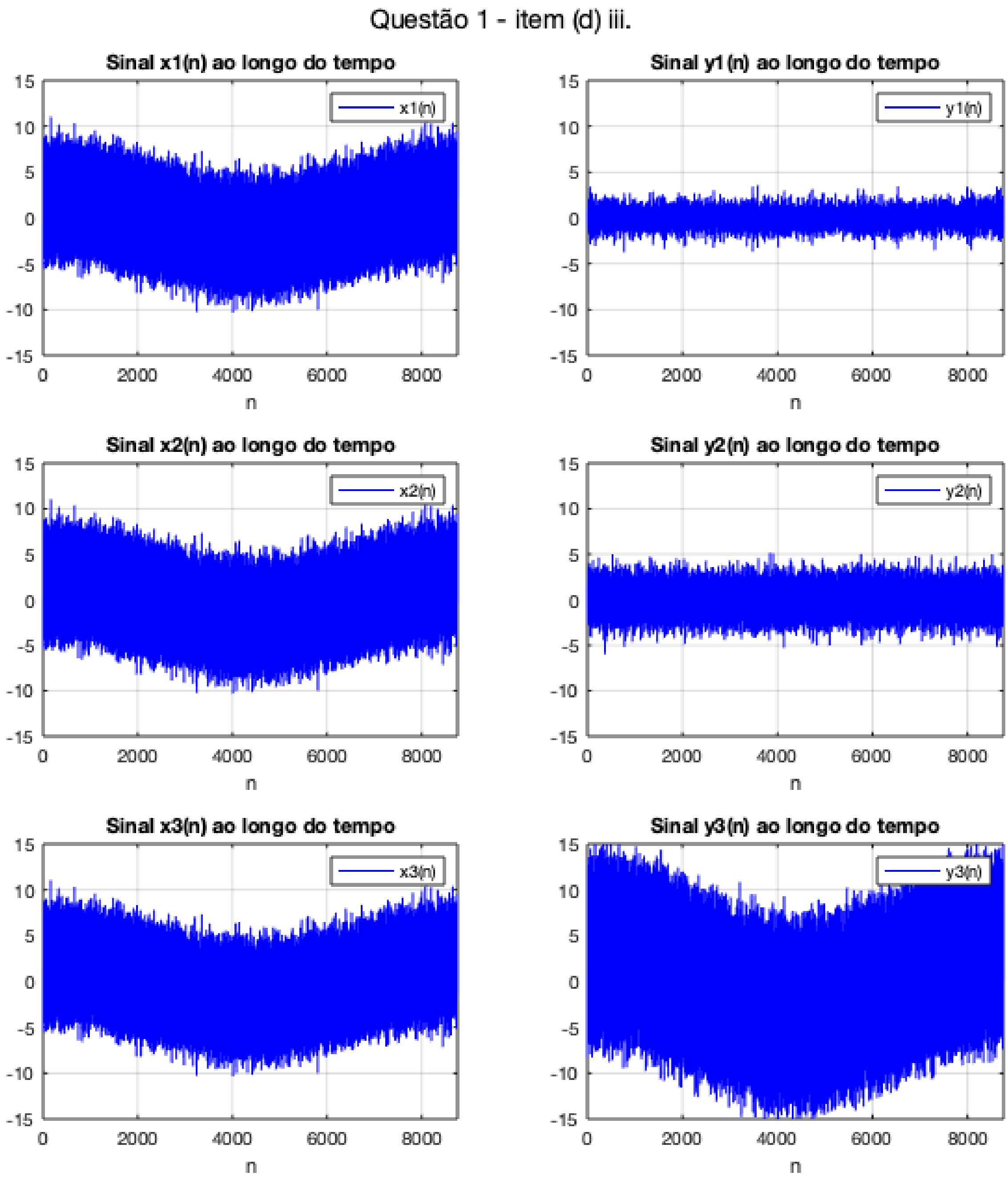

Figura 1. Sinais ao longo do tempo.

iv. Faça também diagramas de dispersão (isto é, *scatter plots*) de $x_1(n)$ por $x_2(n)$, $x_1(n)$ por $x_3(n)$ e $x_2(n)$ por $x_3(n)$. Em cada gráfico, identifique as direções relacionadas às componentes principais. Comente os resultados obtidos.
DICA: Como sugestão, veja os comandos do Exemplo 2 apresentados na aula de 07/12/2021.

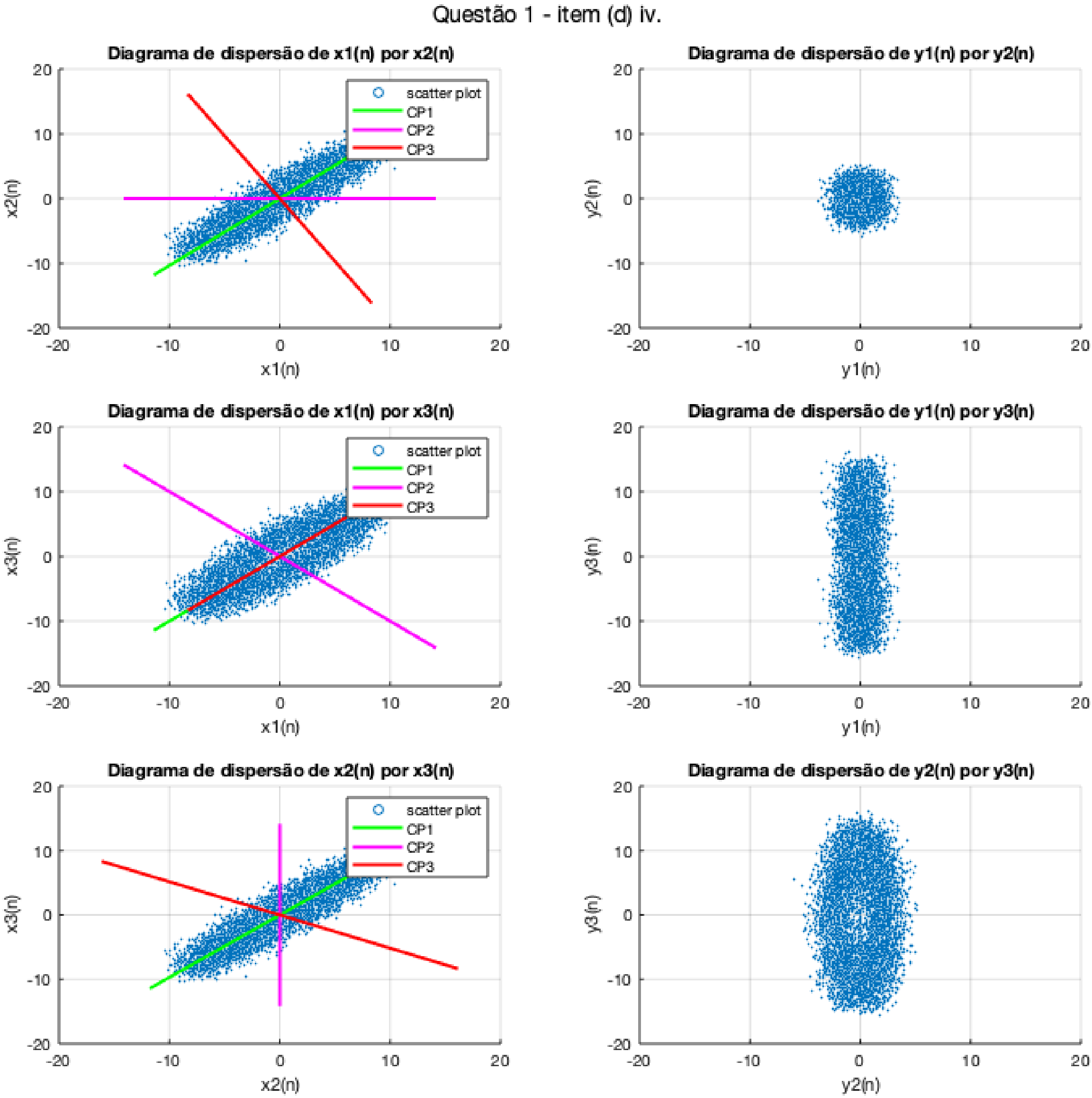

Figura 2. Diagramas de dispersão entre os sinais estudados e Componentes Principais.

Aplicando uma abordagem sistemática a interpretação das figuras, vamos usar:

$$\frac{\Delta y_1}{\Delta y_2} \rightarrow \text{Variação do } y_1 \text{ em relação ao } y_2;$$
$$\frac{\Delta y_1}{\Delta y_3} \rightarrow \text{Variação do } y_1 \text{ em relação ao } y_3;$$
$$\frac{\Delta y_2}{\Delta y_3} \rightarrow \text{Variação do } y_2 \text{ em relação ao } y_3.$$

Assim, é notável que:

$$\frac{\Delta y_1}{\Delta y_2} > \frac{\Delta y_2}{\Delta y_3} > \frac{\Delta y_1}{\Delta y_3}$$

De posse desta verificação e das Variâncias calculadas anteriormente, sendo:

$$Variância\ 1 = 0{,}9975 \text{ e } Proporção\ da\ Variância\ Explicada\ 1 = 1{,}5833\%,$$
$$Variância\ 2 = 3{,}4100 \text{ e } Proporção\ da\ Variância\ Explicada\ 2 = 5{,}4127\%,$$
$$Variância\ 3 = 58{,}5925 \text{ e } Proporção\ da\ Variância\ Explicada\ 3 = 93{,}0040\%.$$

Pode-se afirmar:

- A maior parte da Variância do sinal x está na direção do vetor 3 – Vide Proporção da Variância Explicada acima;
- O descarte do sinal y_1 se mostra a melhor solução quando otimizações são necessárias.

v. Faça gráficos com histogramas de $x_1(n)$, $x_2(n)$, $x_3(n)$, $y_1(n)$, $y_2(n)$ e $y_3(n)$. Relacione esses histogramas com os componentes principais e os diagramas de dispersão obtidos no item iv. Comente os resultados obtidos.

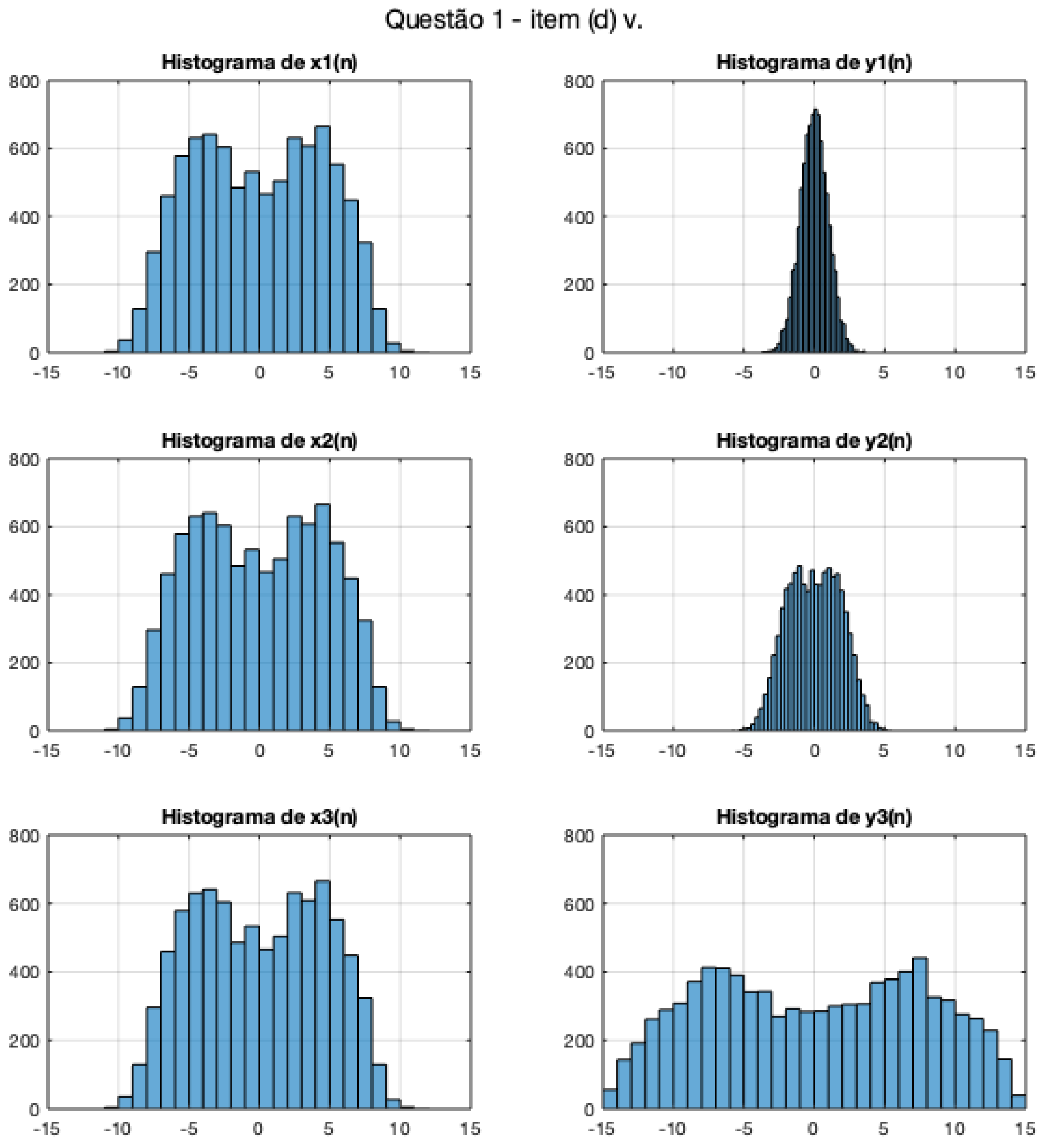

Figura 3. Histograma dos sinais estudados.

Claramente, confrontando a *Figura 2* com a *Figura 3*, nota-se que as distribuições de todos os sinais x são aproximadamente iguais (eles se correlacionam, analisando o plot da *Figura 2 – Diagrama de dispersão*) o que não acontece para y, uma vez que as distribuições desses apresentam maiores divergências entre si (y_1 em particular). Assim, com os Diagramas de dispersão e analisando os efeitos e resultados entre as componentes principais, suas relações ficam obviamente claras se adicionarmos à análise os histogramas com as distribuições que se mostraram na *Figura 3*.

Questão 2 (2,0 pontos)

Para responder a essa questão, considere as amostras da sequência $\boldsymbol{x1}$ do arquivo **dadosAP2.mat** disponível na página da disciplina. Deseja-se obter uma expressão analítica para a sequência $\boldsymbol{x1}$. Pedem-se:

(a) Implemente no MATLAB (ou *software* equivalente) o periodograma com a janela retangular. Considere o comprimento da janela igual ao número total de amostras do arquivo de dados fornecido. Esboce a densidade espectral de potência da sequência $\boldsymbol{x1}$. Não esqueça de incluir os programas usados para esboçar a densidade espectral de potência.

De posse das especificações apresentadas, se aplicou no MATLAB (*breve amostra do script.m que segue em anexo*)

```
...
N=size(x1, 2)
L=size(x1, 2)    %Comprimento da Janela - Pede-se o número de amostras do sinal
K=1              %Existira apenas um seguimento, das especificações e item anterior
Nf=N
Nf2=round(Nf/2)
...
```

obtendo-se o esboço da DEP (densidade espectral de potência) de $x_1(n)$ solicitado:

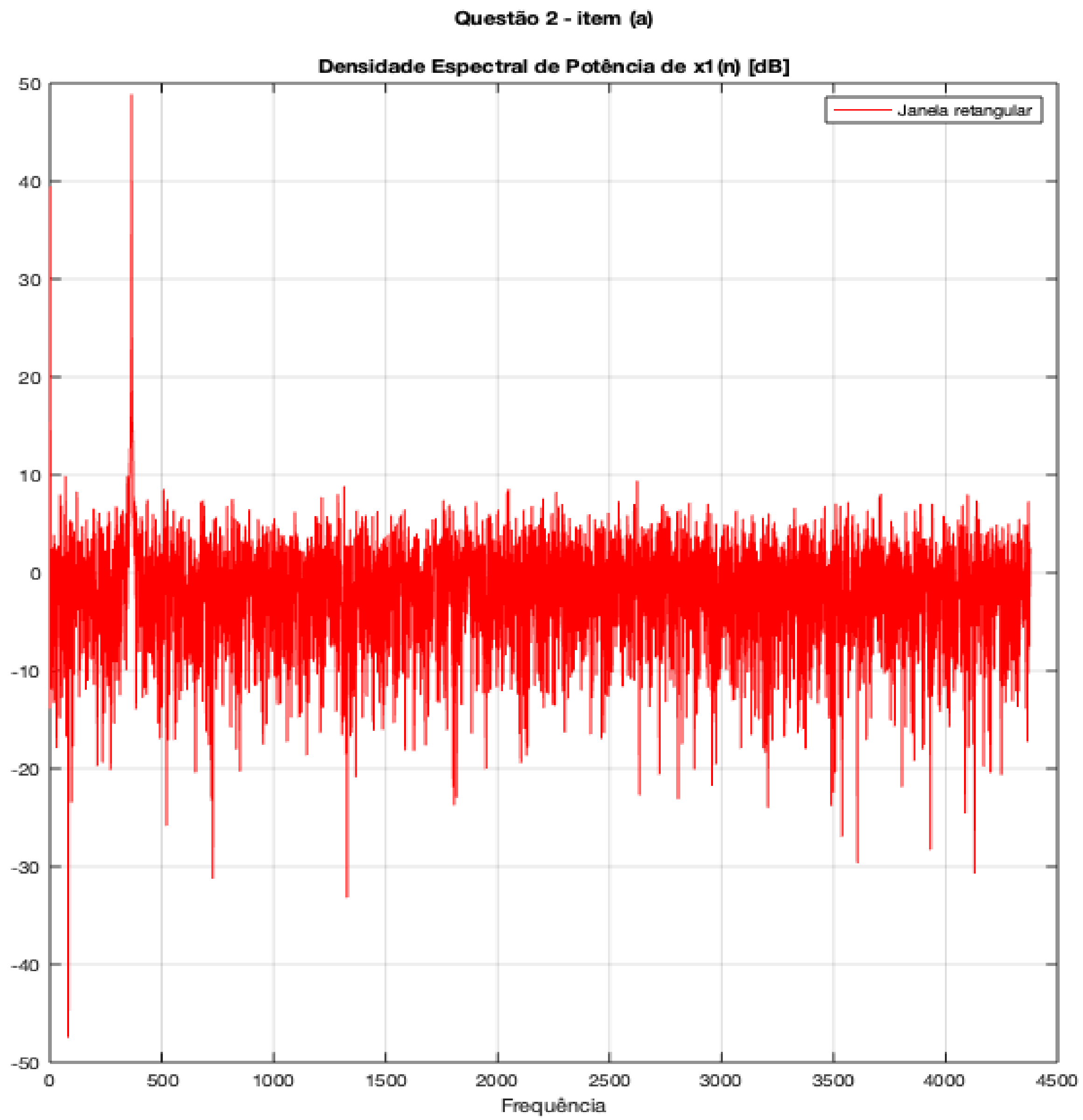

Figura 4. DEP - x1(n)

(b) A partir do esboço, forneça uma expressão analítica para a sequência x_1 com o valor dos coeficientes e especificação do ruído adicionado ao sinal. Justifique adequadamente os valores e a expressão fornecida.

Após processar e avaliar o esboço anterior, notaram-se duas senoides (duas raias) em seu espectro. Assim, com o MATLAB, através da rotina (que segue anexada em sua integridade)

```
...
varR1E = mean(Ipm1(1000:Nf/2))                  %Variância – ruído a partir de 1000
[v11,v12] = max(abs(Ipm1))                      %Intermediário 1 – freq. e amplitude
[v21,v22] = max(abs(Ipm1(Ipm1<max(Ipm1))))      %Intermediário 2 – freq. e amplitude
w1sinalEstimado1 = 2*pi*v12/Nf.                 %Freq. senoide 1
w2sinalEstimado1 = 2*pi*v22/Nf                  %Freq. senoide 2
A1 = sqrt((v11-varR1E)*4*U1*L/(Um1))            %Amplitude senoide 1
A2 = sqrt((v21-varR1E)*4*U1*L/(Um1))            %Amplitude senoide 2
...
```

se chegou aos principais valores que compõe a expressão analítica para o sinal $x_1(n)$, sendo:

$$x_1(n) = 5{,}92 * \cos(0{,}263 * n + \theta_1) + 2{,}00 * \cos(0{,}001 * n + \theta_2) + v(n),$$

com

$$v(n) \rightarrow ruído\ com\ variância\ unitária, \text{ e } \sigma_{x1}^2 \cong 0{,}9952.$$

(c) Na sua opinião, é possível melhorar a estimativa dos parâmetros da expressão analítica para a sequência $\boldsymbol{x_1}$? Justifique adequadamente sua resposta.

Penso que a estimativa feita através das fórmulas sofreriam melhorias caso se use algum outro método de estimação espectral para os parâmetros, como por exemplo, como visto em aula, *Pisarenko*, ou, também, usando um tipo de Janelamento (*Blackman*) mais preciso e/ou comprimentos diferentes.