

IA888 - Análise de Sinais e de Sistemas Lineares

Ortogonalização

Prof. Ricardo C.L.F. Oliveira

Faculdade de Engenharia Elétrica e de Computação
Universidade Estadual de Campinas

2º Semestre 2014

Suponha que se quer aproximar o sinal $y(t)$ por uma **combinação linear** de funções $f_k(t)$

$$y(t) \approx \sum_{k=1}^n \alpha_k f_k(t)$$

A dimensão do espaço \mathcal{S} gerado pela combinação linear das funções $f_k(t)$ é n se as funções $f_k(t)$ forem **linearmente independentes** entre si, isto é, as n funções $f_k(t)$ formam uma base de representação do espaço \mathcal{S} .

Assim, trata-se de encontrar os valores dos coeficientes α_k que **minimizem o erro**

$$\varepsilon(t) = y(t) - \sum_{k=1}^n \alpha_k f_k(t) = y(t) - \alpha' f(t)$$

sendo $f(t)$ o vetor coluna das funções do tempo $f_k(t)$ e $\alpha \in \mathbb{R}^n$ o vetor coluna de coeficientes. O valor quadrático do erro pode ser calculado por

$$\varepsilon^2(t) = (y(t) - \alpha' f(t))(y(t) - \alpha' f(t))' = y^2(t) + \alpha' f(t) f(t)' \alpha - 2\alpha' (f(t) y(t))$$

Ortogonalização – Aproximação por combinação linear de funções

II

sendo $f(t)f(t)'$ uma matriz no $\mathbb{R}^{n \times n}$ na qual cada componente é uma função do tempo resultante do produto dois a dois das funções $f_k(t)$ e $f(t)y(t)$ um vetor coluna no \mathbb{R}^n com elementos dados pelos produtos $f_k(t)y(t)$.

Calculando-se a média temporal no intervalo no qual deseja-se a aproximação de $y(t)$, tem-se

$$\langle \varepsilon^2(t) \rangle = \langle y^2(t) \rangle + \alpha' \langle f(t)f(t)' \rangle \alpha - 2\alpha' \langle f(t)y(t) \rangle$$

A matriz $R = \langle f(t)f(t)' \rangle \in \mathbb{R}^{n \times n}$ de correlação temporal das funções $f_k(t)$ é computada como $R = [r_{k\ell}]$ sendo $r_{k\ell}$ o produto escalar das funções $f_k(t)$ e $f_\ell(t)$, isto é,

$$r_{k\ell} = \langle f_k(t)f_\ell(t) \rangle = \int_{-\infty}^{+\infty} f_k(t)f_\ell(t)dt \quad , \quad k, \ell = 1, 2, \dots, n$$

Propriedade 1 (Critério de Gram)

Se as funções $f_k(t)$ forem linearmente independentes entre si, a matriz R será, por construção, uma matriz definida positiva. R é portanto não-singular, isto é, pode ser invertida, pois

$$v' \langle f(t)f(t)' \rangle v = \langle v' f(t)f(t)' v \rangle = \langle (f(t)' v)' (f(t)' v) \rangle = \langle \beta^2(t) \rangle$$

com $\beta(t) = f(t)' v$.

Como as funções $f_k(t)$ são linearmente independentes, $\beta(t) = 0$ se e somente se $v = 0$. Portanto,

$$v' \langle f(t)f(t)' \rangle v > 0, \quad \forall v \neq 0$$

Propriedade 2 (Erro médio quadrático)

O erro médio quadrático pode ser escrito como

$$\langle \varepsilon^2(t) \rangle = \langle y^2(t) \rangle + \alpha' R \alpha - 2\alpha' \langle f(t)y(t) \rangle$$

cujo valor mínimo é obtido para α solução de

$$\frac{d}{d\alpha} \langle \varepsilon^2(t) \rangle = 0 \Rightarrow 2R\alpha - 2\langle f(t)y(t) \rangle = 0 \Rightarrow \alpha = R^{-1} \langle f(t)y(t) \rangle \quad (1)$$

Exemplo 1.1

Considere os sinais linearmente independentes $f_1(t)$ e $f_2(t)$ (mostrados na Figura 1) dados por

$$f_1(t) = 2G_1(t - 0.5) \quad , \quad f_2(t) = (3t + 1)G_1(t - 0.5)$$

Exemplo II

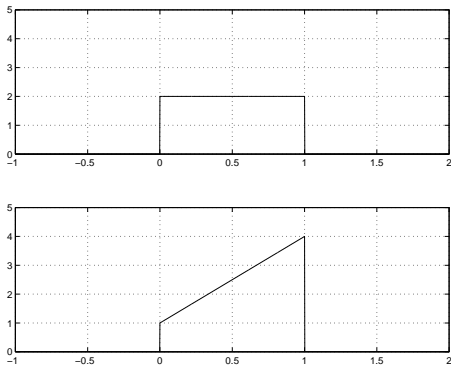


Figura : Funções $f_1(t) = 2G_1(t - 0.5)$ (acima) e $f_2(t) = (3t + 1)G_1(t - 0.5)$ (abaixo).

Exemplo III

A matriz de correlação R é dada por

$$R = \langle f(t)f(t)' \rangle = \begin{bmatrix} 4 & 5 \\ 5 & 7 \end{bmatrix} \Rightarrow R^{-1} = \frac{1}{3} \begin{bmatrix} 7 & -5 \\ -5 & 4 \end{bmatrix}$$

Os sinais $f_1(t)$ e $f_2(t)$ podem ser usados para aproximar funções no intervalo $[0, 1]$. Por exemplo, as funções $x_1(t)$, $x_2(t)$ e $x_3(t)$ e suas aproximações são dadas por

$$x_1(t) = 2 - t \approx \begin{bmatrix} 1.1667 & -0.3333 \end{bmatrix} f(t)$$

$$x_2(t) = \sinh(t) \approx \begin{bmatrix} -0.2102 & 0.3854 \end{bmatrix} f(t)$$

$$x_3(t) = \cosh(t) \approx \begin{bmatrix} 0.3651 & 0.1781 \end{bmatrix} f(t)$$

A Figura 2 mostra os sinais originais (pontilhados) e as aproximações. Observe que $x_1(t)$ é linearmente dependente de $f_1(t)$ e $f_2(t)$ e portanto o erro na aproximação é nulo. Os sinais $\sinh(t)$ e $\cosh(t)$ não são linearmente dependentes das funções $f_1(t)$ e $f_2(t)$, mas puderam ser aproximados com erro pequeno no intervalo considerado.

Exemplo IV

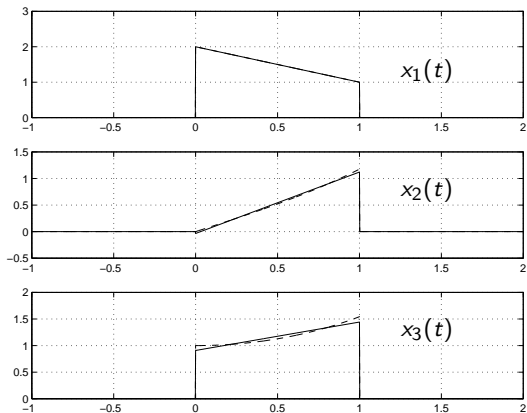


Figura : Funções $x_1(t) = 2 - t$, $x_2(t) = \sinh(t)$ e $x_3(t) = \cosh(t)$.

Os coeficientes α , obtidos pela expressão (1), determinam a aproximação com **erro quadrático mínimo** do sinal $y(t)$ por uma combinação linear das funções linearmente independentes $f_k(t)$.

Se as funções $f_k(t)$ forem ortogonais entre si, R será uma matriz diagonal, resultando no cálculo desacoplado dos coeficientes α_k .

Analisa-se a seguir a projeção de sinais em uma base ortogonal com dois propósitos: explicitar o desacoplamento no cálculo dos coeficientes de projeção e apresentar o algoritmo de ortogonalização de Gram-Schmidt.¹

¹Erhard Schmidt, alemão (1876-1959).

Projeção ortogonal I

Suponha que se quer aproximar o sinal $y(t)$ por uma combinação linear de funções ortogonais $g_k(t)$.

$$y(t) \approx \sum_{k=1}^n c_k g_k(t)$$

O sinal erro é dado por

$$\varepsilon(t) = y(t) - \sum_{k=1}^n c_k g_k(t)$$

resultando em

$$\langle \varepsilon^2(t) \rangle = \langle y^2(t) \rangle + \sum_{k=1}^n c_k^2 \langle g_k^2(t) \rangle - 2 \sum_{k=1}^n c_k \langle y(t) g_k(t) \rangle + \underbrace{\sum_{k=1}^n \sum_{\ell=1}^n c_k c_\ell}_{k \neq \ell} \underbrace{\langle g_k(t) g_\ell(t) \rangle}_{= 0, \text{ ortogonais}}$$

Note que $\langle \varepsilon^2(t) \rangle$ é uma função quadrática estritamente convexa nos coeficientes c_k e, portanto, possui um mínimo global.

$$\frac{\partial}{\partial c_k} \langle \varepsilon^2(t) \rangle = 0 \implies c_k = \frac{\langle y(t)g_k(t) \rangle}{\langle g_k^2(t) \rangle} \quad ; \quad k = 1, 2, \dots, n$$

Observe que o cálculo de cada coeficiente c_k é desacoplado do cálculo dos demais coeficientes, propriedade que deriva diretamente da hipótese de ortogonalidade das funções $g_k(t)$ da base.

Um subproduto importante é que o erro $\varepsilon(t)$ é ortogonal a todos os elementos da base.

$$\langle \varepsilon(t)g_k(t) \rangle = \langle y(t)g_k(t) \rangle - \sum_{\ell=1}^n c_\ell \langle g_\ell(t)g_k(t) \rangle = \langle y(t)g_k(t) \rangle - c_k \langle g_k^2(t) \rangle = 0$$

Note que, impondo $\langle \varepsilon(t)g_k(t) \rangle = 0$ a priori, obtêm-se diretamente os coeficientes c_k .

Exemplo 1.2

Considere os sinais ortogonais

$$x_1(t) = G_2(t) \quad , \quad x_2(t) = tG_2(t)$$

O sinal $x(t)$ dado por

$$x(t) = t^2 G_2(t)$$

pode ser aproximado por

$$x(t) \approx a_1 x_1(t) + a_2 x_2(t) \quad \Rightarrow \quad \varepsilon(t) = x(t) - a_1 x_1(t) - a_2 x_2(t)$$

Portanto,

$$\begin{aligned} \langle \varepsilon^2(t) \rangle = & \langle x^2(t) \rangle + a_1^2 \langle x_1^2(t) \rangle + a_2^2 \langle x_2^2(t) \rangle - 2a_1 \langle x_1(t)x(t) \rangle - \\ & 2a_2 \langle x_2(t)x(t) \rangle + 2a_1 a_2 \langle x_1(t)x_2(t) \rangle \end{aligned}$$

Exemplo II

As condições

$$\frac{\partial}{\partial a_1} \langle \varepsilon^2(t) \rangle = 0 \quad , \quad \frac{\partial}{\partial a_2} \langle \varepsilon^2(t) \rangle = 0$$

implicam

$$\begin{bmatrix} \langle x_1^2(t) \rangle & \langle x_1(t)x_2(t) \rangle \\ \langle x_2(t)x_1(t) \rangle & \langle x_2^2(t) \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \langle x_1(t)x(t) \rangle \\ \langle x_2(t)x(t) \rangle \end{bmatrix}$$

Como $x_1(t)$ e $x_2(t)$ são ortogonais, tem-se

$$\langle x_1^2(t) \rangle = 2 \quad , \quad \langle x_2^2(t) \rangle = \frac{2}{3} \quad , \quad \langle x_1(t)x(t) \rangle = \frac{2}{3} \quad , \quad \langle x_2(t)x(t) \rangle = 0$$

$$a_1 = \frac{\langle x_1(t)x(t) \rangle}{\langle x_1^2(t) \rangle} = \frac{1}{3} \quad , \quad a_2 = \frac{\langle x_2(t)x(t) \rangle}{\langle x_2^2(t) \rangle} = 0$$

$$\varepsilon(t) = \left(t^2 - \frac{1}{3} \right) G_2(t)$$

Note que o erro $\varepsilon(t)$ é ortogonal a $x_1(t)$ e $x_2(t)$.

Exemplo 1.3

Considere os sinais

$$x_1(t) = G_1(t-0.5) \quad , \quad x_2(t) = tG_1(t-0.5)$$

O sinal $x(t)$ dado por

$$x(t) = t^2 G_1(t-0.5)$$

pode ser aproximado por

$$x(t) \approx a_1 x_1(t) + a_2 x_2(t) \quad \Rightarrow \quad \varepsilon(t) = x(t) - a_1 x_1(t) - a_2 x_2(t)$$

As condições de mínimo implicam

$$\begin{bmatrix} \langle x_1^2(t) \rangle & \langle x_1(t)x_2(t) \rangle \\ \langle x_2(t)x_1(t) \rangle & \langle x_2^2(t) \rangle \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1/3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 0.25 \end{bmatrix}$$

$$a_1 = -\frac{1}{6} \quad , \quad a_2 = 1$$

Note que, por $x_1(t)$ e $x_2(t)$ não serem ortogonais, foi necessário resolver numericamente um sistema linear de equações. O erro, ortogonal a $x_1(t)$ e $x_2(t)$, é dado por

$$\varepsilon(t) = \left(t^2 - t + \frac{1}{6} \right) G_1(t - 0.5)$$

A Figura 3 mostra os sinais $x(t)$, $x_1(t)$, $x_2(t)$ e o erro $\varepsilon(t)$. Observe que, como $x_1(t)$ é constante no intervalo, a média de $\varepsilon(t)$ é nula.

Exemplo III

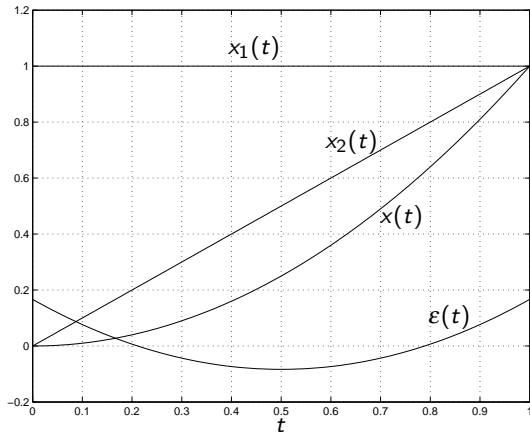


Figura : Sinais $x(t)$, $x_1(t)$, $x_2(t)$ e $\epsilon(t)$.

Discutem-se, a seguir, os procedimentos para se conseguir uma base ortogonal a partir de um conjunto dado de sinais.

Propriedade 3 (Ortogonalização de Gram-Schmidt)

Uma base ortogonal $g_k(t)$ pode ser obtida a partir de um conjunto de funções $f_k(t)$ pelo procedimento a seguir, explorando o fato de o erro de projeção ser sempre ortogonal aos elementos da base.

$$g_1(t) = f_1(t) ; \quad g_k(t) = f_k(t) - \sum_{\ell=1}^{k-1} \frac{\langle f_k(t)g_\ell(t) \rangle}{\langle g_\ell^2(t) \rangle} g_\ell(t) \quad , \quad k = 2, \dots, n$$

Note que $g_2(t)$ é o erro da projeção de $f_2(t)$ sobre $g_1(t)$, $g_3(t)$ é o erro da projeção de $f_3(t)$ sobre $g_1(t)$ e $g_2(t)$ e assim por diante.

A dimensão da base será igual ao número de funções linearmente independentes do conjunto $f_k(t)$.

Exemplo 1.4

Considere as funções $f_1(t)$, $f_2(t)$ e $f_3(t)$ mostradas na Figura 4. Observe que as funções são linearmente independentes, mas não são ortogonais, pois

$$\langle f_1(t)f_2(t) \rangle \neq 0 \quad , \quad \langle f_2(t)f_3(t) \rangle \neq 0$$

Exemplo II

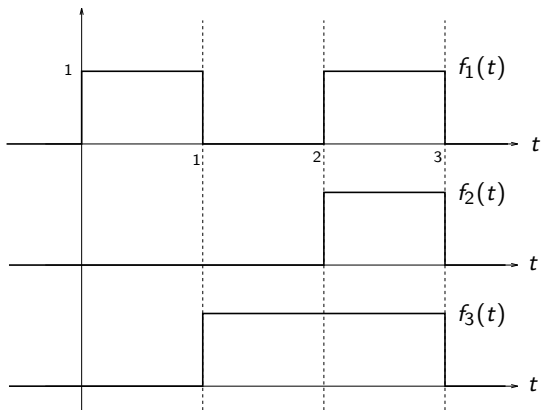


Figura : Funções $f_1(t)$, $f_2(t)$ e $f_3(t)$.

Realizando-se a ortogonalização de Gram-Schmidt, tem-se

$$g_1(t) = f_1(t)$$

$$g_2(t) = f_2(t) - \frac{\langle f_2(t)g_1(t) \rangle}{\langle g_1^2(t) \rangle} g_1(t) = f_2(t) - \frac{1}{2} g_1(t)$$

$$g_3(t) = f_3(t) - \frac{\langle f_3(t)g_1(t) \rangle}{\langle g_1^2(t) \rangle} g_1(t) - \frac{\langle f_3(t)g_2(t) \rangle}{\langle g_2^2(t) \rangle} g_2(t) = f_3 - \frac{1}{2} g_1(t) - \frac{1/2}{1/2} g_2(t)$$

As funções $g_1(t)$, $g_2(t)$ e $g_3(t)$, ortogonais entre si, são mostradas na Figura 5.

Exemplo IV

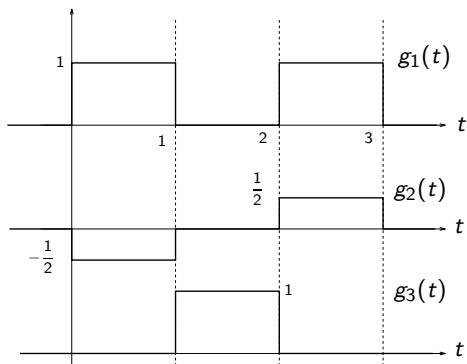


Figura : Funções ortogonais $g_1(t)$, $g_2(t)$ e $g_3(t)$.

Exemplo 1.5

Considere o sinal $x(t)$ mostrado na Figura 6, cuja energia (isto é, o valor da integral do módulo do sinal ao quadrado) é igual a 3. O sinal $x(t)$ pode ser escrito na base $g_1(t)$, $g_2(t)$ e $g_3(t)$, resultando nos coeficientes de projeção dados por

$$\langle x(t)g_1(t) \rangle = 2 \quad , \quad \langle x(t)g_2(t) \rangle = 0 \quad , \quad \langle x(t)g_3(t) \rangle = -1$$

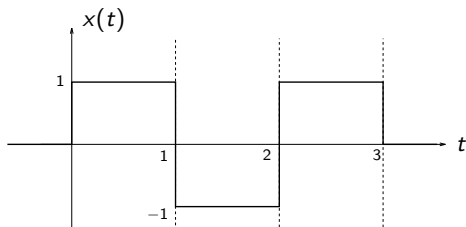


Figura : Sinal $x(t)$ (energia igual a 3).

Portanto,

$$x(t) = \frac{2}{2} g_1(t) + \frac{0}{1/2} g_2(t) + \frac{-1}{1} g_3(t) \quad \Rightarrow \quad x(t) = g_1(t) - g_3(t)$$

Exemplo 1.6

Considere o conjunto de quatro sinais linearmente independentes $f_1(t)$, $f_2(t)$, $f_3(t)$ e $f_4(t)$, nulos fora do intervalo $[0, 1]$.

$$f_1(t) = 2 \quad , \quad f_2(t) = 3t + 1 \quad , \quad f_3(t) = \text{sen}(2\pi t) \quad , \quad f_4(t) = \text{cos}(2\pi t)$$

Aplicando-se o algoritmo de Gram-Schmidt obtêm-se os sinais $g_1(t)$, $g_2(t)$, $g_3(t)$ e $g_4(t)$, mostrados na Figura 7. Observe que, por construção, $g_1(t) = f_1(t)$, enquanto que $g_2(t)$ é alterado para ficar ortogonal a $g_1(t)$. O sinal $f_3(t)$, que dá origem a $g_3(t)$, é alterado apenas por $g_2(t)$, pois já era ortogonal a $g_1(t)$. O sinal $g_4(t)$ é igual a $f_4(t)$, pois já era ortogonal aos três anteriores.

Exemplo II

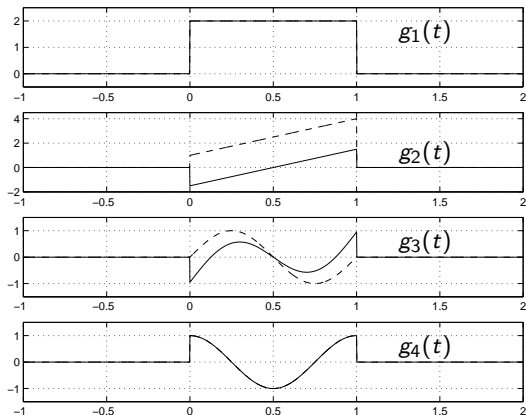


Figura : Sinais $g_1(t)$, $g_2(t)$, $g_3(t)$ e $g_4(t)$ resultantes da ortogonalização de Gram-Schmidt (sinais originais em pontilhado).

O algoritmo de Gram-Schmidt pode ser formulado como o resultado de um problema de triangularização da matriz $R = \langle f(t)f(t)' \rangle$ de correlação temporal das funções $f_k(t)$.

Um conjunto de funções $f_k(t)$ gera, por combinação linear, um espaço \mathcal{S} . Se as n funções $f_k(t)$ forem linearmente independentes, o espaço \mathcal{S} tem dimensão n e $f(t)$ constitui uma base para \mathcal{S} (não necessariamente ortogonal).

Transformações lineares na forma

$$g(t) = Qf(t) \quad , \quad Q \text{ não singular}$$

preservam a representação do espaço \mathcal{S} , isto é, $g(t)$ constitui uma nova base para \mathcal{S} .

Assim, a orthogonalização pode ser definida em termos da escolha da matriz Q tal que

$$\langle g(t)g(t)' \rangle = \langle Qf(t)f(t)'Q' \rangle = QRQ' = \mathbf{I} \quad (2)$$

Ortogonalização de Gram-Schmidt como triangularização II

Note que $\langle g(t)g(t)' \rangle = \mathbf{I}$ impõe uma ortonormalização, que corresponde a um sistema quadrático de equações com n^2 variáveis e $n(n+1)/2$ restrições, indicando que há inúmeras maneiras de ortonormalizar um conjunto de funções linearmente independentes.

A orthogonalização de Gram-Schmidt equivale a uma escolha apropriada de Q triangular inferior, pois $g_1(t) = f_1(t)$, $g_2(t) = af_1(t) + bf_2(t)$, $g_3(t) = af_1(t) + bf_2(t) + cf_3(t)$ e assim por diante.

A transformação de Cholesky² aplicada à matriz R , simétrica e definida positiva, produz L triangular inferior que satisfaz $R = LL'$. Assim,

$$QRQ' = (QL)(QL)' = \mathbf{I}$$

Uma solução trivial, induzida pela decomposição de Cholesky, é dada por

$$Q = L^{-1}$$

Observe que a inversa de uma matriz triangular inferior é, por construção, uma matriz triangular inferior. Assim, a transformação de Cholesky permite obter de forma matricial a ortonormalização de Gram-Schmidt.

A fatorização de Schur³ aplicada à matriz R produz uma matriz unitária V e uma matriz diagonal Ω composta pelos autovalores de R tais que $R = V\Omega V'$, sugerindo como solução

$$Q = V\Omega^{-0.5} V' \triangleq R^{-0.5}$$

²Andre-Louis Cholesky, francês (1875-1918).

³Issai Schur, matemático russo (1875-1941) com atuação na Alemanha.

Exemplo 1.7

Considere os sinais gerados pelo deslocamento de um pulso triangular dados por

$$f_k(t) = \text{Tri}_{2T}(t - kT) ; k = 1, 2, \dots, 5$$

A função $\text{Tri}_{2T}(t)$ é mostrada na Figura 8. Os pulsos $f_k(t)$ não são ortogonais, pois

$$r_{k\ell} = \int_{-\infty}^{+\infty} \text{Tri}_{2T}(t - kT) \text{Tri}_{2T}(t - \ell T) dt ; k, \ell = 1, 2, \dots, 5$$

$$r_{k\ell} = \begin{cases} 2T/3 & ; k = \ell \\ T/6 & ; |k - \ell| = 1 \\ 0 & \text{fora} \end{cases} \Rightarrow R = \frac{T}{6} \begin{bmatrix} 4 & 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

Exemplo II

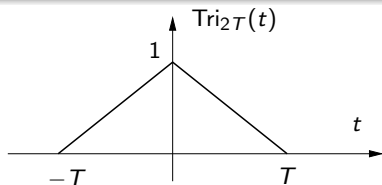


Figura : Função $\text{Tri}_{2T}(t)$.

Note que se as funções $f_k(t)$ fossem ortogonais entre si, a matriz R correspondente seria diagonal.

A aplicação da decomposição de Cholesky na matriz R para $T = 1.5$ dada por

$$R = \begin{bmatrix} 1 & 0.25 & 0 & 0 & 0 \\ 0.25 & 1 & 0.25 & 0 & 0 \\ 0 & 0.25 & 1 & 0.25 & 0 \\ 0 & 0 & 0.25 & 1 & 0.25 \\ 0 & 0 & 0 & 0.25 & 1 \end{bmatrix}$$

resulta na matriz Q

$$Q = \begin{bmatrix} +1.000 & 0 & 0 & 0 & 0 \\ -0.258 & +1.033 & 0 & 0 & 0 \\ +0.069 & -0.276 & +1.035 & 0 & 0 \\ -0.019 & +0.074 & -0.277 & +1.035 & 0 \\ +0.005 & -0.019 & +0.074 & -0.277 & +1.035 \end{bmatrix}$$

A transformação $g = Qf$ produz os sinais mostrados na Figura 9. Note que o primeiro elemento g_1 preservou a forma de f_1 , e os demais elementos foram sendo progressivamente alterados.

Exemplo IV

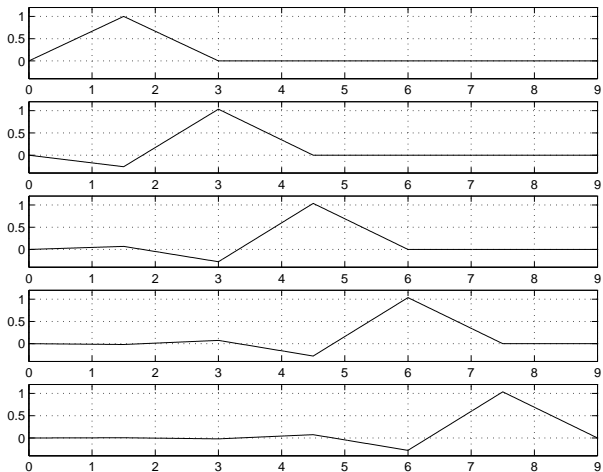


Figura : Sinais ortogonalizados por Gram-Schmidt.

Exemplo 1.8

Uma permutação na ordem das funções $f_k(t)$ do exemplo anterior produz resultados distintos (porém também ortogonais). Considere a seguinte ordem

$$f(t) = [f_1(t) \quad f_3(t) \quad f_5(t) \quad f_2(t) \quad f_4(t)]'$$

que resulta em

$$R = \begin{bmatrix} 1 & 0 & 0 & 0.25 & 0 \\ 0 & 1 & 0 & 0.25 & 0.25 \\ 0 & 0 & 1 & 0 & 0.25 \\ 0.25 & 0.25 & 0 & 1 & 0 \\ 0 & 0.25 & 0.25 & 0 & 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} +1.000 & 0 & 0 & 0 & 0 \\ 0 & +1.000 & 0 & 0 & 0 \\ 0 & 0 & +1.000 & 0 & 0 \\ -0.267 & -0.267 & 0 & +1.069 & 0 \\ -0.019 & -0.287 & -0.268 & +0.077 & +1.072 \end{bmatrix}$$

Exemplo II

A transformação $g = Qf$, com $Q = L^{-1}$ e $R = LL'$, produz os sinais mostrados na Figura 10.

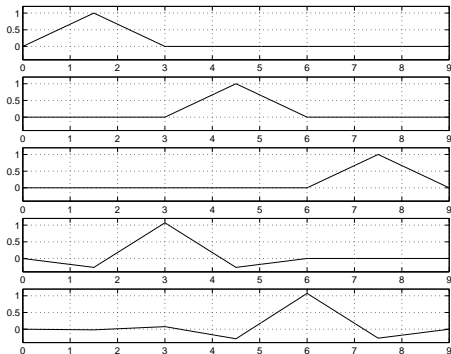


Figura : Sinais $\{f_1(t), f_3(t), f_5(t), f_2(t), f_4(t)\}$ ortogonalizados por Gram-Schmidt.

Observe que esse ordenamento implicou na alteração da forma das funções $f_2(t)$ e $f_4(t)$ e na preservação das funções $f_1(t)$, $f_3(t)$ e $f_5(t)$.