



Instituto de Física Teórica
Universidade Estadual Paulista

DISSERTAÇÃO DE MESTRADO

IFT–D.005/03

Determinismo e Estocasticidade em Séries Temporais Empíricas

Birajara Soares Machado

Orientador

Gerson Francisco

– São Paulo –
Fevereiro de 2003

Resumo

Neste trabalho procurou-se desenvolver e avaliar uma metodologia consistente com a teoria do caos capaz de classificar o mecanismo gerador de séries temporais empíricas de dados. Faz-se, para tal, uma descrição de testes quantitativos na caracterização de séries temporais, bem como de um método para redução de ruído. Por fim, aplica-se a metodologia proposta em sinais experimentais provenientes da atividade elétrica cerebral.

Palavras chave: análise de séries temporais; caos determinístico; redes neurais artificiais; transformada em ondeletas.

Áreas do conhecimento: 1.01.04.00–3; 1.02.02.08–0; 1.05.01.00–2.

Abstract

In this work we sought to develop and to evaluate a methodology consistent with the chaos theory capable to classify the generating mechanism of empirical time series. For such, we will make a description of quantitative tests in the characterization of time series, including a method for noise reduction. Finally, we will apply the methodology here proposed in experimental signals originating from the cerebral electric activity.

Key words: artificial neural network; deterministic chaos; time series analysis; wavelets transforms.

Capítulo 1

Determinismo *vs.* Estocasticidade

1.1 Introdução

Existe profunda diferença entre métodos para previsão de sistemas caóticos e processos estocásticos. Tal classificação é um passo decisivo para estabelecer estratégias de modelagem do sistema.

Tratando-se de caos determinístico, espera-se descrever a dinâmica por meio de equações diferenciais ou mapeamentos para sistemas discretos. Aqui, a modelagem de séries temporais associadas ao sistema dinâmico pode ser feita usando regressões não-lineares e não-paramétricas, onde redes neurais artificiais são casos particulares [1]. Processos estocásticos estão associados a um número muito grande de graus de liberdade; convém mencionar que caos de alta dimensionalidade pode ser interpretado como um processo estocástico, pois nesse caso a modelagem mais adequada deverá ser estocástica.

A questão básica do trabalho é determinar qual o mecanismo que gerou uma série temporal empírica. Evidentemente, uma resposta bem definida a esse problema será de imensa importância em praticamente todas as aplicações científicas nesse contexto. Essa classificação necessita tanto de estudos preliminares para estabelecer a estacionariedade da série temporal [2, 3, 4, 5], quanto de métodos para a redução de ruído [6, 7, 8, 9]. Na ausência desta análise preliminar, compromete-se a aplicação de algoritmos na determinação de parâmetros de diagnóstico úteis na classificação quanto ao mecanismo gerador.

No capítulo presente, inicialmente descreve-se de forma introdutória a teoria dos sistemas dinâmicos e, na sequência, dos processos estocásticos. Logo após apresentam-se três diferentes métodos para a detecção de determinismo em séries temporais baseados nas abordagens métrica, topológica e algorítmica, utilizados na literatura [4, 10, 11, 12, 13, 14].

O método métrico, que utiliza a noção de distância entre pontos do atrator

estranho, fornece medidas a partir da estrutura local deste atrator que são invariantes sob mudanças no sistema de coordenadas, mas que dependem do parâmetro de controle [11, 15]. O método topológico, que fornece informações sobre a organização do atrator estranho, caracteriza-se pelo estudo de propriedades das órbitas periódicas imersas no atrator que independem do sistema de coordenadas e também de variações no parâmetro de controle. A abordagem topológica, que não necessita de métrica, é aconselhável principalmente em sistemas caóticos de baixa dimensionalidade [15, 16]. O terceiro método apresentado baseia-se na teoria da complexidade de algoritmos utilizada na teoria da comunicação [17].

1.2 Sistemas Dinâmicos

Existem dois tipos de equações que geram um sistema dinâmico. O primeiro tipo é definido como

$$\dot{x} = f(t, x), \quad t \in \mathbb{R}, \quad (1.1)$$

onde $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ de classe C^1 e $x : S \subset \mathbb{R} \rightarrow \mathbb{R}^n$, sendo U e S conjuntos abertos em $\mathbb{R}^n$ e $\mathbb{R}$, respectivamente. A função f é um campo vetorial e a solução é uma curva tangente ao campo. Uma solução com condição inicial x_0 no instante t_0 pode ser escrita como $x(t; t_0, x_0) = x_t(t_0, x_0)$. Esta solução define o fluxo

$$\phi : U \times S \rightarrow \mathbb{R}^n, \quad (1.2)$$

onde $x_t(0, x_0) = \phi_t(x_0)$.

O segundo tipo de equação descreve sistemas discretos e pode ser definido como

$$x_{m+1} = g(x_m), \quad m \in \mathbb{Z}, \quad (1.3)$$

onde $g : M \subset \mathbb{R}^n \rightarrow M$ de classe C^1 e M é um conjunto compacto em $\mathbb{R}^n$.

Um sistema dinâmico diferenciável pode ser representado tanto em tempo discreto quanto em tempo contínuo. Fenômenos representados por séries temporais geralmente acontecem de forma contínua e deveriam ter uma representação contínua. Todavia, as respectivas séries temporais são obtidas com uma determinada frequência, atribuindo ao sistema uma representação discreta. Este problema é facilmente contornado na medida em que se pode discretizar processos contínuos por meio de um mapa de Poincaré. Seja Π uma hipersuperfície transversal $(n - 1)$ -dimensional do campo vetorial n -dimensional. Nos casos de interesse é possível encontrar um conjunto aberto $D \subset \Pi$ tal que as trajetórias partindo de D retornem a Π . Pode-se então definir o mapa de Poincaré P como $P : D \subset \Pi \rightarrow \Pi$.

Do ponto de vista geométrico, $\phi_t(x_0)$ é uma curva em $\mathbb{R}^n$ e a equação (1.1) fornece o vetor tangente a cada ponto. Campos vetoriais que não dependem explicitamente do tempo são chamados autônomos e aqueles que dependem explicitamente do tempo são chamados não-autônomos. Um campo não-autônomo pode ser transformado em autônomo mediante a introdução de mais uma dimensão no sistema.¹

Um campo vetorial autônomo é então definido como

$$\dot{x} = f(x), \quad (1.4)$$

onde $x \in U \subset \mathbb{R}^n$ e U é um conjunto aberto em $\mathbb{R}^n$.

Define-se ponto fixo (ou de equilíbrio) de (1.4) como $\bar{x} \in \mathbb{R}^n$ que satisfaz $f(\bar{x}) = 0$. Diz-se que $\bar{x}$ é assintoticamente estável se a resposta do sistema a uma pequena perturbação aproxima-se de $\bar{x}$ quando $t \rightarrow +\infty$ para condições iniciais próximas de $\bar{x}$. O ponto fixo $\bar{x}$ é nesse caso um atrator. O ponto fixo $\bar{x}$ é definido como estável (estabilidade neutra) se a resposta do sistema a uma pequena perturbação permanece pequena quando $t \rightarrow +\infty$. O ponto fixo $\bar{x}$ é instável se a perturbação cresce para $t \rightarrow +\infty$.

As definições de estabilidade são locais. Um ponto fixo é estável na medida em que atrai soluções com condições iniciais próximas dele, mas pode não atrair soluções com condições iniciais distantes. O conjunto de todas as possíveis condições iniciais que convergem para o mesmo ponto fixo $\bar{x}$ é chamado bacia de atração de $\bar{x}$.

Para a classificação quanto à estabilidade de um ponto fixo, usa-se o método de linearização. Seja um sistema não-linear descrito pela equação (1.4), no qual existe um ponto de equilíbrio $\bar{x}$. É possível escrever $\dot{x}$ em função de uma expansão em série de Taylor em torno do ponto $\bar{x}$, obtendo-se então

$$\dot{x} = f(\bar{x} + \xi) = f(\bar{x}) + \left. \frac{\partial f}{\partial x} \right|_{\bar{x}} \xi + \mathcal{O}(\xi^m), \quad \text{para } m \geq 2. \quad (1.5)$$

Desconsiderando os termos que são de ordem m , obtém-se

$$\dot{\xi} = \left. \frac{\partial f}{\partial x} \right|_{\bar{x}} \xi = J(\bar{x})\xi, \quad (1.6)$$

onde $J(\bar{x})$ é a matriz jacobiana do sistema calculada no ponto fixo $\bar{x}$ e $\|\xi\| \ll \|x\|$, para $\xi = x - \bar{x}$.

Deste modo, é necessário verificar a estabilidade do ponto fixo $\xi = 0$. A solução geral da equação (1.6) é escrita como

$$\xi(t) = e^{\lambda(\bar{x})t} \xi(0), \quad (1.7)$$

¹Transforma-se um campo vetorial não-autônomo em autônomo redefinindo o tempo como uma nova variável dependente, isto é, $t = y$, tal que $\dot{y} = 1$.

onde a função $\xi(t)$ é uma aproximação de primeira ordem de $\phi_t(x_0)$ e os autovalores $\lambda(\bar{x})$ são determinados resolvendo a equação secular

$$\det[J(\bar{x}) - \lambda(\bar{x})I_{n \times n}] = 0, \quad (1.8)$$

sendo $\xi(0)$ e $\lambda(\bar{x})$, respectivamente, os autovetores e autovalores da matriz $J(\bar{x})$ e $I_{n \times n}$ é a matriz identidade n -dimensional.

Os autovalores da matriz $J(\bar{x})$ podem ser reais ou complexos. No caso mais geral, para $\lambda(\bar{x}) = \Re(\bar{x}) + i\Im(\bar{x})$, escreve-se $\xi(t)$ como

$$\xi(t) = e^{\Re(\bar{x})t} e^{i\Im(\bar{x})t} \xi(0), \quad (1.9)$$

onde $\Re(\bar{x})$ e $\Im(\bar{x})$ representam, respectivamente, as partes real e imaginária do autovalor $\lambda(\bar{x})$.

Como $e^{i\Im(\bar{x})t}$ é uma função limitada, a estabilidade de $\xi(t)$ vai depender basicamente de $\Re(\bar{x})$. Se $\Re(\bar{x}) \neq 0$ para todos os autovalores da matriz jacobiana, define-se o equilíbrio no ponto fixo como hiperbólico. Quando $[\Re(\bar{x})]_k < 0$ para todo k tem-se estabilidade assintótica e $[\Re(\bar{x})]_k > 0$ para no mínimo um valor de k implica instabilidade. Se $\Re(\bar{x}) = 0$ o equilíbrio no ponto fixo é dito não hiperbólico.

Os subespaços lineares gerados pelos autovetores $\xi(0)$ da matriz $J(\bar{x})$ são definidos como subespaços invariantes. Seja $\bar{x} \in \mathbb{R}^n$ um ponto fixo do sistema não-linear (1.4). Considera-se o sistema linear (1.6) associado, onde $\xi \in \mathbb{R}^n$ e a matriz jacobiana $J(\bar{x})$ é uma matriz constante $n \times n$, representando-se $\mathbb{R}^n$ como uma soma direta de três subespaços invariantes, definidos como:

a) subespaço estável

$$E_{lin}^s = \text{span}\{e_1, e_2, \dots, e_s\};$$

b) subespaço instável

$$E_{lin}^u = \text{span}\{e_{s+1}, e_{s+2}, \dots, e_{s+u}\};$$

c) subespaço central

$$E_{lin}^c = \text{span}\{e_{s+u+1}, e_{s+u+2}, \dots, e_{s+u+c}\};$$

onde $s + u + c = n$, $\{e_1, \dots, e_s\}$ são os autovetores de $J(\bar{x})$ que correspondem aos autovalores com parte real negativa, $\{e_{s+1}, \dots, e_{s+u}\}$ são os autovetores de $J(\bar{x})$ que

correspondem aos autovalores com parte real positiva e $\{e_{s+u+1}, \dots, e_{s+u+c}\}$ são os autovetores de $J(\bar{x})$ que correspondem aos autovalores com parte real nula. As letras s , u e c indicam subespaços estáveis, instáveis e centrais, respectivamente. Estes subespaços são definidos como invariantes desde que a solução (1.7) com condição inicial $\xi(0)$ inteiramente contida em um particular subespaço invariante permaneça neste mesmo subespaço para todo tempo.

Agora é conveniente enunciar dois teoremas que relacionam $J(\bar{x})\xi$ com $f(x)$. O primeiro é o de Hartman-Grobman e o segundo é o da variedade invariante.

Teorema 1.1 (Hartman-Grobman) *Se a matriz $J(\bar{x})$ não possuir autovalores nulos ou puramente imaginários, então existe um homeomorfismo definido em alguma vizinhança U de $\bar{x} \in \mathbb{R}^n$, o qual leva localmente órbitas do fluxo não-linear $\phi_t(x_0)$ a órbitas do fluxo linear $e^{\lambda(\bar{x})t}\xi(0)$. Além disso, o homeomorfismo preserva o sentido das órbitas e pode ser escolhido para preservar a parametrização no tempo.*

Teorema 1.2 (Variedade Invariante) *Supõe-se que (1.4) possui um ponto de equilíbrio hiperbólico $\bar{x}$. Então, existem variedades invariantes locais estáveis e instáveis, $W_{loc}^s(\bar{x})$ e $W_{loc}^u(\bar{x})$, com as mesmas dimensões dos autoespaços E_{lin}^s e E_{lin}^u do sistema linearizado e tangentes a esses espaços em $\bar{x}$. Essas variedades são tão lisas quanto o campo f e definidas como*

$$\begin{aligned} W_{loc}^s(\bar{x}) &= \{x_0 \in U / \lim_{t \rightarrow +\infty} \phi_t(x_0) = \bar{x} \text{ e } \phi_t(x_0) \in U, t \geq 0\}, \\ W_{loc}^u(\bar{x}) &= \{x_0 \in U / \lim_{t \rightarrow -\infty} \phi_t(x_0) = \bar{x} \text{ e } \phi_t(x_0) \in U, t \leq 0\}. \end{aligned}$$

A variedade invariante, no caso linear, será um subespaço vetorial invariante de $\mathbb{R}^n$ ou, para o caso não-linear, uma superfície imersa em $\mathbb{R}^n$ na qual pode ser representada localmente em um gráfico. Agora será definido o conceito de conjunto invariante e, logo após, a condição para que ele seja uma variedade invariante.

Definição 1.1 *Seja $A \subset \mathbb{R}^n$ um conjunto. Então A é invariante sob o campo vetorial (1.4) se para qualquer $x_0 \in A$ a solução $\phi_t(x_0) \in A$, $\forall t \in \mathbb{R}$.*

Definição 1.2 *Um conjunto invariante $A \subset \mathbb{R}^n$ é definido como uma variedade invariante C^r ($r \geq 1$) se A tiver a estrutura de uma variedade diferenciável C^r .*

A caracterização da dinâmica caótica em um atrator² corresponde a uma divergência exponencial local das condições iniciais durante a evolução temporal de

²Um atrator é definido intuitivamente como um conjunto invariante para o qual órbitas próximas convergem a ele depois de um tempo suficientemente longo. Adiante apresenta-se uma definição mais precisa.

um sistema dinâmico dissipativo, mais conhecida como sensibilidade às condições iniciais. Considere duas condições iniciais diferentes, mas muito próximas, definidas como $x_j(0)$ e $x'_j(0) = x_j(0) + \delta x(0)$, para $\|\delta x(0)\| \ll \|x_j(0)\|$. Após um tempo t , a separação entre as duas trajetórias $x_j(t)$ e $x'_j(t)$ será dada por $\delta x_j(t) = x'_j(t) - x_j(t)$. Haverá dependência sensível às condições iniciais se

$$\delta x_j(t) \approx e^{\lambda_j t} \delta x(0) , \quad (1.10)$$

para $\lambda_j > 0$. Essa dependência resulta da não-linearidade presente no sistema, amplificando exponencialmente pequenas diferenças nas condições iniciais. O parâmetro λ_j é denominado expoente característico de Lyapunov. No instante t o elemento de hipervolume no espaço de fases é definido como

$$\delta V(t) = \prod_{j=1}^n \delta x_j(t) = [\delta x(0)]^n \prod_{j=1}^n e^{\lambda_j t} = \delta V(0) \exp \left(\sum_{j=1}^n \lambda_j t \right) .$$

Em sistemas dinâmicos dissipativos, o hipervolume no espaço de fases é contraído com a evolução no tempo, o que não é possível para sistemas conservativos, como consequência do teorema de Liouville.³ Seja $V \subset \mathbb{R}^n$ um pequeno hipervolume no espaço de fases limitado pela hipersuperfície ∂V . Verifica-se então como V varia durante um pequeno período de tempo. Usando a definição do fluxo de um campo, observa-se que esta variação é igual ao fluxo total do campo através de ∂V , tem-se

$$V(t + dt) - V(t) = \oint_{\partial V} dx \cdot \hat{n} d\sigma ,$$

onde $\hat{n}$ é a orientação de ∂V , $d\sigma$ é um elemento de hipersuperfície e $dt \ll t$. A taxa de variação temporal do hipervolume será então dada por

$$\dot{V} = \oint_{\partial V} \dot{x} \cdot \hat{n} d\sigma = \oint_{\partial V} f(x) \cdot \hat{n} d\sigma .$$

A integral de superfície pode ser reescrita de acordo com o teorema da divergência de Gauss, obtendo-se

$$\dot{V} = \int_V \nabla \cdot f(x) dx ,$$

lembrando que V define a região interna de ∂V e $\nabla \cdot f(x)$ é dado por

$$\nabla \cdot f(x) = \sum_{j=1}^n \frac{\partial \dot{x}_j}{\partial x_j} = \sum_{j=1}^n \lambda_j .$$

³O teorema de Liouville afirma que para sistemas conservativos o hipervolume no espaço de fases é conservado com a evolução no tempo.

Se $\sum_{j=1}^n \lambda_j = 0$, o sistema é definido como conservativo; se $\sum_{j=1}^n \lambda_j < 0$, que resulta $\delta V(t) < \delta V(0)$, o sistema é definido como dissipativo.

Em sistemas dinâmicos dissipativos, onde possivelmente observa-se caos determinístico, não há necessidade do hipervolume contrair-se em todas as direções no espaço de fases. Em determinadas direções o hipervolume pode contrair e em outras expandir, ocorrendo necessariamente instabilidade na evolução dinâmica. A única forma pela qual soluções contraem-se em uma direção e expandem-se em outra, permanecendo em uma região finita do espaço de fases, é por um processo de dobra [18]. A condição básica sobre a dimensão do espaço de fases para ocorrência deste processo é de que a dinâmica seja no mínimo tridimensional.⁴ Esta condição não se impõe para mapeamentos, porém o sistema contínuo que este mapa representa deve ter dinâmica tridimensional. Como resultado deste processo de dobra presente na evolução, normalmente o atrator associado ao sistema dissipativo possui um aspecto típico de objeto fractal, enfatizando sua estrutura auto-similar. Segundo Ruelle e Takens [19], um atrator é definido como estranho (caótico) quando a solução $\phi_t(x_0)$ depender sensivelmente das condições iniciais. Esta dependência implica que a posição de uma trajetória dentro de um atrator estranho não é previsível de maneira trivial, uma vez que pequenos desvios nas condições iniciais estão sempre presentes. Entretanto, existem também atratores caóticos que não são fractais, bem como atratores com dimensão fractal que não são caóticos [20].

Um fluxo contínuo autônomo pode ser escrito como

$$\dot{x} = f_{\mu}(x), \quad (1.11)$$

onde μ é um parâmetro capaz de impor condições ao sistema de forma arbitrária (parâmetro de controle). Existem valores do parâmetro μ no qual o campo vetorial é bem definido e o movimento, após um regime transiente, converge para uma região limitada do espaço de fases, indicando que o sistema seja definido como dissipativo. Mudando os valores de μ , a natureza do movimento no regime assintótico pode mudar. Este efeito é denominado bifurcação e os valores de μ que caracterizam essa mudança no regime assintótico são definidos como pontos de bifurcação. Formalmente, uma bifurcação está associada com a perda de estabilidade estrutural do sistema. Um sistema é estruturalmente estável se para qualquer perturbação suficientemente pequena das equações que o define, o fluxo resultante é topologicamente equivalente ao gerado pelas equações sem a perturbação. A função f_{μ} pode incluir

⁴Admite-se que seja possível obter um atrator estranho em duas dimensões. Nesse caso, um dos expoentes de Lyapunov é positivo. Ao longo da direção paralela ao fluxo o expoente associado é nulo. Então se teria $(0, +)$ como sinais para os expoentes, resultando $\sum_{i=1}^2 \lambda_i > 0$ e o elemento de volume no espaço de fases divergiria, o que não é possível.

vários parâmetros de controle. Mas, em geral, utiliza-se apenas um parâmetro, o que significa considerar apenas bifurcações com codimensão 1.⁵

Podem eventualmente ocorrer, para valores apropriados de μ , sistemas que apresentam uma convergência a atratores estranhos, onde há rigorosamente dependência sensível às condições iniciais. Naturalmente, faz-se necessário detectar a presença de caos em séries determinísticas. Para tanto, utiliza-se conceitos como expoentes característicos de Lyapunov e dimensão fractal, por exemplo. Estes conceitos derivam da análise estatística dos movimentos apresentados pelas séries temporais.

Uma teoria estatística sempre se refere a algum tipo de média, em particular, médias temporais. Conseqüentemente, deve-se estar seguro de que regimes transientes na evolução dinâmica foram superados. Uma vez finalizado esse regime, as órbitas de um sistema caótico tendem a um movimento assintótico na vizinhança do atrator estranho. Quanto maior for o tamanho da amostra, maior será a participação do comportamento assintótico em relação ao movimento transiente. O sistema deve encontrar-se em um estado estacionário ou próximo desse estado, situação em que os pontos representativos da evolução temporal do sistema estarão bastante próximos do atrator associado à dinâmica.

A teoria ergódica do caos estabelece que a média temporal equivale à média espacial no espaço de fases. O peso com o qual se mede a média espacial deve ser especificado por meio de uma medida invariante. Uma medida ρ é definida como invariante se obedecer a propriedade

$$\rho[\phi_{-t}(E)] = \rho(E), \quad t > 0,$$

onde E é um subconjunto de pontos do espaço de fases e o conjunto $\phi_{-t}(E)$ é obtido envolvendo-se cada ponto em E durante um tempo t . A medida $\rho(x)$ é definida como uma média temporal de diversas funções delta de Dirac, cada uma delas definida num ponto $\phi_t(x_0)$ visitado pelo sistema num dado instante, ou seja,

$$\rho(x) = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_0^\tau \delta[x - \phi_t(x_0)] dt.$$

Desta forma, uma medida invariante fornece a densidade de probabilidade dos pontos do atrator estranho. Uma medida é temporalmente invariante se representar a totalidade do atrator, e isto dá sentido ao uso da teoria ergódica. Tem-se então o teorema ergódico enunciado a seguir.

Teorema 1.3 (Teorema Ergódico) *Dada uma função contínua φ e um atrator com medida invariante ρ , a média temporal de φ sobre uma órbita originada de*

⁵Codimensão é o número mínimo de parâmetros que devem ser variados de maneira a se observar algum tipo de bifurcação.

uma condição inicial x_0 na bacia de atração é igual à média ponderada pela medida invariante, ou seja,

$$\int_E \varphi(x) d\rho(x) = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_0^\tau \int_E \varphi(x) \delta[x - \phi_t(x_0)] dx dt = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_0^\tau \varphi[\phi_t(x_0)] dt ,$$

onde $d\rho(x) = \rho(x)dx$. Se E é compacto e invariante com relação à dinâmica, então existe uma medida ergódica invariante sobre a dinâmica com suporte em E .

Analisar medidas invariantes é muito mais fácil e conveniente do que calcular médias temporais. A partir do estudo das propriedades estatísticas de medidas invariantes, procede-se à descrição geométrica e caracterização de atratores. Essa caracterização pode ser dinâmica (expoentes característicos de Lyapunov) ou estática (dimensão fractal). Para maiores detalhes ver Eckmann e Ruelle [21].

Para discutir o comportamento assintótico das soluções de um sistema dinâmico, cabem as seguintes definições:

Definição 1.3 Um ponto p é chamado não errante se, para qualquer vizinhança U de p , existir $t \neq 0$ tal que $\phi_t(U) \cap U \neq \{ \}$.

Por exemplo, pontos fixos e órbitas periódicas são não errantes. Seguem-se agora as definições de conjunto atrator e bacia de atração, para introduzir o conceito de atrator.

Definição 1.4 Um conjunto invariante fechado $A \subset \mathbb{R}^n$ é chamado conjunto atrator, se existe alguma vizinhança U de A , tal que $\forall x_0 \in U, \forall t \in \mathbb{R}^+, \phi_t(x_0) \in U$ e $\lim_{t \rightarrow +\infty} \phi_t(x_0) \in A$.

Definição 1.5 Seja $B = \cup_{t \leq 0} \phi_t(U)$, o conjunto U é tal que $\lim_{t \rightarrow +\infty} \phi_t(U) = A$, onde A é o conjunto atrator. Então, B é chamado bacia de atração de A .

Definição 1.6 Um conjunto fechado e conexo M é chamado trapping region, se $\phi_t(M) \subset M, \forall t \geq 0$, ou, equivalentemente, se o campo vetorial no contorno ∂M apontar para o interior de M . O conjunto atrator será dado por $\cap_{t > 0} \phi_t(M) = A$.

Necessariamente, prova-se que um sistema dinâmico possui atrator encontrando-se uma trapping region M no espaço de fases. Agora é possível estabelecer, em definitivo, o conceito de atrator.

Definição 1.7 Um conjunto invariante A é topologicamente transitivo se, dados dois conjuntos abertos quaisquer $\{S, U\} \subset A$, existe $t \in \mathbb{R}$, tal que $\phi_t(S) \cap U \neq \{ \}$.

Definição 1.8 *Um conjunto A é chamado atrator se A é um conjunto atrator e, além disto, for topologicamente transitivo.*

Se o atrator A satisfizer esta última definição e também apresentar sensibilidade às condições iniciais, então A é definido como um atrator estranho, caracterizando a dinâmica caótica do sistema. Como exemplo de um sistema dinâmico caótico, ilustra-se o sistema de Lorenz [22]. Este é um clássico sistema autônomo que apresenta um atrator estranho. Trata-se de um sistema não-linear com três equações diferenciais ordinárias de primeira ordem dado por

$$\begin{cases} \dot{x} = \sigma y - \sigma x \\ \dot{y} = \rho x - y - xz \\ \dot{z} = xy - \beta z \end{cases}, \quad (1.12)$$

com $(x, y, z) \in \mathbb{R}^3$ e os parâmetros σ , ρ e β definidos como positivos.

Com o interesse na previsão do tempo, o modelo foi criado para estudar o movimento convectivo descrito por uma massa de gás atmosférico que, quando estivesse perto do chão, deveria esquentar e subir e, quando estivesse na parte superior da atmosfera, deveria esfriar e descer. O modelo de Lorenz não representa o movimento real descrito pelos gases atmosféricos, mas sua notável contribuição na verificação do fenômeno de alta sensibilidade às condições iniciais não pôde ser ignorada. Foi observada uma divergência exponencial nas soluções do sistema (1.12) para condições iniciais próximas. Isso foi interpretado como evidência da dificuldade de se prever a evolução das soluções devido à sempre presente imprecisão no conhecimento das condições iniciais. As funções $x(t)$, $y(t)$ e $z(t)$ não representam coordenadas espaciais, elas possuem significados físicos precisos no modelo: $x(t)$ é proporcional à intensidade de convecção; $y(t)$ é proporcional à diferença de temperatura entre as correntes ascendente e descendente; e $z(t)$ é proporcional à distorção do perfil de temperatura vertical em relação a um perfil linear.

Os pontos fixos do sistema de Lorenz são $\bar{x}_1 = (0, 0, 0)$ e $\bar{x}_{2,3} = (\pm\kappa, \pm\kappa, \rho - 1)$, onde $\kappa = \sqrt{\beta(\rho - 1)}$. As matrizes jacobianas para o sistema de Lorenz calculadas nos três pontos fixos valem

$$J_{\bar{x}_1} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\beta \end{bmatrix}, \quad J_{\bar{x}_2} = \begin{bmatrix} -\sigma & \sigma & 0 \\ 1 & -1 & -\kappa \\ \kappa & \kappa & -\beta \end{bmatrix}, \quad J_{\bar{x}_3} = \begin{bmatrix} -\sigma & \sigma & 0 \\ 1 & -1 & \kappa \\ -\kappa & -\kappa & -\beta \end{bmatrix}.$$

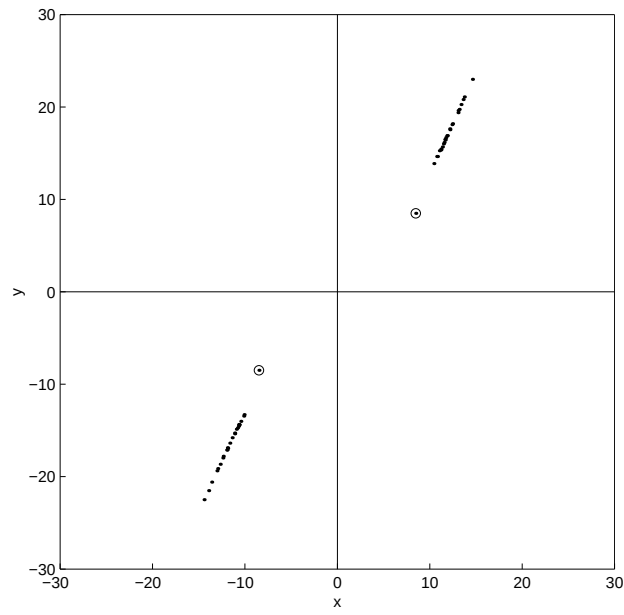


FIGURA 1.1: Seção de Poincaré para o atrator de Lorenz em $z = 27$. Os pontos fixos são indicados por $\odot$.

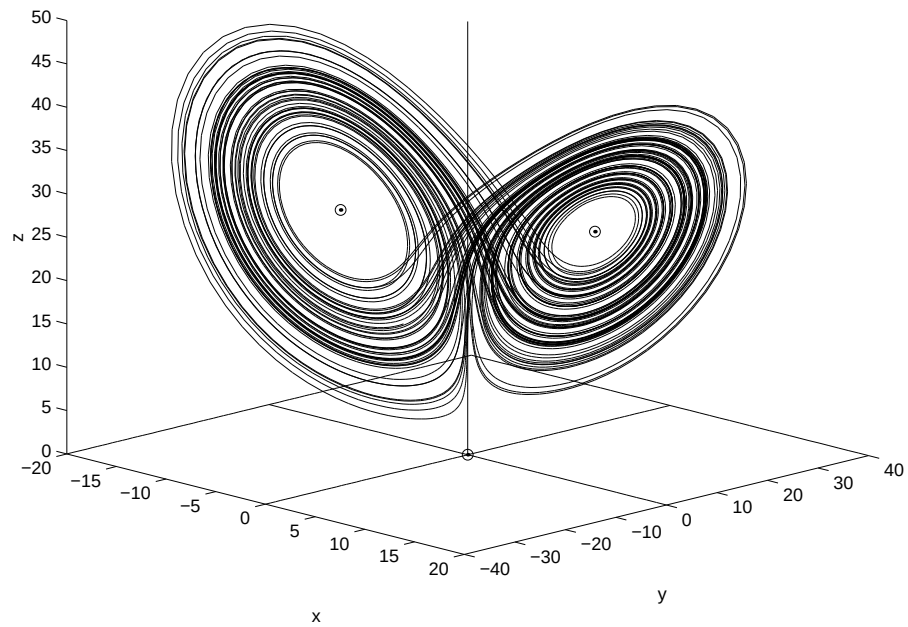


FIGURA 1.2: Atrator estranho de Lorenz para $\rho = 28$, $\sigma = 10$ e $\beta = 8/3$.

Os valores dos parâmetros utilizados por Lorenz são $\sigma = 10$ e $\beta = 8/3$. O parâmetro ρ (número de Rayleigh relativo) é tomado como um parâmetro de controle. O ponto fixo $\bar{x}_1$ é estável para $0 < \rho < 1$, pois todos os autovalores de $J_{\bar{x}_1}$ nesse domínio são negativos. Para os pontos fixos $\bar{x}_2$ e $\bar{x}_3$, as respectivas matrizes jacobianas apresentam os mesmos autovalores. Ainda no domínio $\rho < 1$, o par de soluções $x = y = \pm\kappa$ e $z = \rho - 1$ é instável. Quando $\rho > 1$, o equilíbrio torna-se instável em $\bar{x}_1$ e o par de soluções que descreve os outros dois pontos fixos adquire estabilidade linear. Trata-se de uma bifurcação denominada supercrítica de forquilha em $\rho = 1$ (ponto de *pitchfork*). A bifurcação de forquilha ocorre em sistemas que apresentam algum tipo de simetria. Com efeito, no sistema (1.12) essa bifurcação resulta da invariância do fluxo sob a simetria $(x, y, z) \leftrightarrow (-x, -y, z)$ mostrada na figura 1.1 por meio de uma seção de Poincaré do atrator de Lorenz. Em $\rho_c = \sigma(\sigma + \beta + 3)/(\sigma - \beta - 1)$ ocorrem bifurcações subcríticas de Hopf em $\bar{x}_2$ e $\bar{x}_3$. A bifurcação de Hopf é caracterizada pela existência de um par de autovalores puramente imaginários da matriz jacobiana calculada no ponto de bifurcação (ponto de Hopf). Para $\rho = \rho_c$, os autovalores são $\lambda_1 = -(\sigma + \beta + 1)$ e $\lambda_{2,3} = \pm i[2\sigma(\sigma + 1)/(\sigma - \beta - 1)]^{1/2}$ (admitindo $\sigma > \beta + 1$). Na bifurcação subcrítica de Hopf, a estabilidade é localmente perdida em ρ_c , onde, para qualquer $\rho > \rho_c$, observa-se a existência de órbitas instáveis. Nesse domínio do parâmetro de controle, todos os três pontos fixos são instáveis. Para $\rho \in [24, 74; 30, 10]$, um atrator estranho torna-se a única solução estável do fluxo. Na figura 1.2, ilustra-se este atrator estranho. Observa-se que: a) as trajetórias estão contidas numa região limitada do espaço de fases; b) o fluxo é recorrente⁶ em cada um dos setores do atrator em torno dos pontos fixos $\bar{x}_2$ e $\bar{x}_3$; c) existe dependência sensível às condições iniciais; e d) o atrator não é exatamente uma superfície, mas também não constitui um volume, a dimensão fractal desse atrator é 2,06.

Finalmente, observa-se que a dependência sensível às condições iniciais também implica em conseqüências nos cálculos numéricos. Algoritmos diferentes introduzirão pequenas diferenças nas soluções calculadas que podem levar a grandes desvios. No entanto, a aparência geral da solução e a estrutura do atrator estranho associado independem destes fatores.

1.3 Processos Estocásticos

Este trabalho é baseado na detecção de determinismo e, conseqüentemente, foi dada uma ênfase maior na teoria dos sistemas dinâmicos. Contudo, necessita-se

⁶Isto é, trajetórias que começam em qualquer subconjunto do atrator voltam a esse subconjunto para valores de t suficientemente grandes.

abordar certos aspectos de processos estocásticos e por esse motivo faz-se um resumo sobre este assunto; para maiores detalhes aconselham-se [23, 24, 25].

A definição formal de processo estocástico é dada a seguir.

Definição 1.9 *Um processo estocástico é uma família $\{X_t\}$ de variáveis aleatórias definidas sobre um espaço de probabilidade $(\Omega, \mathcal{C}, \mathbb{P})$, indexadas por elementos de um conjunto de parâmetros T . Quando $T \subset \mathbb{Z}$ diz-se que $\{X_t\}$ é um processo com parâmetro discreto, por outro lado, quando $T \subset \mathbb{R}$ diz-se que $\{X_t\}$ é um processo com parâmetro contínuo.*

A definição rigorosa de variável aleatória requer a noção de espaço de probabilidade. Esse tipo de espaço é dado pelo triplete $(\Omega, \mathcal{C}, \mathbb{P})$, onde Ω é um conjunto não vazio (espaço amostral), $\mathcal{C}$ é uma coleção de subconjuntos contidos em Ω e $\mathbb{P}$ é uma função definida por

$$\begin{aligned} \mathbb{P} : \mathcal{C} &\rightarrow \mathbb{R}^+ \\ \mathcal{C} &\rightarrow \mathbb{P}(\mathcal{C}) \end{aligned} \quad , \quad (1.13)$$

dando a probabilidade do evento $\mathcal{C}$. Mais especificamente, $\mathcal{C}$ é uma σ -álgebra sobre Ω , também chamada conjunto de informação ou experimento, enquanto seus elementos são denominados conjuntos mensuráveis, observáveis ou eventos. Diz-se que $\mathcal{C}$ é uma σ -álgebra se satisfizer

- C1. $\Omega \in \mathcal{C}$,
- C2. $\forall \mathcal{C}, \mathcal{C} \in \mathcal{C} \Rightarrow \mathcal{C}^c \in \mathcal{C}$,
- C3. $\forall \mathcal{C}_k, \mathcal{C}_k \in \mathcal{C} \Rightarrow \bigcup_{k=1}^{+\infty} \mathcal{C}_k \in \mathcal{C}$, para $k \in \mathbb{N}$.

Um exemplo de σ -álgebra é aquela que contém paralelepípedos de $\mathbb{R}^n$. Mais precisamente, essa noção requer a intersecção de todas as σ -álgebras que contém os paralelepípedos, denotada por $\sigma(\mathcal{R}^n)$ e chamada σ -álgebra gerada por $\mathcal{R}^n$.⁷ No caso de espaços topológicos gerais pode-se construir a σ -álgebra gerada pelos abertos desse espaço chamada de álgebra de Borel, cujos elementos são chamados conjuntos de Borel ou boreleanos. No espaço euclidiano $\mathbb{R}^n$ com álgebra de paralelepípedos $\mathcal{R}^n$ e topologia usual τ definida por alguma noção de distância, tem-se $\sigma(\mathcal{R}^n) = \sigma(\tau)$. Isso significa que os boreleanos de $\mathbb{R}^n$ nada mais são que elementos da σ -álgebra gerada pelos paralelepípedos. Uma notação conveniente para a álgebra de Borel sobre $\mathbb{R}^n$ é $\mathcal{B}(\mathbb{R}^n)$, ou simplesmente $\mathcal{B}^n$. Para maiores detalhes, aconselham-se [26, 27].

Apresenta-se agora a definição de variável aleatória.

⁷Neste contexto, $\mathcal{R}^n$ é uma álgebra definida por todas as uniões finitas e disjuntas de paralelepípedos em $\mathbb{R}^n$, seus elementos são representados pelo produto cartesiano de intervalos semi-abertos, isto é, um elemento típico dessa álgebra é expresso por $(a_1, b_1] \times \dots \times (a_n, b_n]$.

Definição 1.10 *Variável aleatória é uma função mensurável $X : \Omega \rightarrow \mathbb{R}$, onde $(\Omega, \mathcal{C}, \mathbb{P})$ é um espaço de probabilidade e $(\mathbb{R}, \mathcal{B}, \lambda)$ o espaço de Borel, que satisfaz $X^{-1}(\mathcal{B}) \in \mathcal{C}$, para $\forall \mathcal{B} \in \mathcal{B}$. A função $\lambda(a, b] = b - a$ é denominada medida de Lebesgue.*

Portanto, a variável aleatória X é uma característica numérica atribuída a experimentos envolvendo o espaço amostral Ω e o objetivo é determinar qual a probabilidade da imagem de X pertencer a certos subconjuntos de $\mathbb{R}$. É comum utilizar intervalos $(a, b]$, para $-\infty \leq a < b < +\infty$, de modo que a probabilidade de $X(\omega) \in (a, b]$ seja dada por $\mathbb{P}\{X^{-1}(a, b]\}$. Na figura 1.3, mostra-se um esquema que resume os conceitos aqui utilizados para definição de variável aleatória.

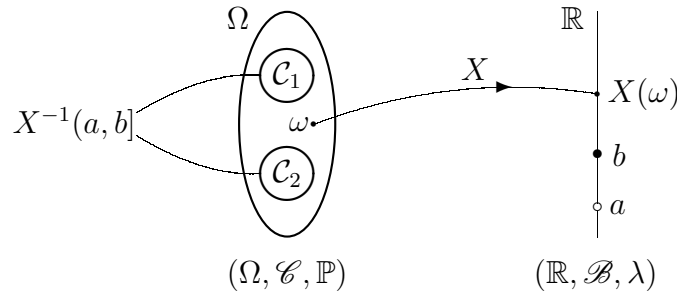


FIGURA 1.3: Esquema que caracteriza a relação entre os espaços de probabilidade e de Borel por meio da função variável aleatória. A probabilidade da variável aleatória X assumir valores no intervalo $(a, b]$ é idêntica a probabilidade do conjunto $X^{-1}(a, b] = C_1 \cup C_2$.

Uma variável aleatória é freqüentemente descrita por sua função de distribuição. A função de distribuição $f : \mathbb{R} \rightarrow \mathbb{R}^+$ da variável aleatória X é dada por

$$f_X(x) = \mathbb{P}\{\omega/X(\omega) \leq x\}. \quad (1.14)$$

Assume-se que existe uma função $g : \mathbb{R} \rightarrow \mathbb{R}^+$ chamada densidade de X , satisfazendo

$$f_X(x) = \int_{-\infty}^x g_X(x') dx'.$$

Nesse caso

$$\mathbb{P}\{X^{-1}(a, b]\} = \int_a^b g_X(x') dx'. \quad (1.15)$$

Por exemplo, caso a variável aleatória X tiver densidade gaussiana, escreve-se

$$\mathbb{P}\{X^{-1}(a, b]\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-(x'-\mu)^2/2\sigma^2} dx',$$

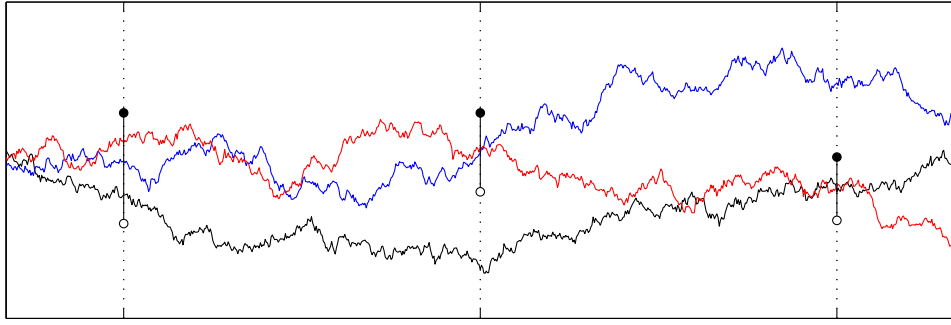


FIGURA 1.4: Um processo estocástico interpretado como uma família de trajetórias. A probabilidade de uma trajetória passar pelos boreleanos $\mathcal{B}_1$, $\mathcal{B}_2$ e $\mathcal{B}_3$ nos respectivos instantes t_1 , t_2 e t_3 , é definida como $\mathbb{P}[X_{t_1}^{-1}(\mathcal{B}_1) \cap X_{t_2}^{-1}(\mathcal{B}_2) \cap X_{t_3}^{-1}(\mathcal{B}_3)]$.

a função média (ou valor esperado) e variância de X são definidas, respectivamente, como

$$\mu = \langle X(\omega) \rangle = \int_{-\infty}^{+\infty} x' g_X(x') dx' , \quad (1.16)$$

$$\sigma^2 = \langle [X(\omega) - \mu]^2 \rangle = \int_{-\infty}^{+\infty} (x' - \mu)^2 g_X(x') dx' . \quad (1.17)$$

Diz-se que as variáveis aleatórias X_1 e X_2 são independentes se

$$\mathbb{P}[\{\omega/X_1(\omega) \in \mathcal{B}_1, X_2(\omega) \in \mathcal{B}_2\}] = \mathbb{P}[\{\omega/X_1(\omega) \in \mathcal{B}_1\}] \mathbb{P}[\{\omega/X_2(\omega) \in \mathcal{B}_2\}] ,$$

os intervalos $\mathcal{B}_1$ e $\mathcal{B}_2$ pertencem à álgebra de Borel dos intervalos de $\mathbb{R}$, maiores informações em [28, 29].

Dessa discussão, vem que um processo estocástico $\{X_t\}$ é uma variável aleatória para cada instante t , ou seja, trata-se de uma função de dois argumentos, $X_t(\omega)$, para $\omega \in \Omega$. Essa dependência indica que em t , obtém-se uma variável aleatória descrita por sua função de distribuição $f_{X_t}(x)$. Se X_t e X_τ forem variáveis aleatórias independentes para $t \neq \tau$ e, além disso, apresentarem a mesma distribuição, então define-se o processo como *i.i.d.*, isto é, independente e identicamente distribuído. Por outro lado, para cada $\omega \in \Omega$ fixo, obtém-se uma realização ou trajetória do processo. As trajetórias $X_\omega(t)$ do processo evoluem de acordo com as distribuições $f_{X_t}(x)$ e portanto são similares, no sentido de apresentarem as mesmas propriedades estatísticas. O conjunto de valores $\{X(\omega, t)\}$ é chamado espaço dos estados do processo estocástico. Na figura 1.4, ilustra-se trajetórias estocásticas em três instantes diferentes.

Uma suposição desejável, tratando-se de processos estocásticos, é a de estacionariedade. A estacionariedade permite a obtenção de propriedades estatísticas importantes para caracterização do processo. O processo estocástico $\{X_t\}$ é estacionário se as características estatísticas de X_τ , para todo $\tau \in T$, são as mesmas de X_t . Portanto, medidas feitas a partir da função média $\mu_t = \langle X_t(\omega) \rangle$ e da função variância $\sigma_t^2 = \langle [X_t(\omega) - \mu_t]^2 \rangle$, por exemplo, são constantes, ou seja, $\mu_t = \mu$ e $\sigma_t^2 = \sigma^2$, para todo $t \in T$. Uma sequência *i.i.d.* é estacionária e a notação reservada para esse processo quando as variáveis aleatórias tiverem densidade gaussiana será $X_t \sim i.i.d. N(\mu, \sigma^2)$. O problema na análise e caracterização da estacionariedade é que usualmente tem-se somente uma trajetória $X_\omega(t)$ do processo estocástico, sendo então necessário assumir a hipótese de ergodicidade.⁸

1.4 Detecção de Determinismo em Séries Temporais

1.4.1 Dimensão de Correlação

A dimensão de correlação (D_2) é uma das dimensões fractais associadas a um atrator estranho não homogêneo. Outras noções importantes são a dimensão fractal ou capacidade (D_0) e a dimensão de informação (D_1), mas a dimensão de correlação é a mais utilizada devido ao algoritmo desenvolvido por Grassberger e Procaccia [12].

Pode-se definir uma família de dimensões D_q , para $-\infty \leq q \leq +\infty$, e que servem para medir o grau de não homogeneidade do atrator estranho. Essa característica está associada com o fato de que algumas regiões do atrator são mais densas, ou seja, visitadas com uma maior frequência. Diz-se neste caso que o atrator é um multifractal. As dimensões D_q são conhecidas como as dimensões generalizadas de Renyi e satisfazem $D_{q'} \geq D_q$, para $q' < q$. De acordo com a desigualdade $D_0 \neq D_2$, a dimensão de correlação pode ser interpretada como um limite inferior para a dimensão fractal. A igualdade entre as infinitas dimensões acontecerá quando o atrator for homogêneo e, conseqüentemente, sua auto-similaridade for completamente descrita por uma única regra.

Inicialmente, quando se trabalha com séries temporais, é necessário construir vetores $\xi_m(t_i)$ m -dimensionais a partir da série $\{\tilde{x}(t_i)\}_{i=1}^\ell$, fazendo

$$\xi_m(t_i) = \{\tilde{x}_{t_i}, \tilde{x}_{t_i-\tau}, \tilde{x}_{t_i-2\tau}, \dots, \tilde{x}_{t_i-(m-1)\tau}\}, \quad (1.18)$$

onde m é a dimensão de imersão, τ é o parâmetro de atraso (*time-delay*) e cada $\xi_m(t_i)$ representa um ponto no espaço euclidiano m -dimensional. Takens [30], demonstrou

⁸A ergodicidade permite que a média μ_t e a variância σ_t^2 do processo estocástico possam ser substituídas, respectivamente, pela média e variância temporais de uma realização.

que se o sistema (1.4) fornece um fluxo n -dimensional, onde $\tilde{x}$ é uma medida feita sobre a trajetória x do sistema. Por meio dos vetores ξ_m obtém-se um espaço de imersão contínuo para esse fluxo e as propriedades métricas invariantes, tais como dimensão fractal e expoentes característicos de Lyapunov, em ambos os espaços (o espaço n -dimensional de x e m -dimensional de ξ_m) são as mesmas.

Este método é denominado usualmente como reconstrução de Takens ou método dos atrasos temporais e permite a reconstrução do atrator estranho associado a $\{x(t_i)\}_{i=1}^{\ell}$, embora o atrator reconstruído não seja idêntico ao original.

Se o movimento ocorrer sobre um atrator estranho, então a trajetória voltará para a ε -vizinhança de qualquer ponto $\xi_m(t_i)$ após algum tempo. Como efeito disso, os pontos $\xi_m(t_i)$ do atrator estão espacialmente organizados. A medida desta organização espacial é a integral de correlação, que pode ser escrita como

$$C_{m,\ell}(\varepsilon) = \frac{2}{\ell(\ell-1)} \sum_{i=1}^{\ell} \sum_{j=i+t_{\min}}^{\ell} \Theta[\varepsilon - \|\xi_m(t_i) - \xi_m(t_j)\|] , \quad (1.19)$$

onde $t_{\min}$ elimina a correlação entre pontos medidos consecutivamente (que ocasionam uma falsa estrutura com baixa dimensionalidade), $\Theta(x)$ é a função degrau de Heaviside, definida por

$$\Theta(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases} ,$$

e ε é o raio da hipersfera que cerca o ponto $\xi_m(t_i)$. A dimensão de correlação pode ser calculada fazendo

$$D_2(m) = \lim_{\varepsilon \rightarrow 0} \lim_{\ell \rightarrow +\infty} \frac{\partial \log_{10} C_{m,\ell}(\varepsilon)}{\partial \log_{10} \varepsilon} , \quad (1.20)$$

isto é, a inclinação da reta $\log_{10} C_{m,\ell}(\varepsilon) \times \log_{10} \varepsilon$.

A dimensão de correlação é estimada sobre uma região de escala apropriada do parâmetro ε . Para ε pequeno, espera-se que $\log_{10} C_{m,\ell}(\varepsilon) \rightarrow -\infty$, para valores intermediários de ε , a curva $\log_{10} C_{m,\ell}(\varepsilon) \times \log_{10} \varepsilon$ pode eventualmente apresentar uma região com comportamento linear e, para valores grandes de ε , identifica-se uma região onde $\log_{10} C_{m,\ell}(\varepsilon) \rightarrow 0$, pois esta é uma situação em que todos os pontos do atrator estranho estão dentro da hipersfera de raio ε .

Em princípio, uma série temporal gerada por um processo estocástico possui uma dimensão de correlação infinita, pois não é esperado que sua órbita tenha alguma estrutura espacial. Tratando-se de processos estocásticos não existem atratores e a dinâmica, mesmo em condições assintóticas, necessita da totalidade do espaço de fases. Um simples teste de determinismo consiste no aumento da dimensão de

imersão até uma possível convergência da dimensão de correlação. Se a dimensão de correlação não convergir em relação aos incrementos na dimensão de imersão, a série temporal será estocástica, caso contrário será determinística. De fato, a dimensão de imersão do sistema é o limite no qual a dimensão de correlação não cresce consideravelmente com o aumento dos graus de liberdade.

Apesar dessa técnica ter sido usada extensamente, há razões para questionar conclusões baseadas apenas na dimensão de correlação para caracterização de sistemas determinísticos. Dentre elas, destacam-se: a) dificuldade de se obter resultados confiáveis para conjuntos com número reduzido de dados; b) precauções devem ser adotadas na técnica de reconstrução do espaço de fases; c) a determinação de uma região de escala, essencial para o cálculo da dimensão de correlação é, em geral, subjetiva; e d) existência de ruído faz com que cálculos baseados na dimensão de correlação sejam inúteis. Um nível de ruído de 2% já é considerado prejudicial [9].

1.4.2 Diagrama de Recorrência

O diagrama de recorrência (*recurrence plot*), introduzido por Eckmann *et al.* [31], consiste em um método qualitativo que detecta quando uma série temporal pode exibir comportamento caótico, e é baseado em evidências de recorrência em órbitas periódicas instáveis imersas no atrator estranho [11, 16]. Este método topológico preserva a ordem temporal da série, analisando a forma na qual os mecanismos responsáveis pela expansão e contração do atrator estranho influenciam as órbitas associadas ao sistema, entrelaçando-as de forma muito específica. Órbitas periódicas instáveis existem com abundância em atratores estranhos; elas são densas em casos que o atrator possui estrutura geométrica hiperbólica [32].⁹

Apresenta-se agora a descrição do método. Se uma observação x_i da série temporal $\{x_i\}_{i=1}^n$ ocorrer próxima a uma órbita periódica instável, então observações posteriores irão evoluir próximas desta órbita enquanto não forem repelidas. Se as observações evoluírem por um tempo suficientemente longo no atrator estranho, elas irão retornar para a ε -vizinhança de x_i após algum intervalo τ , onde τ indica o período da órbita. Para detectar as regiões de *close returns* (regiões de retornos próximos) em uma série temporal é utilizado um diagrama de cores onde representam-se todas as diferenças $|x_i - x_{i+\tau}|$ calculadas. Se a diferença indexada por (i, τ) for menor que um limiar ε , rotula-se a região correspondente de preto; se a diferença for maior do que ε , rotula-se a região de branco. O eixo horizontal do diagrama indica o número da observação i , para $i = 1, 2, \dots, n$, e o eixo vertical do

⁹A estrutura hiperbólica é devida ao fato de existirem direções estáveis e instáveis presentes simultaneamente no atrator estranho.

diagrama indica τ , para $\tau = 1, 2, \dots, n - i$. As regiões de *close returns* são indicadas por segmentos de linhas horizontais. Por exemplo, para um segmento horizontal limitado por t e t' (para $t' > t$), a observação inicial x_t indica onde a série temporal começa a seguir a órbita periódica instável; a observação final é $x_{t'+\tau}$, que representa o ponto pertencente à ε -vizinhança de $x_{t'}$, após um período τ . Na figura 1.5 ilustra-se este exemplo.

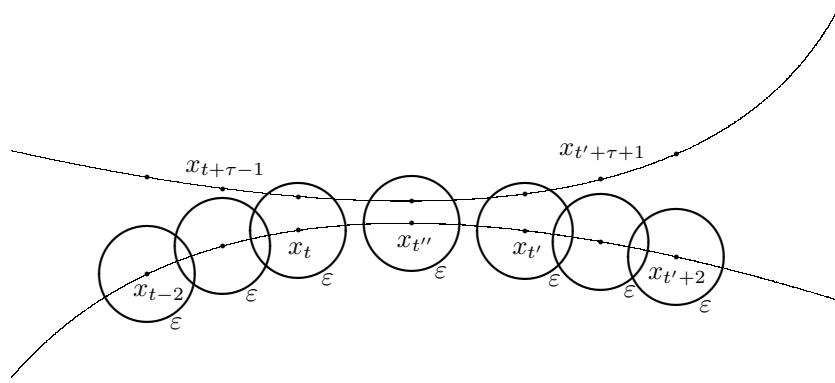


FIGURA 1.5: Região de *close returns* definida no intervalo $[t, t']$. O instante t'' é definido como $t'' = (t' + t)/2$.

Se a série temporal for caótica, segmentos de linhas horizontais serão visualizados no diagrama. Todavia, se a série temporal for estocástica, uma variedade uniformemente distribuída de pontos pretos será observada em todo o diagrama. A figura 1.6 mostra o diagrama de recorrência para uma série temporal estocástica *i.i.d.* $N(0, 1)$ e a figura 1.7 mostra o diagrama de recorrência para a série temporal associada ao sistema não-linear de equações diferenciais ordinárias proposto por Rössler [33], definido como

$$\begin{cases} \dot{x} = -y - z \\ \dot{y} = x + ay \\ \dot{z} = b + z(x - c) \end{cases}, \quad (1.21)$$

com $a = 0, 2$; $b = 0, 2$ e $c = 5, 7$.

Em comparação com o método métrico baseado na dimensão de correlação, o diagrama de recorrência apresenta algumas importantes vantagens, dentre as quais destacam-se: a) o método topológico pode ser aplicado em séries temporais com tamanho relativamente pequeno; b) o método é robusto mesmo com a existência de ruído; e c) a aplicação do método preserva a ordem temporal da série.

A implementação do algoritmo utilizado neste trabalho é devido a Claire G. Gilmore [11], não necessitando de técnicas para a reconstrução da dinâmica no

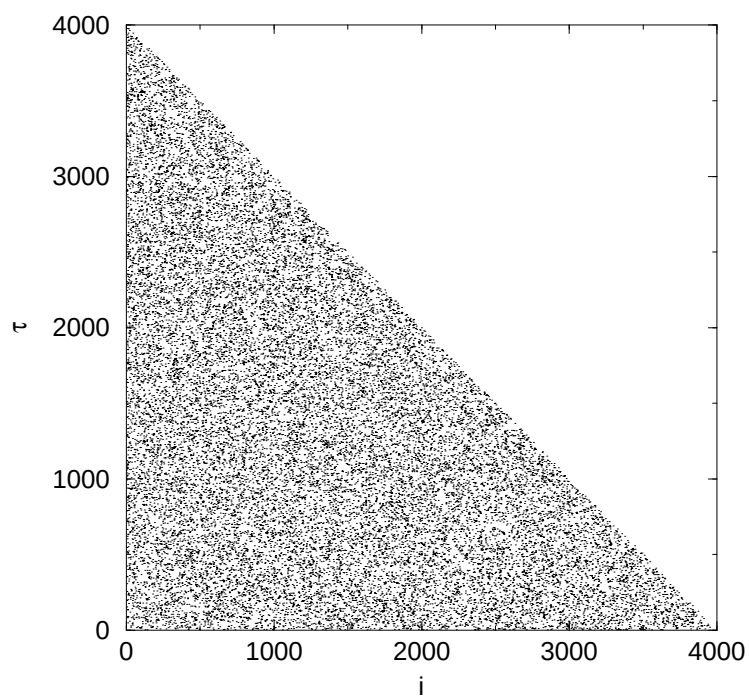
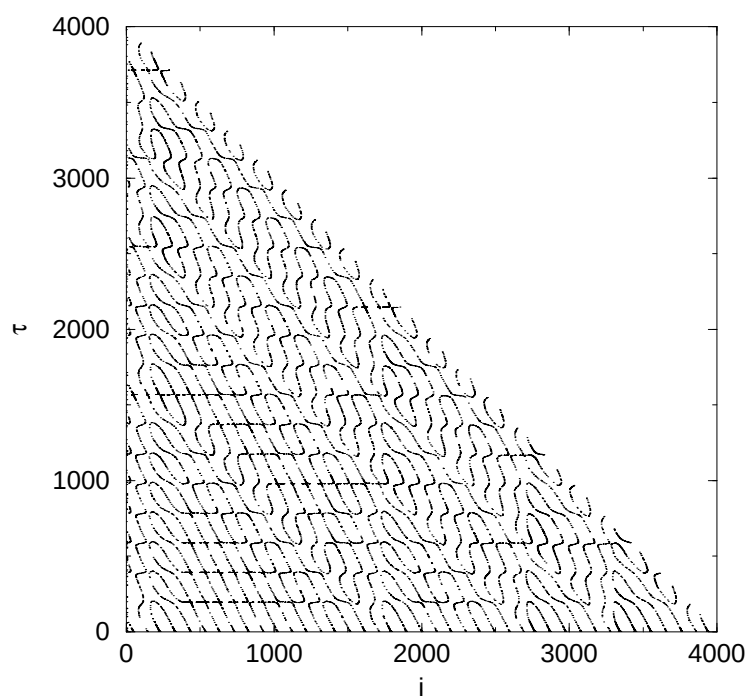
FIGURA 1.6: Diagrama de recorrência para uma série estocástica *i.i.d.* $N(0,1)$.

FIGURA 1.7: Diagrama de recorrência para a série associada ao sistema de Rössler.

espaço de fases. Esta adaptação do diagrama de recorrência é denominada *close returns test*. O limiar ε foi determinado a partir de uma pequena fração da distância média entre todas as observações da série temporal. Se ε for muito pequeno, dificilmente caracteriza-se algum padrão devido ao número insuficiente de pontos pretos e, se ε for muito grande, pode-se ter um padrão escondido no diagrama.

1.4.3 Complexidade de Lempel-Ziv

A complexidade de Lempel-Ziv é um terceiro método utilizado para a classificação de séries temporais. Este parâmetro é de natureza totalmente distinta dos métodos métrico e topológico já apresentados. Esta medida de complexidade introduzida por Lempel e Ziv [34, 35], não necessita de nenhuma dimensão de imersão e a série é interpretada como um sinal binário gerado por algum tipo de fonte. Esta idéia está presente na teoria da comunicação, onde deseja-se determinar o alfabeto mínimo necessário para codificar uma fonte cujo sinal será enviado por um canal com ruído.

De acordo com Kolmogorov [17], a complexidade algorítmica (ou de Kolmogorov) de uma seqüência binária de tamanho n é dada pelo comprimento $K(n)$ do menor programa de computador que gera esta seqüência. Segundo Solomonoff [36], não existe um algoritmo geral que determine tal programa. Lempel e Ziv escolheram então, a partir de todos os programas possíveis, uma classe que permite somente duas operações: copiar e inserir. Além disso, eles não calcularam o comprimento $K(n)$ do programa, mas um número $n_w(n)$ que é uma medida útil deste comprimento. Trata-se de uma medida apropriada da complexidade de Kolmogorov [23].

O algoritmo de Lempel-Ziv é construído por meio de um particionamento que separa a seqüência original em subseqüências (ou palavras) de comprimentos menores. Considere a seqüência $s_1 s_2 \dots s_n$ reconstruída por um programa até o dígito s_r , e que s_r é um dígito inserido, ou seja, não foi obtido simplesmente por um processo de cópia de $s_1 s_2 \dots s_{r-1}$. A subseqüência limitada por s_r será denotada por $S = s_1 s_2 \dots s_r$!, onde o uso do apóstrofo indica que s_r foi inserido. Na condição de verificar se o resto da seqüência, isto é, $s_{r+1} s_{r+2} \dots s_n$ pode ser reconstruído por meio de uma simples cópia de S , Lempel e Ziv procederam com os seguintes passos. Definir $Q = s_{r+1}$ e verificar se este termo está contido no vocabulário da subseqüência S , de modo que Q pode ser simplesmente obtido copiando-se alguma palavra de S . Isto é equivalente a verificar se Q está contido no vocabulário $v(SQ\pi)$ de $SQ\pi$, onde $SQ\pi$ denota a subseqüência que é composta por S e Q , e π indica que o último dígito foi retirado da subseqüência (no contexto atual, $SQ\pi = S$). Esta última formulação da questão pode ser estendida para situações onde Q contém dois (isto

Capítulo 2

Estacionariedade e Linearidade

2.1 Introdução

Muitos métodos utilizados na literatura para análise de séries temporais, lineares ou não-lineares, assumem algum tipo de estacionariedade, de modo que estudos preliminares quanto a este critério é uma etapa essencial para uma futura modelagem da série. Todavia, a estacionariedade não é somente um problema técnico na análise de séries temporais; o estudo e compreensão de sinais não-estacionários são tópicos atuais de pesquisa em muitas áreas da ciência.

No capítulo presente, apresentam-se dois métodos para análise de estacionariedade. O primeiro método, o teste da raiz unitária de Dickey-Fuller, baseia-se na análise dos coeficientes obtidos de um ajuste linear da série temporal, sendo então necessária uma breve descrição das especificações utilizadas na modelagem linear, bem como a proposta de dois métodos para análise de linearidade da série: a estatística BDS e o teste *surrogate data* [4, 37, 38, 39]. O segundo método para análise de estacionariedade, o método *cross prediction error*, baseia-se na predição da série temporal com base na similaridade entre segmentos obtidos do particionamento da própria série. Este método requer como condição básica que a série analisada seja determinística [4].

2.2 Modelagem de Séries Temporais Lineares

2.2.1 Modelos Auto-Regressivos

A série temporal $\{x_t\}_{t=1}^N$ é realização de um processo auto-regressivo de ordem p e escreve-se $AR(p)$, se a observação x_t satisfizer a equação

$$x_t = \eta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \epsilon_t, \quad (2.1)$$

onde $\epsilon_t \sim i.i.d. N(0, \sigma_\epsilon^2)$ e está descorrelacionado com x_{t-j} , para $1 \leq j \leq p$.

Supondo $\eta = 0$ e com o auxílio do operador retroativo Λ , definido pela operação $\Lambda^m x_t = x_{t-m}$, para $m \geq 1$, reescreve-se x_t como

$$\alpha(\Lambda)x_t = \epsilon_t, \quad (2.2)$$

onde

$$\alpha(\Lambda) = 1 - \alpha_1 \Lambda - \alpha_2 \Lambda^2 - \dots - \alpha_p \Lambda^p,$$

define o operador auto-regressivo de ordem p .

Em Box *et al.* [40], demonstrou-se que a condição para que $\{x_t\}_{t=1}^N$ seja estacionária, onde estacionariedade significa um processo com média e variância constantes, é que todas as raízes de $\alpha(\Lambda) = 0$ estejam fora do círculo unitário. Em particular, quando $p = 1$ obtém-se $\Lambda = \alpha^{-1}$, e a condição implica $|\alpha| < 1$.

A demonstração proposta por esses autores consiste na análise das funções autocovariância e autocorrelação de um processo estacionário, respectivamente definidas como

$$\gamma_\tau = \langle (x_t - \mu)(x_{t+\tau} - \mu) \rangle, \quad (2.3)$$

$$\rho_\tau = \frac{\langle (x_t - \mu)(x_{t+\tau} - \mu) \rangle}{\sqrt{\langle (x_t - \mu)^2 \rangle \langle (x_{t+\tau} - \mu)^2 \rangle}}, \quad (2.4)$$

onde $\langle (x_t - \mu)^2 \rangle = \langle (x_{t+\tau} - \mu)^2 \rangle = \gamma_0$, indicando $\rho_0 = 1$. Na realidade, a autocovariância é uma função de $|\tau|$ de modo que $\gamma_\tau = \gamma_{-\tau}$. Com o auxílio da desigualdade de Cauchy-Schwarz, dada por

$$|\langle (x_t - \mu)(x_{t+\tau} - \mu) \rangle| < \sqrt{\langle (x_t - \mu)^2 \rangle \langle (x_{t+\tau} - \mu)^2 \rangle}, \quad \text{para } \tau \neq 0,$$

verifica-se $|\gamma_\tau| < \gamma_0$ como condição de estacionariedade do processo.

Em particular, usando o modelo AR(1), descrito pelo operador auto-regressivo $\alpha(\Lambda) = 1 - \alpha\Lambda$, multiplicado por $x_{t-\tau}$ e logo após tomando o valor médio, obtém-se a relação

$$\rho_\tau = \alpha \rho_{\tau-1} = \dots = \alpha^\tau \rho_0, \quad \tau \geq 0,$$

ou seja, reescrevendo a condição de estacionariedade como $|\rho_\tau| < 1$, a função de autocorrelação decairá exponencialmente em duas formas diferentes dependendo do sinal de α e necessariamente $|\alpha| < 1$ para que o processo seja estacionário.

2.2.2 Modelos de Médias Móveis

A série temporal $\{x_t\}_{t=1}^N$ é realização de um processo de médias móveis de ordem q e escreve-se $\text{MA}(q)$, se a observação x_t satisfizer a equação

$$x_t = \eta - v_1\epsilon_{t-1} - v_2\epsilon_{t-2} - \dots - v_q\epsilon_{t-q} + \epsilon_t, \quad (2.5)$$

onde $\epsilon_t \sim i.i.d. N(0, \sigma_\epsilon^2)$. Segue-se que x_t é estacionária, podendo ser interpretada como uma média ponderada de $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$.

Define-se o operador de médias móveis de ordem q como

$$v(\Lambda) = 1 - v_1\Lambda - v_2\Lambda^2 - \dots - v_q\Lambda^q,$$

onde $\Lambda^m\epsilon_t = \epsilon_{t-m}$, se $m \geq 1$. Supondo $\eta = 0$, o modelo $\text{MA}(q)$ pode ser descrito na forma

$$x_t = v(\Lambda)\epsilon_t. \quad (2.6)$$

Em particular, para o modelo $\text{MA}(1)$, tem-se $v(\Lambda) = 1 - v\Lambda$, de tal forma que se escreve x_t como

$$x_t = (1 - v\Lambda)\epsilon_t, \quad (2.7)$$

de onde, formalmente, segue

$$\epsilon_t = (1 - v\Lambda)^{-1}x_t = (1 + v\Lambda + v^2\Lambda^2 + \dots)x_t, \quad (2.8)$$

ou seja, reescreve-se x_t na forma

$$x_t = -vx_{t-1} - v^2x_{t-2} - \dots + \epsilon_t. \quad (2.9)$$

Para que a série do lado direito convirja, é necessário que $|v| < 1$. Nessa equação, tem-se x_t escrita como um modelo auto-regressivo de ordem infinita. Diz-se que $|v| < 1$ é uma condição de invertibilidade para o modelo $\text{MA}(1)$.

2.2.3 Modelos Auto-Regressivos e de Médias Móveis

Um modelo auto-regressivo e de médias móveis de ordem (p, q) , denotado por $\text{ARMA}(p, q)$, é definido por

$$x_t = \eta + \alpha_1x_{t-1} + \dots + \alpha_px_{t-p} - v_1\epsilon_{t-1} - \dots - v_q\epsilon_{t-q} + \epsilon_t, \quad (2.10)$$

onde $\epsilon_t \sim i.i.d. N(0, \sigma_\epsilon^2)$.

Supondo $\eta = 0$ e usando os operadores auto-regressivo e de médias móveis, definidos anteriormente, reescreve-se x_t na forma

$$\alpha(\Lambda)x_t = v(\Lambda)\epsilon_t. \quad (2.11)$$

Para um modelo ARMA(p, q) genérico, a condição de estacionariedade é a mesma que para um modelo AR(p), ou seja, as raízes de $\alpha(\Lambda) = 0$ devem estar fora do círculo unitário, e a condição de invertibilidade é a mesma que para um modelo MA(q), ou seja, as raízes de $v(\Lambda) = 0$ também devem estar fora do círculo unitário.

2.2.4 Modelos Auto-Regressivos Integrados e de Médias Móveis

Em muitos casos, séries temporais empíricas são não-estacionárias. Uma série temporal $\{x_t\}_{t=1}^N$ não-estacionária pode ser transformada em estacionária por meio da diferenciação efetuada d vezes a partir de seus valores originais. Assim, quando $\{x_t\}_{t=1}^N$ torna-se estacionária por diferenciação, diz-se que a série é integrada de ordem d , ou $\mathcal{I}(d)$. A série temporal $\{x_t\}_{t=1}^N$ é não-estacionária homogênea de ordem d se

$$\{w_t\}_{t=1}^{N-d} = \{\Delta^d x_t\}_{t=d+1}^N, \quad (2.12)$$

é uma série estacionária, onde Δ denota a diferenciação, ou seja,

$$\Delta x_t = x_t - x_{t-1}, \quad \Delta^2 x_t = \Delta x_t - \Delta x_{t-1}, \quad \dots, \quad \Delta^d x_t = \Delta^{d-1} x_t - \Delta^{d-1} x_{t-1}. \quad (2.13)$$

Um modelo ARMA estimado com base numa série integrada de ordem d é denominado modelo auto-regressivo integrado e de médias móveis. Genericamente, esse tipo de modelo costuma ser indicado por ARIMA(p, d, q), onde p refere-se à ordem em que a série temporal será auto-regredida, d é a ordem de integração e q a ordem do operador de médias móveis.

2.3 Testes de Estacionariedade

2.3.1 Teste da Raiz Unitária de Dickey-Fuller

Dickey e Fuller [2], desenvolveram um teste para verificar a existência de raiz unitária em uma série temporal. O procedimento básico para a realização desse teste para uma série $\{x_t\}_{t=1}^N$ consiste em regredi-la contra seus valores defasados de uma unidade, x_{t-1} . Logo após, testar a significância estatística do coeficiente $\hat{\rho}$ associado a x_{t-1} , estimado por mínimos quadrados, utilizando-se de um teste de hipóteses.

Do ponto de vista estatístico, um teste de hipóteses, admitindo-se correta a hipótese sob prova, consiste em determinar certa região Ω_0 do espaço amostral Ω , tal que a probabilidade de um evento q pertencer a Ω_0 seja igual a um número positivo, ns , denominado nível de significância, compreendido entre 0 e 1, isto é, $\mathbb{P}(q \in \Omega_0) = ns$. Desse modo, $1 - ns$ é o nível de confiança nc do teste, Ω_0 é a região crítica e $\Omega - \Omega_0$ é a região de confiança. As hipóteses estabelecidas podem ser de dois tipos: a) hipótese nula H_0 , quando se admite não haver diferença entre a informação fornecida pela evidência amostral e a afirmação da hipótese; e b) hipótese alternativa H_1 , quando se admite haver diferença entre a informação fornecida pela evidência amostral e a afirmação da hipótese estabelecida para realização do teste.

Assim, o processo do teste de hipóteses consiste em aceitar ou rejeitar a hipótese nula H_0 . O critério adotado, com base na evidência amostral, não garante uma conclusão correta do teste. Dois tipos de erro podem ocorrer: a) rejeição da hipótese nula H_0 , quando esta for verdadeira (ns é a probabilidade de ocorrência deste erro); e b) aceitação da hipótese nula H_0 , quando esta for falsa. O nível de confiança nc corresponde à probabilidade de aceitação de H_0 , quando esta for verdadeira.

Uma série temporal é não-estacionária se possui raiz unitária, ou seja, quando a hipótese $\rho = 1$ for aceita. As hipóteses do teste da raiz unitária podem ser formuladas como hipótese nula $H_0 : \rho = 1$ (presença de raiz unitária ou série não-estacionária) e hipótese alternativa $H_1 : -1 < \rho < 1$ (ausência de raiz unitária ou série estacionária).

Há três especificações para aplicação do teste de Dickey-Fuller. Formalmente, escreve-se as especificações como:

a) equação sem intercepto e sem tendência

$$x_t = \rho x_{t-1} + \epsilon_t ; \quad (2.14)$$

b) equação com intercepto e sem tendência

$$x_t = \eta + \rho x_{t-1} + \epsilon_t ; \quad (2.15)$$

c) equação com intercepto e com tendência

$$x_t = \eta + \beta t + \rho x_{t-1} + \epsilon_t ; \quad (2.16)$$

onde $\epsilon_t \sim i.i.d. N(0, \sigma_\epsilon^2)$.

O critério que decide se a hipótese nula H_0 deve ser rejeitada ou não, com base na evidência amostral, consiste em definir um limite crítico t_c que divida o espaço amostral em duas partes: uma região de rejeição (ou crítica) e uma região de aceitação (ou de confiança). Tal limite é definido de modo que permita confrontar a

hipótese nula H_0 com a evidência amostral. A região de rejeição é um subconjunto do espaço amostral, no qual, se contiver o coeficiente indicativo da evidência amostral, a hipótese nula H_0 é rejeitada. Já a região de confiança é um subconjunto do espaço amostral que, contendo o coeficiente indicativo da evidência amostral, permite a aceitação da hipótese nula H_0 .

A estatística $t_{\hat{\rho}}$, que mede o coeficiente indicativo associado ao $\hat{\rho}$ estimado da evidência amostral com o auxílio do teste t de Student, pode ser definida de acordo com [41] como

$$t_{\hat{\rho}} = \frac{\hat{\rho} - 1}{s_{\hat{\rho}}}, \quad (2.17)$$

onde $s_{\hat{\rho}}$ é o erro-padrão do coeficiente $\hat{\rho}$ estimado associado a x_{t-1} e sua definição depende da especificação utilizada no teste. Por exemplo, quando a especificação linear for a equação (2.15), $s_{\hat{\rho}}$ é estimado fazendo

$$s_{\hat{\rho}}^2 = \frac{1}{(N-2)} \frac{\sum_{t=2}^N (x_t - \hat{\eta} - \hat{\rho}x_{t-1})^2}{\sum_{t=2}^N (x_{t-1} - \langle x_{t-1} \rangle)^2}, \quad (2.18)$$

onde N é o tamanho da série e o termo $(N-2)$ no denominador serve para obter um estimador não viesado.

A regra de decisão será dada fazendo a comparação de $t_{\hat{\rho}}$ com o limiar t_c , se $|t_{\hat{\rho}}| > |t_c|$; rejeita-se H_0 no nível de significância adotado e para $|t_{\hat{\rho}}| \leq |t_c|$, aceita-se H_0 no nível de significância adotado. Os valores críticos t_c para as três especificações alternativas foram construídos com base em simulações de Monte Carlo e encontram-se em [42].

Foi apresentado o teste simples de Dickey-Fuller. No entanto, quando os resíduos estão correlacionados, escreve-se x_t para o caso mais geral como

$$x_t = \eta + \beta t + \rho_1 x_{t-1} + r_t, \quad (2.19)$$

onde

$$r_t = \rho_2 x_{t-2} + \dots + \rho_p x_{t-p} + \epsilon_t, \quad (2.20)$$

para $\epsilon_t \sim i.i.d. N(0, \sigma_\epsilon^2)$.

Supondo, por exemplo, resíduos correlacionados de modo que a observação x_t seja auto-regredida até uma ordem $p = 2$, a especificação mais geral para a aplicação do teste fica como

$$x_t = \eta + \beta t + \rho_1 x_{t-1} + \rho_2 x_{t-2} + \epsilon_t,$$

e o modelo pode ser reescrito, mediante a adição e a subtração de $\rho_2 x_{t-1}$, na forma

$$x_t = \eta + \beta t + \gamma x_{t-1} + \xi_1 \Delta x_{t-1} + \epsilon_t ,$$

onde $\Delta x_{t-1} = x_{t-1} - x_{t-2}$, $\gamma = \rho_1 + \rho_2$ e $\xi_1 = -\rho_2$.

Este é o teste ampliado de Dickey-Fuller e a hipótese nula é formulada como $H_0 : \gamma = 1$ (presença de raiz unitária ou série não-estacionária). No caso em que é necessário auto-regredir x_t até uma ordem p para que os resíduos não estejam mais correlacionados, a especificação mais geral pode ser reescrita como

$$x_t = \eta + \beta t + \gamma x_{t-1} + \sum_{j=1}^{p-1} \xi_j \Delta x_{t-j} + \epsilon_t . \quad (2.21)$$

Neste caso, $\gamma = \sum_{m=1}^p \rho_m$ e $\xi_j = -\sum_{k=j+1}^p \rho_k$. Essa transformação em x_t que resulta na equação (2.21) reescreve o operador auto-regressivo de ordem p , definido como $\rho(\Lambda) = 1 - \rho_1 \Lambda - \rho_2 \Lambda^2 - \dots - \rho_p \Lambda^p$, na forma simplificada $\gamma(\Lambda) = 1 - \gamma \Lambda$, onde facilmente determina-se as raízes de $\gamma(\Lambda) = 0$. A única (e suficiente) condição de estacionariedade a ser analisada é a de que todas as raízes do operador auto-regressivo, que caracterizam a modelagem da série, estejam fora do círculo unitário, não importando a ordem do operador. Reescreve-se então qualquer operador auto-regressivo de ordem superior a 1, ou seja, $p > 1$, como um operador de primeira ordem sem o comprometimento do teste, aplicando a transformação apresentada.

Finalmente, estima-se a equação acima e obtém-se a estatística $t_{\hat{\gamma}}$ associada a x_{t-1} , realizando-se o teste normalmente. Se não for detectada raiz unitária, há estacionariedade, ou diz-se que a série é integrada de ordem zero, ou seja, $\mathcal{I}(0)$. Caso contrário, há não-estacionariedade, e repete-se o procedimento com o uso dos termos $w_{t-1} = \Delta x_t = x_t - x_{t-1}$ obtidos fazendo-se a primeira diferença. Se a série $\{w_t\}_{t=1}^{N-1}$ for estacionária, a série temporal será $\mathcal{I}(1)$ e assim por diante. Desse modo, uma série temporal $\mathcal{I}(d)$, para $d > 0$, tem pelo menos uma raiz unitária.

2.3.2 Método *Cross Prediction Error*

Schreiber [5], introduziu um método para detectar e analisar sinais potencialmente não-estacionários por meio de variações dinâmicas na série temporal. O método *cross prediction error* baseia-se na predição de um segmento da série temporal utilizando outros segmentos pertencentes à série como uma base de dados.

Este teste tem uma contribuição conceitual muito importante, pois é baseado na similaridade entre segmentos da série temporal, em vez da análise de parâmetros estatísticos estimados a partir de médias locais, por exemplo. A aplicação do teste é particularmente útil se a não-estacionariedade estiver associada com mudanças

na forma do atrator enquanto que invariantes dinâmicos permaneçam efetivamente constantes [4]. Enquanto o teste de Dickey-Fuller aplica-se à séries lineares, o presente método detecta formas sutis de não-estacionariedade em sistemas determinísticos não-lineares.

Considere a série temporal $\{x_n\}_{n=1}^N$ particionada por segmentos de comprimento ℓ , de tal forma que o i -ésimo segmento será definido como $\mathcal{S}_i^\ell = \{x_{(i-1)\ell+1}, \dots, x_{i\ell}\}$. Usualmente calcula-se uma dada estatística γ_i para cada segmento $\mathcal{S}_i^\ell$, verificando se a sequência $\{\gamma_i\}_{i=1}^{N/\ell}$ é constante (desconsidera-se as flutuações estatísticas). O método *cross prediction error* difere dessa abordagem, pois utiliza pares de segmentos da série temporal, ou seja, $\gamma_{ij} = \gamma(\mathcal{S}_i^\ell, \mathcal{S}_j^\ell)$. Esta metodologia aumenta o número de parâmetros calculados de N/ℓ para $(N/\ell)^2$.

Apresenta-se agora o cálculo da estatística utilizada no teste. Inicialmente, define-se $\tilde{x} = \{x_n\}_{n=1}^{N_x}$ e $\tilde{y} = \{y_n\}_{n=1}^{N_y}$ como duas séries temporais quaisquer e m como um pequeno número inteiro representando a dimensão de imersão. De ambas as séries temporais, constroem-se vetores imersos $\{\vec{x}_n\}_{n=m}^{N_x-1}$ e $\{\vec{y}_n\}_{n=m}^{N_y-1}$, respectivamente, no mesmo espaço de fases m -dimensional, onde $\vec{x}_n = \{x_{n-(m-1)}, \dots, x_n\}$. Para cada vetor $\vec{y}_n = \{y_{n-(m-1)}, \dots, y_n\}$, pretende-se estimar y_{n+1} , usando, entretanto, como banco de dados a série temporal $\tilde{x}$. Utiliza-se um estimador que relaciona $\vec{x}_n$ com x_{n+1} , definido por

$$\hat{y}_{n+1}^{\tilde{x}} = \frac{1}{[\mathcal{W}_{\varepsilon}^{\tilde{x}}(\vec{y}_n)]_{\#}} \sum_{\vec{x}_{n'} \in \mathcal{W}_{\varepsilon}^{\tilde{x}}(\vec{y}_n)} x_{n'+1}, \quad (2.22)$$

onde $\mathcal{W}_{\varepsilon}^{\tilde{x}}(\vec{y}_n) = \{\vec{x}_{n'} : \|\vec{x}_{n'} - \vec{y}_n\| < \varepsilon\}$ é uma ε -vizinhança de $\vec{y}_n$, formada, porém, dentro do conjunto $\tilde{x}$. O termo $[\mathcal{W}_{\varepsilon}^{\tilde{x}}(\vec{y}_n)]_{\#}$ denota o número de elementos que estão dentro da hipersfera de $\vec{y}_n$ com raio ε . Para pontos isolados com a vizinhança vazia, calcula-se $\hat{y}_{n+1}^{\tilde{x}}$ por meio da média amostral do segmento $\tilde{x}$. A estatística $\gamma(\tilde{x}, \tilde{y})$ da sequência $\tilde{y}$, dado $\tilde{x}$, é então definida como

$$\gamma(\tilde{x}, \tilde{y}) = \left[\frac{1}{N_y - m} \sum_{n=m}^{N_y-1} (\hat{y}_{n+1}^{\tilde{x}} - y_{n+1})^2 \right]^{1/2}, \quad (2.23)$$

e investiga localmente como a dinâmica de $\tilde{x}$ é adequada para a predição de valores em $\tilde{y}$. Em particular, se o atrator de $\tilde{y}$ estiver imerso no atrator de $\tilde{x}$, pontos em $\tilde{y}$ podem ser determinados usando $\tilde{x}$ como uma base de dados. Todavia, $\tilde{y}$ não contém informação suficiente para estimar todos os pontos de $\tilde{x}$. Enquanto a estatística anti-simétrica γ_{ij} fornece informações importantes quanto à estacionariedade do ponto de vista dinâmico, pode também ser confusa em alguns casos, sendo então útil o uso da estatística simétrica $\gamma_{ij} + \gamma_{ji}$. Como exemplo, ilustra-se o método utilizado em duas séries temporais, uma estacionária e outra não-estacionária. Na figura 2.1,

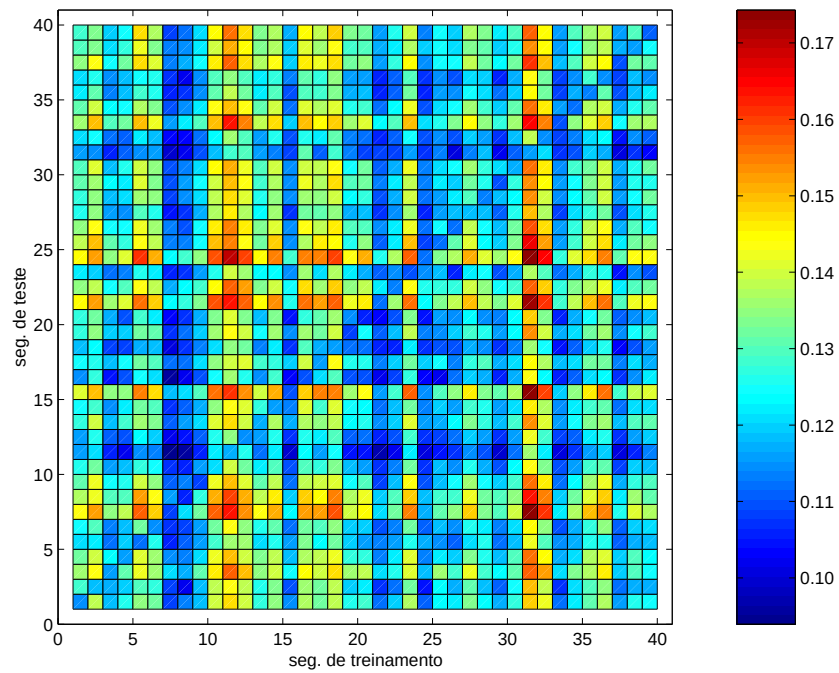
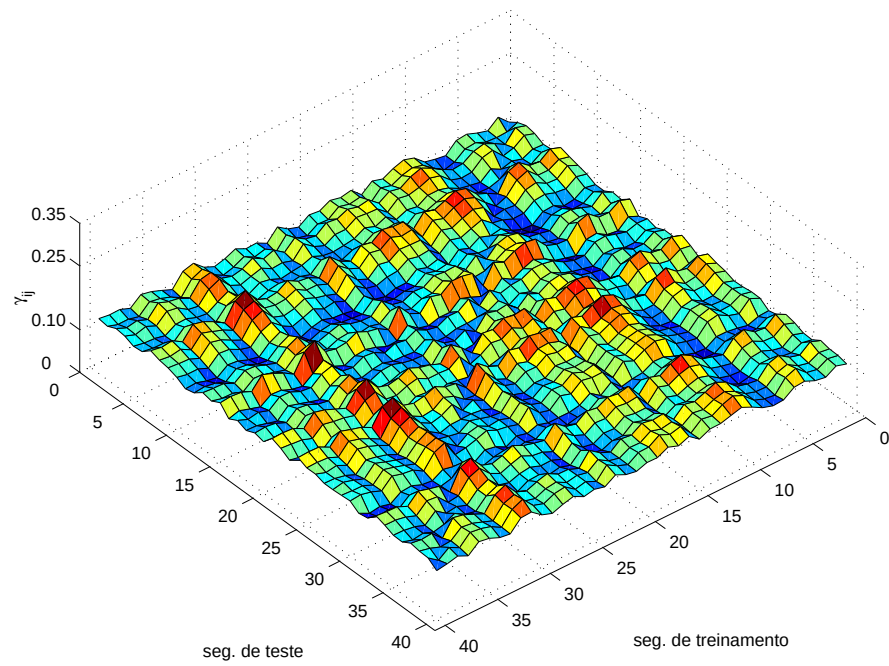


FIGURA 2.1: *Cross prediction error* aplicado nos segmentos obtidos a partir da série temporal associada ao mapeamento estacionário de Hénon.

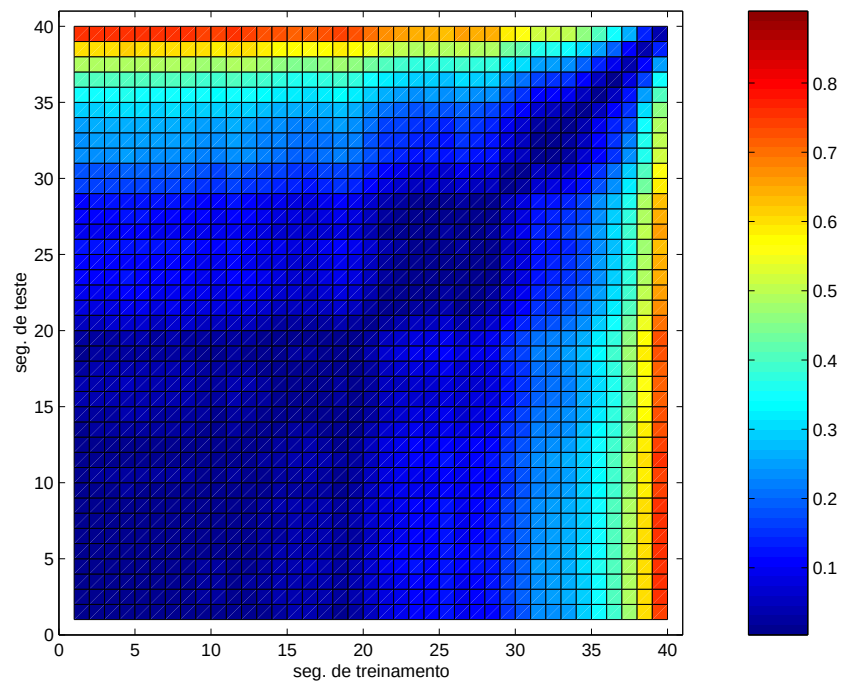
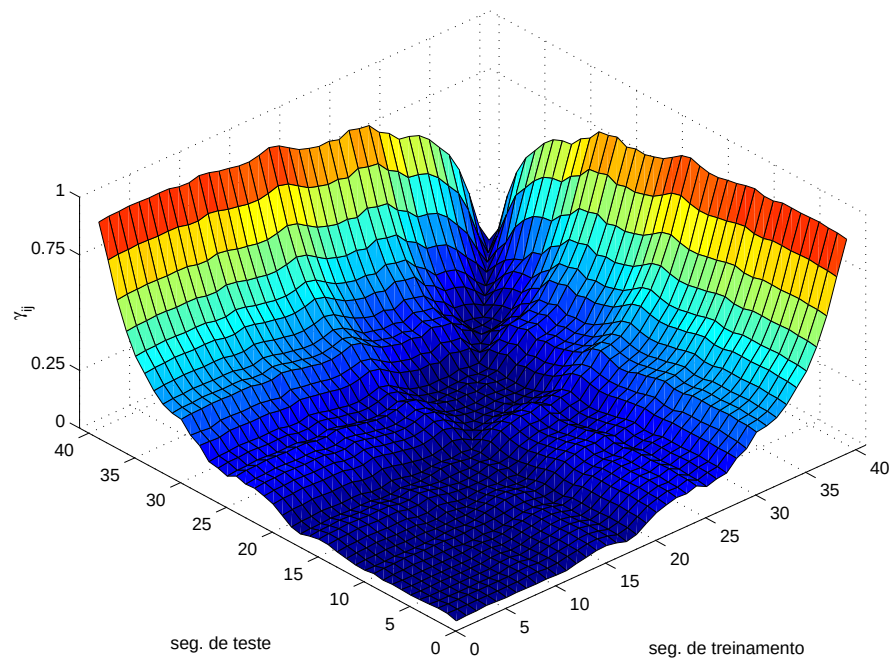


FIGURA 2.2: *Cross prediction error* aplicado nos segmentos obtidos a partir da série temporal associada ao mapeamento não-estacionário de Baker.

ilustra-se o *cross prediction error* aplicado em segmentos obtidos a partir da série $\{x_n\}_{n=1}^N$ associada ao mapeamento estacionário de Hénon [43], definido como

$$\begin{cases} x_{n+1} = 1 - ax_n^2 + y_n \\ y_{n+1} = bx_n \end{cases}, \quad (2.24)$$

onde $a = 1,4$ e $b = 0,3$. Na figura 2.2, mostra-se o método aplicado em segmentos obtidos a partir da série $\{u_n + v_n\}_{n=1}^N$ associada ao mapeamento não-estacionário de Baker [5], definido como

$$\text{se } v_n \leq \alpha : \begin{cases} u_{n+1} = \beta u_n \\ v_{n+1} = v_n / \alpha \end{cases} ; \quad \text{se } v_n > \alpha : \begin{cases} u_{n+1} = 0,5 + \beta u_n \\ v_{n+1} = (v_n - \alpha) / (1 - \alpha) \end{cases}, \quad (2.25)$$

onde $v_n \in [0, 1]$, $\alpha = 0,4$ e $\beta = n/N$.

Observa-se que os termos $\gamma(\mathcal{S}_i^\ell, \mathcal{S}_j^\ell)$ são tipicamente menores que os com $i \neq j$, pois os segmentos de treinamento e teste são idênticos. A estatística γ_{ij} é então representada em uma superfície de erro, onde valores superiores a 0,25 são considerados altos.¹⁰ Este limiar é válido nos dois casos, pois as séries temporais passam por um processo de normalização antes da aplicação do teste. Para uma série não-estacionária, do ponto de vista dinâmico, a previsibilidade decairá com a distância temporal entre os segmentos. Verifica-se este comportamento na aplicação do método para o mapeamento de Baker. Os parâmetros utilizados na aplicação do teste para os dois exemplos são $\ell = N/40$, $m = 2$ e $\varepsilon = 0,15$.

2.4 Testes de Linearidade

2.4.1 Estatística BDS

Originalmente, a estatística BDS (Brock, Dechert e Sheinkman) teve como motivação o interesse em detectar observações geradas por mecanismos determinísticos. Todavia, o teste proposto acabou revelando-se útil como um teste diagnóstico de resíduos para se verificar o ajuste de modelos lineares.

No capítulo 1, apresentou-se o cálculo da dimensão de correlação por meio da integral de correlação definida pela equação (1.19). Esta medida satisfaz a condição assintótica

$$\lim_{\ell \rightarrow +\infty} C_{m,\ell}(\varepsilon) = C_m(\varepsilon).$$

¹⁰Para esta escolha, considerou-se aplicações do método realizadas em séries dinamicamente estacionárias e a fragilidade do estimador (2.22) utilizado no teste.

Se a série $\{\tilde{x}_t\}$ for realização de um processo estacionário de variáveis aleatórias independentes com distribuição f , escreve-se

$$C_m(\varepsilon) = \int_{\vec{x}} \int_{\vec{x}'} \Theta_{\varepsilon}(\|\vec{x} - \vec{x}'\|) df_m(\vec{x}) df_m(\vec{x}') ,$$

onde

$$\Theta_{\varepsilon}(\|\vec{x} - \vec{x}'\|) = \Theta(\varepsilon - \|\vec{x} - \vec{x}'\|) = \prod_{i=1}^m \Theta_{\varepsilon}(|x_i - x'_i|) ,$$

define a função degrau de Heaviside, $\vec{x}$ é um vetor m -dimensional determinado a partir da reconstrução de Takens da série, f_m denota a função de distribuição do vetor $\vec{x}$ e satisfaz $f_m(\vec{x}) = f_m(x_1, x_2, \dots, x_m) = \prod_{k=1}^m f(x_k)$. Desta forma, obtém-se

$$\begin{aligned} C_m(\varepsilon) = & \int_{x_1} \int_{x_2} \dots \int_{x_m} \int_{x'_1} \int_{x'_2} \dots \int_{x'_m} \Theta_{\varepsilon}(|x_1 - x'_1|) \Theta_{\varepsilon}(|x_2 - x'_2|) \dots \\ & \dots \Theta_{\varepsilon}(|x_m - x'_m|) df(x_1) df(x_2) \dots df(x_m) df(x'_1) df(x'_2) \dots df(x'_m) , \end{aligned}$$

que pode ser reescrita como

$$C_m(\varepsilon) = \int_{x_1} \int_{x'_1} \Theta_{\varepsilon}(|x_1 - x'_1|) df(x_1) df(x'_1) \dots \int_{x_m} \int_{x'_m} \Theta_{\varepsilon}(|x_m - x'_m|) df(x_m) df(x'_m) .$$

No que, finalmente, implica

$$C_m(\varepsilon) = [C_1(\varepsilon)]^m . \quad (2.26)$$

A BDS tem como hipótese nula que o processo gerador dos dados é independente e identicamente distribuído (*i.i.d.*); a hipótese alternativa não é especificada. O teste baseia-se no fato de que se os dados forem *i.i.d.*, em condições assintóticas, tem-se

$$\sqrt{\ell} \{C_{m,\ell}(\varepsilon) - [C_1(\varepsilon)]^m\} \xrightarrow{d} N\{0, [\sigma_m(\varepsilon)]^2\} ,$$

e “ $\xrightarrow{d}$ ” significa convergência em distribuição (teorema de Denker e Keller [44]). A estatística BDS para amostras finitas é então dada por

$$W_{m,\ell}(\varepsilon) = \frac{\sqrt{\ell} \{C_{m,\ell}(\varepsilon) - [C_{1,\ell}(\varepsilon)]^m\}}{V_{m,\ell}(\varepsilon)} , \quad (2.27)$$

onde $[V_{m,\ell}(\varepsilon)]^2$ é a estimativa amostral e não viesada da variância $[\sigma_m(\varepsilon)]^2$.

Aplica-se a estatística BDS sobre a série dos resíduos obtidos a partir do modelo linear ARIMA (modelo auto-regressivo integrado e de médias móveis) que melhor se ajusta à série temporal analisada. Brock *et al.* [37], demonstraram que se os resíduos forem *i.i.d.*, situação em que a modelagem linear é adequada no ajuste da série temporal, tem-se que $C_{m,\ell}(\varepsilon) \rightarrow [C_{1,\ell}(\varepsilon)]^m$. Verifica-se a aceitação ou a rejeição da

hipótese nula, comparando-se $W_{m,\ell}(\varepsilon)$ com os valores críticos determinados a partir da distribuição normal $N(0, 1)$. Caso a hipótese nula seja rejeitada, indica-se uma não-linearidade presente na série temporal analisada (ou uma forma sutil de não-estacionariedade resistente à diferenciação). Em qualquer especificação de modelo linear ARIMA(p, d, q), somente são permitidos resíduos que sejam independentes e identicamente distribuídos; conclui-se então que o ajuste linear não é adequado para a modelagem da série, sendo necessário então um ajuste não-linear.¹¹

Para aplicação da estatística BDS nas séries temporais utilizadas neste trabalho, o parâmetro ε foi definido como a metade do desvio padrão da série dos resíduos provenientes da modelagem linear. A implementação do algoritmo utilizado para o cálculo de $W_{m,\ell}(\varepsilon)$ é devido a LeBaron [45].

2.4.2 Método *Surrogate Data*

O método *surrogate data* foi introduzido por Theiler *et al.* [39], e consiste em verificar, utilizando-se do espectro de potências de Fourier, se a dinâmica de uma série temporal é descrita por uma especificação linear.

Para um processo linear, o espectro de potências é unicamente descrito pelos coeficientes que caracterizam os operadores auto-regressivo e de médias móveis e, ainda mais importante, todos os momentos são completamente determinados pelo espectro. Portanto, o espectro de potências de Fourier pode ser considerado como uma completa descrição estatística do processo [4].

O teste consiste em utilizar algum observável não-linear λ da série original e logo após estimar uma distribuição de probabilidade $f(\lambda)$, usando réplicas lineares estocásticas (*surrogates*) obtidas a partir de um processo de randomização da transformada de Fourier da série original. Assumindo a hipótese nula de linearidade, constrói-se um *ensemble* de *surrogates* impondo com que eles preservem o espectro de potências e que sejam especificados pela hipótese nula, ou seja, obtém-se séries temporais lineares que possuem as mesmas propriedades estatísticas que caracterizam o processo original linear. O método *surrogate data* é basicamente uma aplicação da técnica de “*bootstrap*” proveniente da estatística moderna.

Os passos para aplicação do teste de hipóteses são descritos a seguir. Calcula-se a estatística λ_0 para o sinal original e $\{\lambda_j\}_{j=1}^K$ para os K *surrogates*, verificando logo após se λ_0 pertence à distribuição de λ obtida a partir dos valores λ_j . Se λ seguir uma distribuição gaussiana, estima-se a média $\langle \lambda \rangle = K^{-1} \sum_{j=1}^K \lambda_j$ e o desvio padrão da média $\delta \langle \lambda \rangle = [(K - 1)^{-1} \sum_{j=1}^K (\lambda_j - \langle \lambda \rangle)^2]^{1/2}$, transformando a distribuição por

¹¹Para esta finalidade, optou-se pela utilização de redes neurais artificiais. No capítulo 4, descreve-se a aplicação desta técnica para modelagem de séries temporais não-lineares.

meio de

$$\lambda'_j = \frac{\lambda_j - \langle \lambda \rangle}{\delta \langle \lambda \rangle} \sim t_{K-1} ,$$

que finalmente permite comparar λ'_0 com o limiar crítico da distribuição t de Student com $K - 1$ graus de liberdade. Caso λ'_0 for maior que o limiar crítico do intervalo de confiança, rejeita-se a hipótese nula de linearidade no nível de significância adotado.

Os *surrogates* são construídos a partir da transformada discreta de Fourier da série temporal original $\{x_n\}_{n=1}^N$, definida como

$$\tilde{x}_k = \frac{1}{\sqrt{N}} \sum_{n=1}^N x_n e^{-i2\pi nk/N} , \quad (2.28)$$

para $0 < k \leq N$. Multiplicando as componentes $\tilde{x}_k$ por fases aleatórias, obtém-se $\tilde{x}'_k = \tilde{x}_k e^{i\phi_k}$, onde ϕ_k é uma variável uniformemente distribuída em $[0, 2\pi)$ e necessariamente $\phi_{N-k} = -\phi_k$, para que a transformada inversa seja real. Calcula-se então a transformada discreta inversa de Fourier, fazendo

$$x'_n = \frac{1}{\sqrt{N}} \sum_{k=1}^N \tilde{x}'_k e^{i2\pi nk/N} , \quad (2.29)$$

obtendo-se assim uma série linear com o espectro prescrito. Esta técnica é denominada como *phase randomization* e diferentes realizações $\{\phi_k\}_{k=1}^N$ geram novos *surrogates* com as mesmas propriedades espectrais da série temporal original $\{x_n\}_{n=1}^N$.

Finalmente, observa-se que a rejeição da hipótese nula de linearidade não serve para a detecção de caos. Este resultado somente indica uma evidência estatística de não-linearidade em uma série temporal estacionária, não necessariamente indicando determinismo na série temporal aplicada ao teste. Este método mostrou-se eficiente na detecção de não-linearidade em sistemas caóticos de baixa dimensionalidade [39]. Para maiores detalhes sobre as estatísticas utilizadas para aplicação do teste, aconselha-se [46].

Capítulo 3

Transformada em Ondeletas

3.1 Introdução

Nos anos 80, foi desenvolvida uma nova forma de decomposição de funções, baseada nas representações quadráticas integráveis e na teoria de grupos, denominada transformada em ondeletas. Matematicamente, o termo “ondeletas” significa o conjunto de funções em forma de ondas (limitadas no espaço) geradas a partir de dilatações e translações de uma função geradora simples $\psi(t)$, que é a ondeleta-mãe.

A inovação revolucionária dessa transformada consiste em introduzir uma “janela variável” que pode dilatar-se ou comprimir-se, dependendo da escala de análise. Desde então, suas propriedades têm sido amplamente exploradas por vários autores, tanto do ponto de vista teórico, veja por exemplo [47], quanto de aplicações à análise de sinais [48].

A principal característica da transformada em ondeletas é o fato de ela ser local, pois tem suporte compacto e é bem definida em uma região do espaço, em oposição à transformada de Fourier, que é global. Este fato torna a transformada em ondeletas ideal para analisar sinais não-estacionários, contendo transitoriedades e estruturas tipo fractais [4]. Bases de Fourier são localizadas em frequência, mas não no tempo: pequenas mudanças nas observações podem provocar mudanças em todas as componentes de uma expansão em Fourier, o que não acontece com uma expansão em série de ondeletas. A diferença é que as funções de uma base de ondeletas são indexadas por dois parâmetros, ao passo que na base de Fourier temos um único parâmetro, ω , que tem a interpretação física de frequência.

A idéia, tanto na análise de Fourier quanto na análise em ondeletas, consiste em aproximar qualquer função por uma combinação linear de senos e cossenos ou ondeletas, respectivamente. Funções com descontinuidades e singularidades necessitarão de menos ondeletas do que senos e cossenos, para uma aproximação comparável.

A transformada de Fourier de um sinal $f(t)$ pode ser escrita como

$$\mathcal{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt, \quad (3.1)$$

indicando o peso relativo com que a frequência angular ω comparece na composição de $f(t)$. O espectro de potências de Fourier $P(\omega)$ é definido como o módulo quadrado de $\mathcal{F}(\omega)$, ou seja,

$$P(\omega) = |\mathcal{F}(\omega)|^2, \quad (3.2)$$

indicando que uma dada distribuição de densidade espectral associa determinada quantidade particular de energia a cada frequência do espectro, porém não traz informação da sua localização temporal.

A transformada em ondeletas contínua de um sinal $f(t)$ é dada por

$$\mathcal{W}_\psi(\gamma, \tau) = \int_{-\infty}^{+\infty} f(t) \psi_{\gamma\tau}^*(t) dt, \quad (3.3)$$

onde

$$\psi_{\gamma\tau}(t) = \frac{1}{\sqrt{\gamma}} \psi\left(\frac{t-\tau}{\gamma}\right). \quad (3.4)$$

Nesta representação, $\gamma \in \mathbb{R}^+$ é o parâmetro de dilatação de escala que corresponde à largura da ondeleta e, $\tau \in \mathbb{R}$ é o parâmetro de translação, que indica a posição da ondeleta. O conjunto de funções $\{\psi_{\gamma\tau}(t)\}$ é a família de ondeletas transladadas e dilatadas continuamente, originadas a partir da ondeleta-mãe $\psi(t)$.

A transformada inversa é então definida como

$$f(t) = \int_{0^+}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{W}_\psi(\gamma, \tau) \psi_{\gamma\tau}(t) d\tau \frac{d\gamma}{\gamma^2}. \quad (3.5)$$

Percebe-se que a transformada em ondeletas atua como uma análise harmônica, expressando a função $f(t)$ como uma superposição de contribuições que têm todas a mesma forma que a ondeleta geradora.

A transformada de Fourier não é suficiente para caracterizar sinais cuja frequência varia temporalmente. O estudo espectral desses processos requer o emprego de transformadas que permitam obter a frequência como função do tempo. O objetivo deste tipo de análise é expandir o sinal em formas de ondas cujas propriedades estejam bem adaptadas à sua estrutura local; a análise em ondeletas cumpre muito bem esta tarefa. A figura 3.1 ilustra as transformadas de Fourier e ondeletas para dois sinais compostos por senóides de frequências ω_1 e ω_2 , porém com comportamentos temporais distintos.

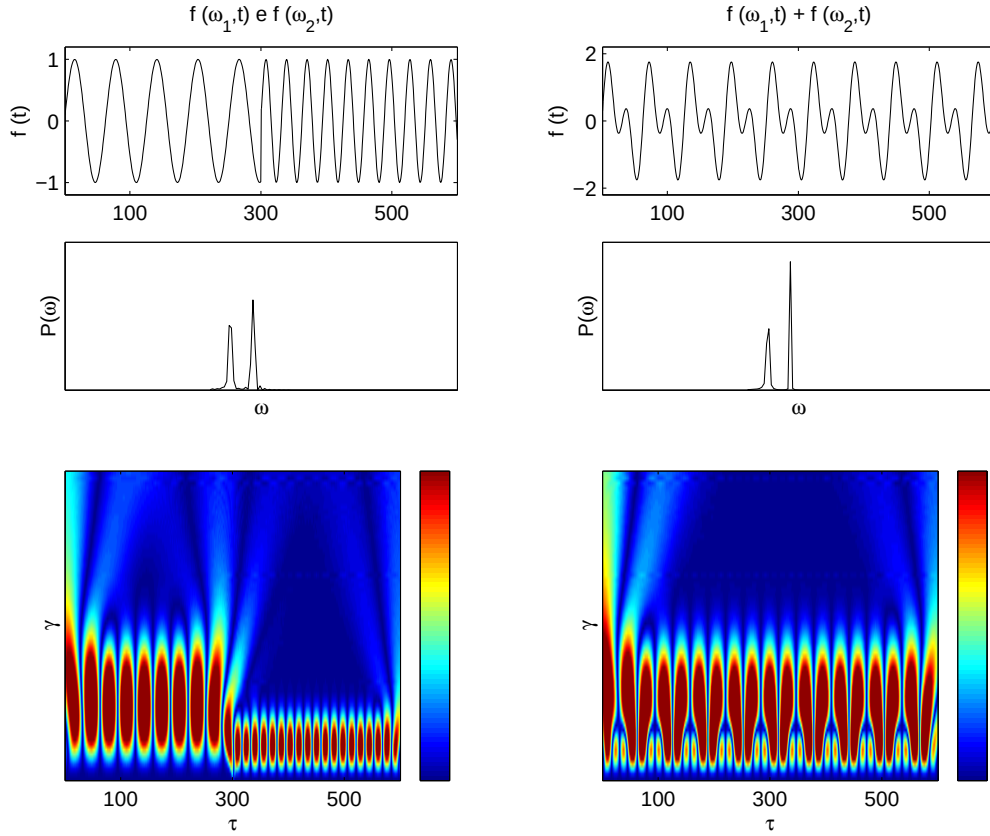


FIGURA 3.1: Comparação ilustrativa entre as transformadas de Fourier e ondeletas para dois sinais compostos por senóides de frequências ω_1 e ω_2 com comportamentos temporais distintos. Neste exemplo $\omega_2 = 2\omega_1$.

A transformada em ondeletas possui as seguintes propriedades: a) conservação de energia, tanto local quanto globalmente, o que garante que não existe perda de informação quando se transforma o sinal; b) localização tempo-escala, no que difere da transformada de Fourier; c) núcleo reprodutor, cuja estrutura depende da escolha da ondeleta-mãe; e d) análise de regularidade local, se $\{\psi_{\gamma\tau}(t)\}$ é uma família de funções regulares, então, a transformada inversa também será regular.

A ondeleta-mãe $\psi(t)$, de acordo com [49], satisfaz três propriedades básicas: a) similaridade, a família de ondeletas é gerada pela translação e dilatação de uma única função $\psi(t)$ e todas as ondeletas serão mutuamente similares; b) admissibilidade, para uma função integrável significa que sua média é nula; e c) $\psi(t)$ deve ser normalizável, ou seja,

$$\int_{-\infty}^{+\infty} \psi(t) \cdot \psi^*(t) dt = 1.$$

Define-se a variância de ondeleta como a integral dos coeficientes quadráticos de ondeleta sobre o parâmetro de translação τ , ou seja,

$$V_\psi(\gamma) = \int_{-\infty}^{+\infty} |\mathcal{W}_\psi(\gamma, \tau)|^2 d\tau . \quad (3.6)$$

Essa quantidade também pode ser denominada espectro de potências de ondeleta, sendo que o nome tem sua origem na comparação com o espectro de potências de Fourier.

A energia total da função $f(t)$, que é conservada, denota-se por

$$E_f = \int_{-\infty}^{+\infty} |f(t)|^2 dt = \int_{0+}^{+\infty} \int_{-\infty}^{+\infty} |\mathcal{W}_\psi(\gamma, \tau)|^2 d\tau \frac{d\gamma}{\gamma^2} , \quad (3.7)$$

ou, em outras palavras, a energia total pode ser reescrita na forma

$$E_f = \int_{0+}^{+\infty} V_\psi(\gamma) \frac{d\gamma}{\gamma^2} . \quad (3.8)$$

Observa-se que a variância de ondeleta $V_\psi(\gamma)$ pode ser interpretada como a contribuição da energia correspondente da escala γ à energia total da função $f(t)$. A representação gráfica de $|\mathcal{W}_\psi(\gamma, \tau)|^2$ no semiplano (γ, τ) é denominada usualmente de escalograma.

Nota-se que, como a transformada em ondeletas mede o grau de similaridade entre a ondeleta $\psi_{\gamma\tau}(t)$ e a função $f(t)$, a escolha de ondeletas-mãe diferentes geram resultados diferentes quando aplicados à mesma $f(t)$.

3.2 Análise de Multirresolução

3.2.1 Transformada em Ondeletas Discreta

A transformada em ondeletas, na sua forma discreta, pode ser escrita como

$$d_{jk} = \int_{-\infty}^{+\infty} f(t) \psi_{jk}^*(t) dt , \quad (3.9)$$

onde

$$\psi_{jk}(t) = \gamma_0^{j/2} \psi(\gamma_0^j t - k\tau_0) , \quad (3.10)$$

o que corresponde a tomar $\gamma = \gamma_0^{-j}$ e $\tau = k \times \tau_0 \times \gamma_0^{-j}$, indicando que a translação τ depende do parâmetro de dilatação γ .

Na análise de Fourier, toda função periódica $f(t)$, com período 2π e de quadrado integrável, ou seja, $f(t) \in \mathcal{L}^2(0, 2\pi)$, é gerada por uma superposição de exponenciais

complexas obtidas por dilatações da função $e^{i\omega t}$. O objetivo é estender essa idéia para $\mathcal{L}^2(\mathbb{R})$, isto é, gerar esse espaço a partir de uma única função $\psi(t)$. Considere o espaço $\mathcal{L}^2(\mathbb{R})$ de todas funções mensuráveis de quadrado integrável sobre $\mathbb{R}$. Neste espaço, as funções $\{\psi_{jk}(t)\}$ devem cair rapidamente a zero quando $t \rightarrow \pm\infty$, logo exponenciais do tipo $e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$ não pertencem a esse espaço. A idéia é considerar dilatações e translações de uma única função $\psi(t)$, de modo a cobrir $\mathbb{R}$. Ou seja, considera-se ondeletas do tipo

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k) . \quad (3.11)$$

A ondeleta $\psi_{jk}(t)$ é obtida de uma ondeleta-mãe $\psi(t)$ por dilatação binária 2^{-j} ($\gamma_0 = 2$) e translação diádica $k \times 2^{-j}$ ($\tau_0 = 1$). As funções $\{\psi_{jk}(t)\}$ formam uma base ortonormal completa no espaço $\mathcal{L}^2(\mathbb{R})$, tem-se então

$$\int_{-\infty}^{+\infty} \psi_{jk}(t) \cdot \psi_{mn}^*(t) dt = \delta_{jm} \cdot \delta_{kn} , \quad (3.12)$$

onde δ_{jm} e δ_{kn} são funções delta de Kronecker. Desta forma, escreve-se qualquer função $f(t) \in \mathcal{L}^2(\mathbb{R})$ como uma combinação linear de todas as dilatações e translações da ondeleta-mãe $\psi(t)$, obtém-se então

$$f(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(t) . \quad (3.13)$$

Diz-se que a função $f(t)$ é caracterizada pelos coeficientes de ondeletas $\{d_{jk}\}$ expressos na base $\{\psi_{jk}(t)\}$, que é normalmente denominada base de Riesz.

Uma maneira de gerar ondeletas é pela função escala $\phi(t)$, que é uma solução da equação

$$\phi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} \ell_k \phi(2t - k) . \quad (3.14)$$

Essa função gera uma família ortonormal em $\mathcal{L}^2(\mathbb{R})$, definida como

$$\phi_{jk}(t) = 2^{j/2} \phi(2^j t - k) . \quad (3.15)$$

Nessas condições, define-se a ondeleta-mãe $\psi(t)$ como

$$\psi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \phi(2t - k) . \quad (3.16)$$

Os coeficientes ℓ_k e h_k são geralmente conhecidos como coeficientes de filtro. Estes são dados por

$$\ell_k = \langle \phi(t), \phi_{1k}(t) \rangle = \sqrt{2} \int_{-\infty}^{+\infty} \phi(t) \cdot \phi^*(2t - k) dt , \quad (3.17)$$

$$h_k = \langle \psi(t), \phi_{1k}(t) \rangle = \sqrt{2} \int_{-\infty}^{+\infty} \psi(t) \cdot \phi^*(2t - k) dt . \quad (3.18)$$

As equações (3.14) e (3.16) são chamadas equações de dilatação e serão fundamentais para decomposição e reconstrução dos subespaços de aproximação e detalhe apresentados na seqüência do capítulo. Para fixar a notação utilizada, observa-se que as funções $\phi(t)$ e $\psi(t)$ são definidas como $\phi_{00}(t)$ e $\psi_{00}(t)$, respectivamente.

3.2.2 Subespaços de Aproximação

Para efetuar a representação multirresolução de uma função em $\mathcal{L}^2(\mathbb{R})$, necessita-se de uma seqüência crescente de subespaços fechados que verificam

$$\dots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots ,$$

estes subespaços V_j , para $j \in \mathbb{Z}$, devem ter uma intersecção trivial e uma união que é densa em $\mathcal{L}^2(\mathbb{R})$, ou seja,

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \text{ e } \overline{\bigcup_{j \in \mathbb{Z}} V_j} = \mathcal{L}^2(\mathbb{R}) .$$

A função escala $\phi(t)$ e as suas translações $\phi(t - k)$, com $k \in \mathbb{Z}$, formam uma base do subespaço V_0 . Da mesma maneira, a família de funções $\{\phi_{1m}(t) : m \in \mathbb{Z}\}$ forma uma base para o subespaço V_1 . Como V_0 está contido em V_1 , qualquer função definida em V_0 pode ser expressa como combinação linear das funções que constituem a base do subespaço V_1 , obtém-se então

$$\phi(t) = \sum_{k \in \mathbb{Z}} \langle \phi(t), \phi_{1k}(t) \rangle \phi_{1k}(t) .$$

Essa equação é a mesma que a (3.14).

Para obter a aproximação $f_{j-1}(t)$ de uma função $f_j(t) \in V_j \subset \mathcal{L}^2(\mathbb{R})$ efetua-se a projeção de $f_j(t)$ no subespaço V_{j-1} , obtém-se desta forma

$$f_{j-1}(t) = \sum_{k \in \mathbb{Z}} a_{j-1,k} \phi_{j-1,k}(t) , \quad (3.19)$$

esta função representa uma aproximação com escala indexada por $j - 1$ da função com escala indexada por j . Os coeficientes $\{a_{j-1,k}\}$, denominados coeficientes de aproximação, são determinados fazendo

$$a_{j-1,k} = \int_{-\infty}^{+\infty} f_j(t) \phi_{j-1,k}^*(t) dt , \quad (3.20)$$

ou

$$a_{j-1,k} = \sum_{m \in \mathbb{Z}} a_{jm} \int_{-\infty}^{+\infty} \phi_{jm}(t) \cdot \phi_{j-1,k}^*(t) dt ,$$

que pode ser reescrita como

$$a_{j-1,k} = \sum_{m \in \mathbb{Z}} a_{jm} \int_{-\infty}^{+\infty} 2^{(2j-1)/2} \phi(2^j t - m) \cdot \phi^*(2^{j-1} t - k) dt .$$

Fazendo a mudança de coordenadas $2^{j-1} t - k \rightarrow t'$, obtém-se

$$a_{j-1,k} = \sum_{m \in \mathbb{Z}} a_{jm} \left\{ \sqrt{2} \int_{-\infty}^{+\infty} \phi(t') \cdot \phi^*[2t' - (m - 2k)] dt' \right\}^* ,$$

e finalmente, usando a equação (3.17), a relação geral para os coeficientes de aproximação será dada por

$$a_{j-1,k} = \sum_{m \in \mathbb{Z}} \ell_{m-2k}^* a_{jm} . \quad (3.21)$$

3.2.3 Subespaços de Detalhe

A análise de multirresolução conduz diretamente a uma decomposição do espaço $\mathcal{L}^2(\mathbb{R})$. Como o próprio nome indica, tal análise consiste em representar uma dada função em diferentes escalas de resolução. Permite também obter, de uma forma eficaz, a informação necessária para passar de uma escala para outra. Para tanto, define-se agora o subespaço W_{j-1} como sendo o complemento ortogonal de V_{j-1} em V_j . Desta forma, tem-se

$$V_j = V_{j-1} \oplus W_{j-1} , \quad (3.22)$$

com $V_{j-1} \perp W_{j-1}$ e $\oplus$ representa uma soma ortogonal direta. Daqui, verifica-se que os subespaços W_j , com $j \in \mathbb{Z}$, são ortogonais e

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = \bigoplus_{j \in \mathbb{Z}} W_j = \mathcal{L}^2(\mathbb{R}) .$$

Esta soma possibilita a equação (3.13).

A ondeleta-mãe $\psi(t)$ em conjunto com todas as suas translações inteiras deverão constituir uma base do subespaço W_0 . Como $W_0 \subset V_1$, a função $\psi(t)$ pode ser expressa como combinação linear das funções que constituem a base de V_1 . Escreve-se

$$\psi(t) = \sum_{k \in \mathbb{Z}} \langle \psi(t), \phi_{1k}(t) \rangle \phi_{1k}(t) .$$

Essa equação é a mesma que a (3.16).

Os coeficientes de filtro $\{\ell_k\}$ e $\{h_k\}$ devem estar relacionados de modo a garantir a desejada ortogonalidade entre os subespaços, $V_0 \perp W_0$. Adiante, mostra-se a

relação necessária entre os coeficientes de filtro no caso das famílias de ondeletas de Daubechies.¹²

Para uma determinada aproximação $f_j(t) \in V_j \subset \mathcal{L}^2(\mathbb{R})$, não há a necessidade de expansão nos infinitos subespaços de detalhe que compõem $\mathcal{L}^2(\mathbb{R})$; expande-se $f_j(t)$ somente nos subespaços que compõem V_j . Define-se V_j como

$$V_j = W_{j-1} \oplus W_{j-2} \oplus W_{j-3} \oplus \dots, \quad (3.23)$$

essa equação é a mesma que a (3.22) e obedece à condição $V_{j-1} \subset V_j$. Esta condição implica que ao passar do nível de resolução $j-1$ para o nível de resolução j , aumenta-se a resolução (ou adiciona-se “detalhes”). Na medida em que o nível de resolução aumenta, a função aproximada converge para a função original, ou seja, $\lim_{j \rightarrow +\infty} f_j(t) = f(t)$. Por outro lado, quando $j \rightarrow -\infty$, a aproximação converge para a função nula. Na figura 3.2, é apresentada uma decomposição ortogonal da função $f_j(t) \in V_j$ nos subespaços V_{j-1} e W_{j-1} .

Para qualquer função $f(t) \in \mathcal{L}^2(\mathbb{R})$ e pelo modo de como é construído o espaço $\mathcal{L}^2(\mathbb{R})$, escreve-se $f(t)$ como

$$f(t) = \dots + g_{-2}(t) + g_{-1}(t) + g_0(t) + g_1(t) + g_2(t) + \dots, \quad (3.24)$$

onde $g_j(t) \in W_j$ para todos $j \in \mathbb{Z}$. A função $g_j(t)$, de acordo com a equação (3.13), pode ser escrita como

$$g_j(t) = \sum_{k \in \mathbb{Z}} d_{jk} \psi_{jk}(t). \quad (3.25)$$

As funções $\{\psi_{jk}(t)\}$, com $k \in \mathbb{Z}$, formam, para um dado valor do parâmetro j , uma base do subespaço W_j . Os subespaços fechados W_j e W_m pertencentes ao espaço $\mathcal{L}^2(\mathbb{R})$ são mutuamente ortogonais, ou seja,

$$\langle g_j(t), g_m(t) \rangle = \sum_{k \in \mathbb{Z}} d_{jk} d_{mk}^* \delta_{jm},$$

para $g_j(t) \in W_j$, $g_m(t) \in W_m$ e $W_j \perp W_m$ quando $j \neq m$.

A diferença $g_{j-1}(t) \in W_{j-1}$ entre a função $f_j(t) \in V_j$ e a aproximação $f_{j-1}(t) \in V_{j-1}$ pode ser expressa em termos da projeção de $f_j(t)$ no subespaço de detalhe W_{j-1} . A função $g_{j-1}(t)$ é escrita como

$$g_{j-1}(t) = \sum_{k \in \mathbb{Z}} d_{j-1,k} \psi_{j-1,k}(t). \quad (3.26)$$

¹²Daubechies [50], introduziu as ondeletas ortonormais de suporte compacto, denominadas *daubelets* dbN. Estas ondeletas não possuem expressão explícita e são usualmente denotadas por $_N\psi(t)$. O parâmetro $N \in \mathbb{Z}$ controla a regularidade da função.

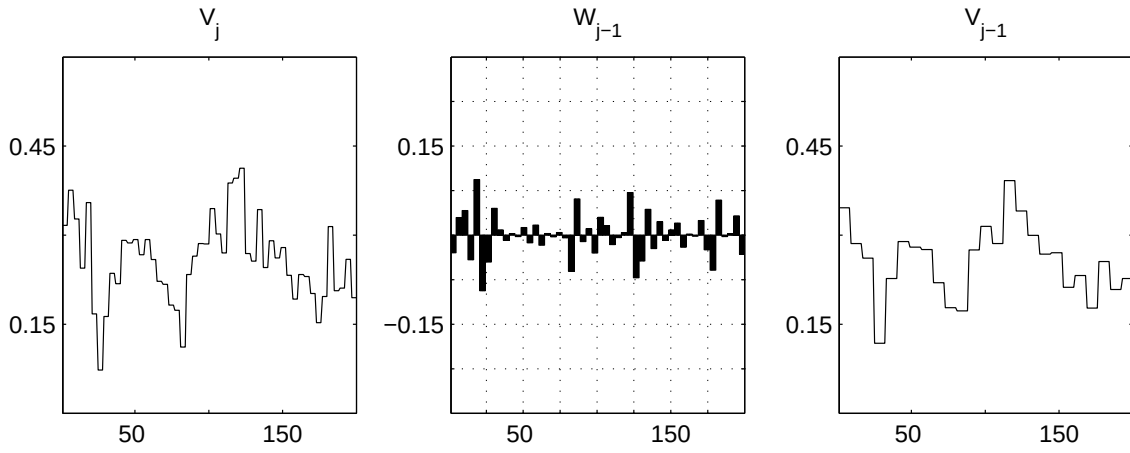


FIGURA 3.2: Decomposição ortogonal da função $f_j(t) \in V_j$ nos subespaços V_{j-1} e W_{j-1} . Observa-se que a função $f_j(t)$ apresenta maior resolução que a função $f_{j-1}(t)$.

A família de ondeletas $\{\psi_{j-1,k}(t) : k \in \mathbb{Z}\}$ forma uma base ortonormal em W_{j-1} e a função $g_{j-1}(t)$ representa o detalhe de $f_j(t)$ em relação à aproximação $f_{j-1}(t)$. Os coeficientes $\{d_{j-1,k}\}$, denominados coeficientes de ondeleta, são determinados fazendo o produto interno

$$d_{j-1,k} = \int_{-\infty}^{+\infty} f_j(t) \psi_{j-1,k}^*(t) dt, \quad (3.27)$$

ou

$$d_{j-1,k} = \sum_{m \in \mathbb{Z}} a_{jm} \int_{-\infty}^{+\infty} \phi_{jm}(t) \cdot \psi_{j-1,k}^*(t) dt,$$

que pode ser reescrita como

$$d_{j-1,k} = \sum_{m \in \mathbb{Z}} a_{jm} \int_{-\infty}^{+\infty} 2^{(2j-1)/2} \phi(2^j t - m) \cdot \psi^*(2^{j-1} t - k) dt.$$

Fazendo a mudança de coordenadas $2^{j-1} t - k \rightarrow t'$, obtém-se

$$d_{j-1,k} = \sum_{m \in \mathbb{Z}} a_{jm} \left\{ \sqrt{2} \int_{-\infty}^{+\infty} \psi(t') \cdot \phi^*[2t' - (m - 2k)] dt' \right\}^*,$$

e por fim, usando a equação (3.18), a relação geral para os coeficientes de ondeleta será dada por

$$d_{j-1,k} = \sum_{m \in \mathbb{Z}} h_{m-2k}^* a_{jm}. \quad (3.28)$$

Finalmente, verifica-se a equação (3.9) fazendo

$$\begin{aligned} f_{j+1}(t) &= f_{j+2}(t) - g_{j+1}(t) , \\ f_{j+1}(t) &= f_{j+3}(t) - g_{j+2}(t) - g_{j+1}(t) , \\ &\vdots \\ f_{j+1}(t) &= f_{+\infty}(t) - \sum_{m \geq j+1} g_m(t) . \end{aligned}$$

Como já mencionado anteriormente, na análise de multirresolução verifica-se que $\lim_{j \rightarrow +\infty} f_j(t) = f(t)$ e o produto interno $d_{jk} = \langle f_{j+1}(t), \psi_{jk}(t) \rangle$ pode ser reescrito como $\langle f(t), \psi_{jk}(t) \rangle$. Isto é possível, pois $\langle \psi_{mn}(t), \psi_{jk}(t) \rangle = 0$, para $m \geq j + 1$.

3.2.4 Ondeletas de Daubechies

O suporte compacto das funções ondeletas é uma propriedade importante, pois está relacionado com o fato de as ondeletas serem localizadas no tempo. Quando pretende-se obter ondeletas com suporte compacto é necessário que apenas um número finito de coeficientes de filtro seja diferente de zero. Os sistemas de ondeletas de Daubechies são organizados em diferentes famílias, cada uma das quais caracterizada por um número diferente de coeficientes de filtro não nulos. A função escala para cada uma das famílias de ondeletas é então dada por

$${}_N\phi(t) = \sqrt{2} \sum_{k=0}^{2N-1} \ell_k {}_N\phi(2t - k) . \quad (3.29)$$

Funções escala assim definidas tomam valores diferentes de zero apenas num determinado intervalo. A este intervalo costuma-se denominar “suporte” da função. Para a função escala ${}_N\phi(t)$, escreve-se

$$\text{supp}_{{}_N\phi(t)} = [0, 2N - 1] .$$

Desta forma, observa-se que a regularidade da função ${}_N\phi(t)$ aumenta à medida que cresce o valor de N .

É por meio da imposição de determinadas condições sobre os valores que os coeficientes de filtro podem assumir que obtém-se certas propriedades para sistemas em ondeletas, tal como, por exemplo, a ortonormalidade. Para esta propriedade tem-se

$$h_m^* = (-1)^m \ell_{2N-1-m} , \quad (3.30)$$

e define a chamada *quadrature mirror filter relation*.

A ondeleta-mãe para cada uma das famílias de ondeletas pode ser definida como

$${}_N\psi(t) = \sqrt{2} \sum_{k=0}^{2N-1} h_k {}_N\phi(2t - k) , \quad (3.31)$$

a relação (3.30) é condicionada pela necessidade de garantir a ortonormalidade entre a função escala e a ondeleta-mãe. Verifica-se então

$$\begin{aligned} \langle {}_N\phi(t), {}_N\psi(t) \rangle &= 2 \int_{-\infty}^{+\infty} \sum_{k=0}^{2N-1} \ell_k {}_N\phi(2t - k) \sum_{m=0}^{2N-1} h_m^* {}_N\phi^*(2t - m) dt , \\ \langle {}_N\phi(t), {}_N\psi(t) \rangle &= 2 \int_{-\infty}^{+\infty} \sum_{k=0}^{2N-1} \ell_k {}_N\phi(2t - k) \sum_{m=0}^{2N-1} (-1)^m \ell_{2N-1-m} {}_N\phi^*(2t - m) dt , \\ \langle {}_N\phi(t), {}_N\psi(t) \rangle &= 2 \sum_{k=0}^{2N-1} \ell_k \sum_{m=0}^{2N-1} (-1)^m \ell_{2N-1-m} \int_{-\infty}^{+\infty} {}_N\phi(2t - k) \cdot {}_N\phi^*(2t - m) dt . \end{aligned}$$

Se durante a geração de ${}_N\phi(t)$ tiver sido garantida a ortonormalidade entre todas as translações inteiras da função, tem-se

$$2 \int_{-\infty}^{+\infty} {}_N\phi(2t - k) \cdot {}_N\phi^*(2t - m) dt = \int_{-\infty}^{+\infty} {}_N\phi(t' - k) \cdot {}_N\phi^*(t' - m) dt' = \delta_{km} ,$$

no que implica

$$\langle {}_N\phi(t), {}_N\psi(t) \rangle = \sum_{k=0}^{2N-1} (-1)^k \ell_k \ell_{2N-1-k} = 0 ,$$

como se pretendia demonstrar.

O suporte da ondeleta-mãe ${}_N\psi(t)$, definida em (3.31), é igual ao da função escala correspondente, ou seja,

$$\sup_{{}_N\psi(t)} = [0, 2N - 1] .$$

O sistema completo de funções é obtido considerando translações e dilatações da função escala e da ondeleta-mãe. Define-se, desta forma

$${}_N\phi_{jk}(t) = 2^{j/2} {}_N\phi(2^j t - k) \text{ e } {}_N\psi_{jk}(t) = 2^{j/2} {}_N\psi(2^j t - k) ,$$

de modo que

$$\sup_{{}_N\phi_{jk}(t)} = \sup_{{}_N\psi_{jk}(t)} = [k \times 2^{-j}, (2N - 1 + k) \times 2^{-j}] .$$

A família de ondeletas com $N = 4$ é utilizada para ilustrar os conceitos e exemplificar algumas das operações envolvendo ondeletas. Desta forma, com o auxílio

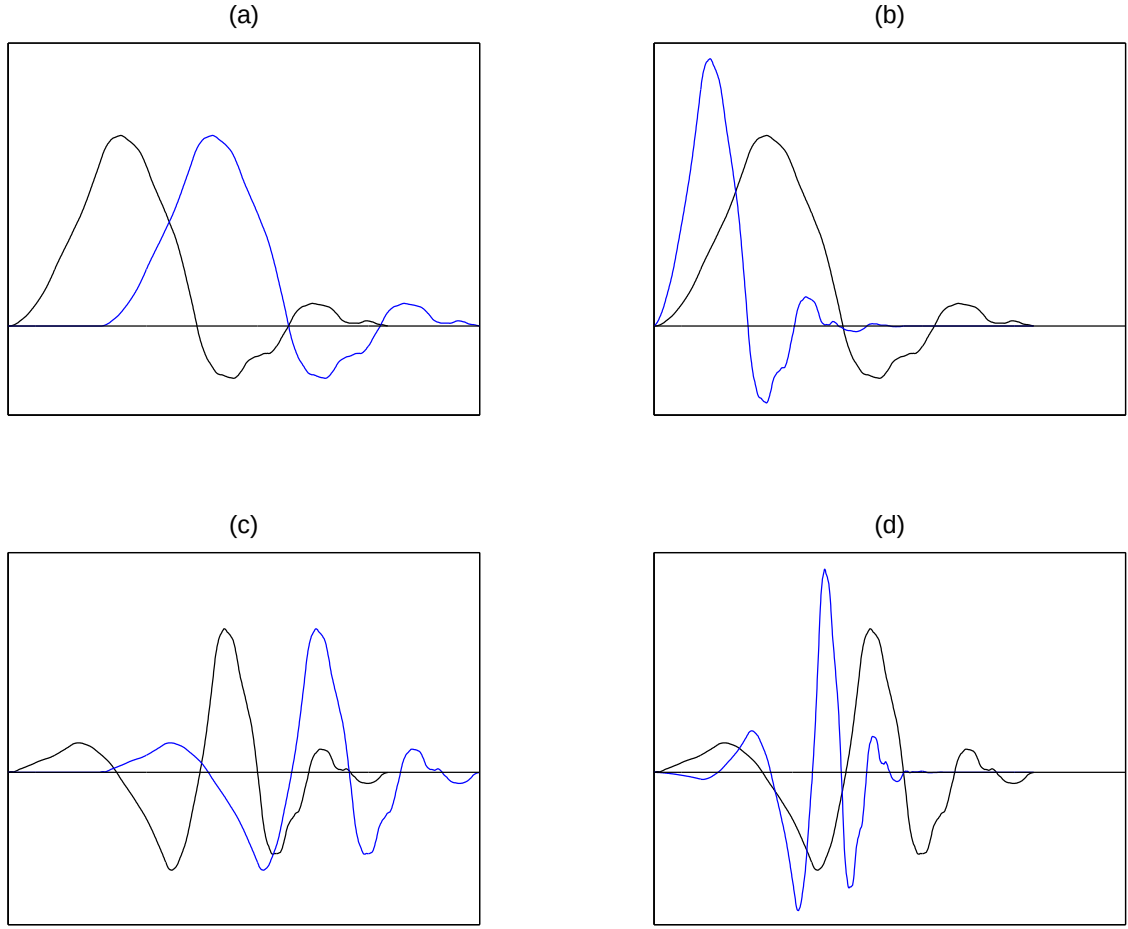


FIGURA 3.3: Translação (a) e dilatação (b) da função escala ${}_4\phi(t)$; translação (c) e dilatação (d) da ondeleta-mãe ${}_4\psi(t)$.

da relação (3.29), a função escala desta família de ondeletas é obtida por meio da resolução da equação

$$\begin{aligned} {}_4\phi(t) = & \sqrt{2} [\ell_0 {}_4\phi(2t) + \ell_1 {}_4\phi(2t-1) + \ell_2 {}_4\phi(2t-2) + \ell_3 {}_4\phi(2t-3) \\ & + \ell_4 {}_4\phi(2t-4) + \ell_5 {}_4\phi(2t-5) + \ell_6 {}_4\phi(2t-6) + \ell_7 {}_4\phi(2t-7)] . \end{aligned}$$

Os valores dos coeficientes ℓ_k , para $0 \leq k \leq 7$, são determinados de modo a garantir algumas propriedades para o sistema de ondeletas. No apêndice A, demonstra-se quais são as condições para obtenção destas propriedades e descreve-se o procedimento numérico proposto por Strang [51] na geração de funções escala.

Uma vez calculada a função escala, a ondeleta-mãe pode ser obtida por meio das

relações (3.30) e (3.31). Tem-se então

$${}_4\psi(t) = \sqrt{2} \sum_{k=0}^7 h_k {}_4\phi(2t-k) = \sqrt{2} \sum_{k=0}^7 (-1)^k \ell_{7-k}^* {}_4\phi(2t-k) ,$$

que pode ser reescrita como

$$\begin{aligned} {}_4\psi(t) = & \sqrt{2} [\ell_7^* {}_4\phi(2t) - \ell_6^* {}_4\phi(2t-1) + \ell_5^* {}_4\phi(2t-2) - \ell_4^* {}_4\phi(2t-3) \\ & + \ell_3^* {}_4\phi(2t-4) - \ell_2^* {}_4\phi(2t-5) + \ell_1^* {}_4\phi(2t-6) - \ell_0^* {}_4\phi(2t-7)] . \end{aligned}$$

Na figura 3.3 apresentam-se as funções ${}_4\phi_{01}(t)$ e ${}_4\phi_{10}(t)$. Estas funções foram obtidas a partir de uma translação inteira de ${}_4\phi(t)$ e de uma operação de dilatação efetuada sobre a mesma função escala, respectivamente. Na mesma figura, ilustram-se as mesmas transformações para a ondeleta-mãe ${}_4\psi(t)$.

3.2.5 Algoritmo Piramidal

A transformada em ondeletas discreta é efetivamente calculada por um algoritmo chamado piramidal, devido essencialmente a Mallat [52]. O algoritmo usa as equações (3.21) e (3.28), decompondo uma função $f(t) \in \mathcal{L}^2(\mathbb{R})$, na hipótese de que $f(t)$ possui nível de resolução k , de modo que $f(t) \rightarrow f_k(t)$, onde $f_k(t) \in V_k$, para $k \in \mathbb{Z}$. Desta forma, é possível decompor $f_k(t)$ como

$$f_k(t) = f_{k-1}(t) + g_{k-1}(t) , \quad (3.32)$$

onde $f_{k-1}(t) \in V_{k-1}$ e $g_{k-1}(t) \in W_{k-1}$.

Supõe-se agora que p seja o número máximo de decomposições realizadas em $f_k(t)$; obtém-se assim, sucessivamente

$$\begin{aligned} f_k(t) &= f_{k-2}(t) + g_{k-2}(t) + g_{k-1}(t) , \\ f_k(t) &= f_{k-3}(t) + g_{k-3}(t) + g_{k-2}(t) + g_{k-1}(t) , \\ &\vdots \\ f_k(t) &= f_{k-p}(t) + g_{k-p}(t) + \dots + g_{k-3}(t) + g_{k-2}(t) + g_{k-1}(t) . \end{aligned}$$

As funções que expressam $f_k(t)$ podem ser representadas como superposições de funções dilatadas e transladadas discretamente, geradas a partir da função escala $\phi(t)$ e da ondeleta-mãe $\psi(t)$. Tem-se, então

$$f_k(t) = \sum_{m \in \mathbb{Z}} a_{k-p,m} \phi_{k-p,m}(t) + \sum_{i=1}^p \sum_{m \in \mathbb{Z}} d_{k-i,m} \psi_{k-i,m}(t) . \quad (3.33)$$

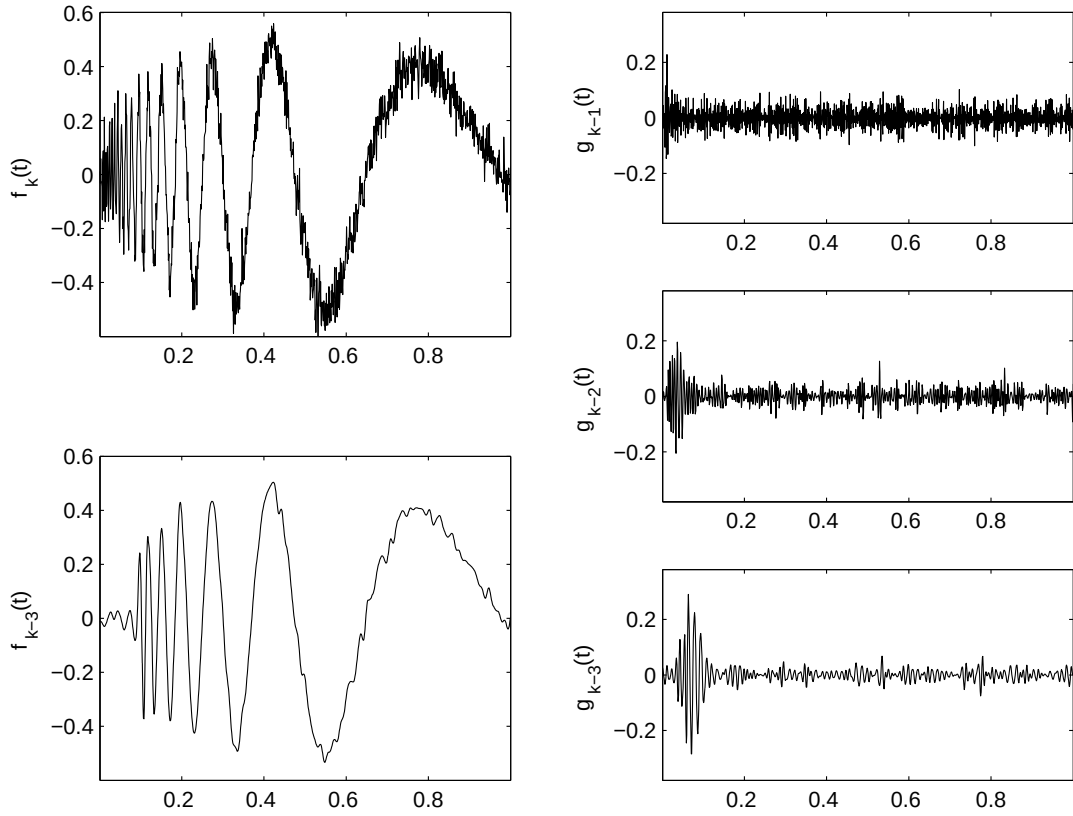


FIGURA 3.4: Decomposição da função $f_k(t)$ na aproximação $f_{k-3}(t)$ e nos detalhes $g_{k-1}(t)$, $g_{k-2}(t)$ e $g_{k-3}(t)$.

A figura 3.4 mostra a aproximação $f_{k-3}(t)$ da série temporal $f_k(t)$ composta pela função $f(t) = \sqrt{t(1-t)} \sin[2, 1\pi/(t+0,05)]$, para $0 \leq t \leq 1$, adicionada de ruído gaussiano *i.i.d.* $N(0; 2, 5 \times 10^{-3})$, em função dos coeficientes de aproximação $a_{k-3,m}$ na base ${}_4\phi_{k-3,m}(t)$ e os detalhes obtidos nesta decomposição, em função dos coeficientes de ondeleta $d_{k-i,m}$ na base ${}_4\psi_{k-i,m}(t)$, para $i = 1, 2, 3$, usando como ondeleta-mãe a função *daublet* db4. A função $f(t)$ é conhecida na literatura como Doppler.

Finalmente, obtém-se os coeficientes de aproximação a_{jm} para reconstruir a função original, fazendo

$$a_{jm} = \langle f_{j-1}(t), \phi_{jm}(t) \rangle + \langle g_{j-1}(t), \phi_{jm}(t) \rangle ,$$

onde $a_{jm} = \langle f_{j+1}(t), \phi_{jm}(t) \rangle = \langle f_j(t), \phi_{jm}(t) \rangle + \langle g_j(t), \phi_{jm}(t) \rangle = \langle f_j(t), \phi_{jm}(t) \rangle$. Esta expressão pode ser reescrita como

$$a_{jm} = \sum_{k \in \mathbb{Z}} a_{j-1,k} \langle \phi_{j-1,k}(t), \phi_{jm}(t) \rangle + \sum_{k \in \mathbb{Z}} d_{j-1,k} \langle \psi_{j-1,k}(t), \phi_{jm}(t) \rangle .$$

Fazendo novamente a transformação $2^{j-1}t - k \rightarrow t'$ e usando as equações (3.17) e (3.18), obtém-se

$$a_{jm} = \sum_{k \in \mathbb{Z}} \ell_{m-2k} a_{j-1,k} + \sum_{k \in \mathbb{Z}} h_{m-2k} d_{j-1,k} . \quad (3.34)$$

3.3 Filtragem de Ruído

O método de redução de ruído no domínio da transformada em ondeletas discreta é normalmente conhecido como método *de-noising*. Este método consiste basicamente em calcular a transformada em ondeletas discreta de um sinal determinístico contaminado por ruído, efetuar uma operação de ceifamento sobre os coeficientes de ondeleta determinados a partir de uma análise de multirresolução e, em seguida, reconstruir o sinal com os coeficientes de ondeleta aprovados pela análise de ceifamento, obtendo-se assim o sinal descontaminado de ruído.

Considere o modelo

$$w_i = f(t_i) + \epsilon_i , \quad \text{para } i = 1, 2, 3, \dots, n , \quad (3.35)$$

onde $\epsilon_i \sim i.i.d. N(0, \sigma_\epsilon^2)$. Supõe-se que $f(t_i)$ pertença a determinada classe de funções, satisfazendo certas propriedades de regularidade.

O procedimento de redução de ruído, de acordo com [53], consiste nos três estágios seguintes: a) calcular a transformada em ondeletas discreta de $w_1, \dots, w_n$, obtendo os coeficientes de ondeleta $\{d_{jk}\}$, que são contaminados por ruído; b) aplicar limiares (*thresholds*) para anular aqueles coeficientes que estão abaixo de certo valor, obtendo nesse estágio os coeficientes $\{d'_{jk}\}$ desprovidos de ruído; e c) calcular a transformada em ondeletas inversa com os coeficientes do estágio anterior para obter a função estimada $f(t_i)$. Na figura 3.5, é apresentado um esquema de decomposição, redução de ruído e reconstrução da função $f_j(t)$, usando a análise de multirresolução.

O esquema utilizado para a seleção dos coeficientes é o chamado *soft threshold*, definido como

$$\delta^s(x) = \begin{cases} 0, & \text{se } |x| \leq \lambda \\ \text{sign}(x)(|x| - \lambda), & \text{se } |x| > \lambda \end{cases} . \quad (3.36)$$

Esta abordagem baseia-se na suposição de que alguns dos coeficientes de ondeleta do sinal original sejam aproximadamente iguais a zero e que, portanto, no sinal contaminado por ruído puro aditivo, tais coeficientes correspondam a ruído puro.¹³

¹³Outro esquema que pode ser utilizado é o chamado *hard threshold* definido como: $\delta^h(x) = 0$, se $|x| \leq \lambda$ e $\delta^h(x) = x$, se $|x| > \lambda$.

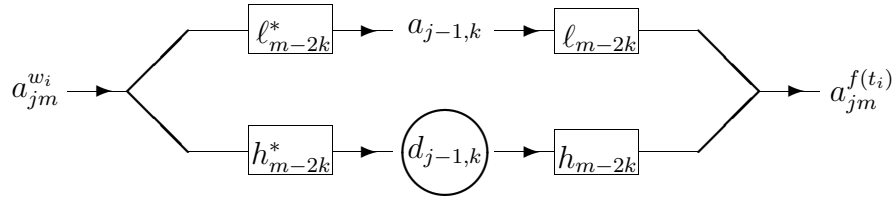


FIGURA 3.5: Esquema de decomposição, redução de ruído e reconstrução da função $f_j(t)$. A operação de ceifamento é indicada por $\bigcirc$. No diagrama, mostra-se somente uma decomposição, podendo ser necessária sucessivas decomposições em aplicações práticas.

No caso de um sinal unidimensional formado por n amostras, testa-se então, para cada coeficiente de ondeleta d_{jk} , a hipótese de que tal coeficiente seja menor que um limiar λ . Os coeficientes de aproximação da última decomposição são sempre incluídos na reconstrução do sinal.

Donoho e Johnstone [54], propõem a utilização de um limiar λ aplicável a todos os níveis j . Este parâmetro é definido como

$$\lambda = \sigma \sqrt{2 \log_e n}, \quad (3.37)$$

que não depende da escala e é chamado universal. Para o nível de ruído σ , pode-se usar o estimador definido por

$$\hat{\sigma} = \frac{\text{median}(|d_{J_0 1}|, |d_{J_0 2}|, \dots, |d_{J_0 m}|)}{0,6745}, \quad (3.38)$$

onde J_0 representa o nível associado à escala de maior resolução, m é o tamanho da série composta pelos coeficientes de ondeleta $\{d_{J_0 k}\}_{k=1}^m$ e o fator $0,6745$ no denominador reescala o numerador para que $\hat{\sigma}$ também seja um estimador adequado para o desvio padrão de um ruído gaussiano [49].

Lembra-se que o método de redução de ruído baseado na transformada em ondeletas não exige estudos prévios quanto à estacionariedade da série a ser filtrada. Os resultados obtidos na aplicação deste método em séries temporais contaminadas com ruído são mostrados posteriormente no capítulo 5. Inspeção visual ou um algoritmo que determine o nível de ruído para uma comparação entre as séries contaminada e filtrada é aconselhável após o uso desta técnica.

Capítulo 4

Redes Neurais Artificiais

4.1 Introdução

Na busca pela representação do conhecimento, inspirado na organização física do cérebro, com suas potencialidades e limitações, situa-se a escola conexionista. O conexionismo denota uma forma particular de processamento da informação a partir de modelos computacionais baseados na estrutura e funcionamento do cérebro humano, caracterizados pela reunião de uma quantidade de células de processamento interligadas por conexões. Estes modelos apresentam “inteligência”, sendo possível aprender, errar e aprender com o erro. Com base neste tipo de estrutura, denominam-se adequadamente estes modelos por redes neurais artificiais, em alusão à tentativa de imitação dos conjuntos formados pelas células neurais, suas características básicas e organização. Todavia, o conhecimento do funcionamento do cérebro, bem como da interconectividade entre os neurônios, ainda é bastante incompleto. Portanto, tais modelos computacionais nada mais são que imitações “grosseiras” e parciais das redes de neurônios do cérebro.

O neurônio, que é a unidade fundamental do cérebro, isoladamente pode ser considerado análogo a uma unidade de processamento; recebe estímulos (pulsos elétricos) provenientes de várias outras células neurais por meio de muitas entradas, e transmite estímulos produzidos nele através de uma única saída, como ilustra a figura 4.1. Os constituintes que formam o neurônio são: a) dendritos, que são um conjunto de terminais de entrada com a função de receber os estímulos; b) corpo celular, que é responsável pela coleta e combinação dos estímulos fornecidos pelos dendritos; e c) axônio, que é um terminal de saída responsável por transmitir estímulos gerados no neurônio para outras células. Este último pode ser considerado como um dispositivo de disparo que produz um pulso elétrico toda vez que o somatório dos sinais dentro do corpo celular atingir um certo limiar crítico [55]. Geralmente, o axônio divide-se na sua extremidade terminal em muitos ramos e assim ele pode passar a sua

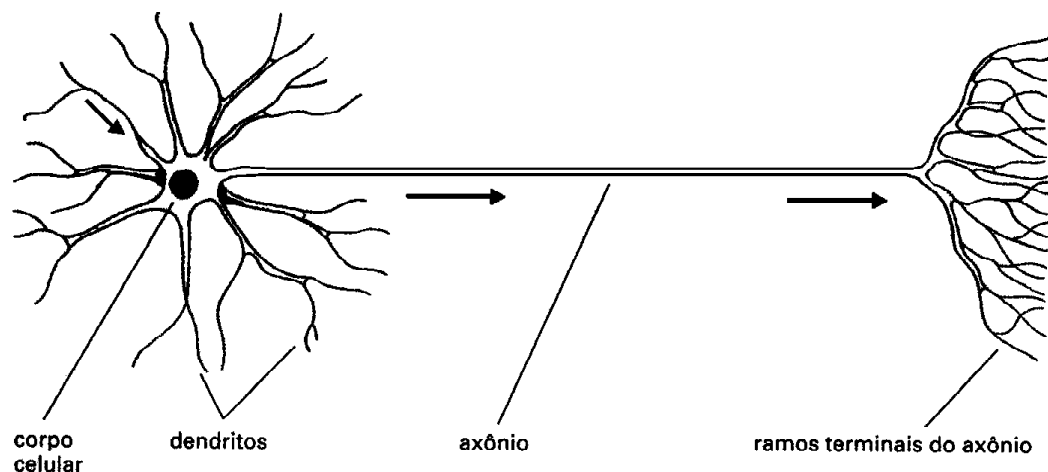


FIGURA 4.1: Esquema dos constituintes do neurônio. As flechas indicam a direção em que os sinais são internamente transmitidos na célula neural.

mensagem simultaneamente a outras células neurais.

O axônio termina em um tipo de contato chamado sinapse, que conecta-o com o dendrito de outro neurônio. Esta conexão não é uma ligação direta; trata-se de uma junção química temporária. A sinapse libera moléculas sinalizadoras, denominadas neurotransmissores, em função do pulso elétrico disparado pelo axônio, sendo possível a necessidade de vários outros disparos antes que a sinapse libere estas substâncias que, por sua vez, rapidamente difundem-se pela fenda sináptica.¹⁴ Os neurotransmissores provocam estímulos elétricos nas células pós-sinápticas por meio de dendritos quimicamente ativos (neuroreceptores), que são conduzidos ao próximo corpo celular, e assim sucessivamente. Para maiores detalhes, ver em [56].

O primeiro modelo de redes neurais artificiais foi apresentado no trabalho pioneiro de McCulloch e Pitts [57]. Com um número suficiente de neurônios e com conexões sinápticas ajustadas apropriadamente, estes autores mostraram que uma rede assim constituída realizaria, em princípio, o cálculo de qualquer função computável. Este era um resultado muito significativo e com ele é geralmente aceito o surgimento da pesquisa em inteligência artificial [58]. O próximo desenvolvimento significativo das redes neurais vem com Hebb [59], no qual foi apresentada pela primeira vez uma formulação explícita de uma regra de aprendizagem fisiológica

¹⁴Quando o axônio de um neurônio conecta-se a uma outra célula para enviar sua mensagem, ele fica em proximidade com esta célula. Entretanto, ele deixa um espaço que é chamado de fenda sináptica.

para a modificação sináptica. O processo de aprendizado ocorre quando acontecem repetidas e efetivas modificações nas sinapses que interconectam os neurônios, em função da maior ou menor liberação de neurotransmissores. Na medida que novos eventos são apresentados, determinadas ligações entre os neurônios são reforçadas, enquanto que outras enfraquecidas. A interconectividade do cérebro é continuamente modificada conforme o organismo vai aprendendo tarefas funcionais diferentes e que agrupamentos neurais são criados por tais modificações. Este ajuste nas ligações entre os neurônios durante o processo de aprendizado é uma das mais importantes características das redes neurais (artificiais).

Um outro passo importante foi dado por Rosenblatt [60], em seu trabalho sobre o *perceptron*. Esta arquitetura foi proposta como uma nova abordagem para o problema na classificação de padrões linearmente separáveis.¹⁵ Basicamente, a arquitetura do *perceptron* consiste de um único neurônio com função de ativação *threshold* e uma camada de pesos ajustáveis nas conexões sinápticas, limitando-se apenas na classificação de padrões com apenas duas classes (hipótese). Rosenblatt provou que se os padrões utilizados para treinar o *perceptron* são retirados de duas classes linearmente separáveis, então o algoritmo converge e posiciona a superfície de decisão na forma de um hiperplano entre as duas classes. Suponha então que as variáveis de entrada tenham sua origem em duas classes linearmente separáveis $\mathcal{C}_1$ e $\mathcal{C}_2$. Sendo $\mathcal{T}_1 = x_1(1), x_1(2), \dots, x_1(m)$ e $\mathcal{T}_2 = x_2(1), x_2(2), \dots, x_2(n)$ subconjuntos de vetores de treinamento pertencentes às classes $\mathcal{C}_1$ e $\mathcal{C}_2$, respectivamente. A união de $\mathcal{T}_1$ e $\mathcal{T}_2$ forma o conjunto de vetores de treinamento completo $\mathcal{T}$. O processo de treinamento consiste no ajuste do vetor de pesos nas conexões sinápticas, dados os $m + n$ vetores do conjunto $\mathcal{T}$ para treinar o classificador, de tal forma que os vetores pertencentes às classes diferentes estejam linearmente separados.

Durante o período do *perceptron* nos anos 60, parecia que as redes neurais artificiais poderiam realizar qualquer tarefa, utilizando a aprendizagem proposta por Rosenblatt. Mas, então, Minsky e Papert [61], demonstraram a existência de limites fundamentais devidos à padrões não linearmente separáveis, gerando desinteresse pela pesquisa nesta área nos anos seguintes. Também a publicação de diversos trabalhos que descreviam modelos de previsão pouco confiáveis tirou quase toda credibilidade das redes neurais artificiais.

A partir dos anos 80, com o avanço da tecnologia provendo suporte computacional e o fracasso do simbolismo¹⁶ na solução de determinados problemas, as redes neurais artificiais passaram a atrair substancial atenção novamente. Nesta mesma

¹⁵Isto é, padrões que se encontram em lados opostos de um hiperplano.

¹⁶A escola simbolista tenta simular o comportamento inteligente desconsiderando a estrutura e os mecanismos responsáveis por tal, ao contrário da escola conexionista.

década, Rumelhart *et al.* [62], desenvolveram o eficiente algoritmo *error backpropagation* a partir do aperfeiçoamento da idéia do *perceptron*. Este algoritmo é baseado na regra de aprendizagem por correção de erro e foi utilizado com sucesso em arquiteturas de rede denominadas *perceptrons* de múltiplas camadas (de pesos sinápticos). Tipicamente, a rede consiste de um conjunto de unidades sensoriais que constituem a entrada da rede, um ou mais conjuntos ocultos de neurônios artificiais e mais um conjunto de neurônios artificiais na saída da rede. O sinal de entrada propaga-se para frente através da rede, camada por camada.

As redes neurais artificiais certamente trilharam por um longo caminho desde os dias de McColluch e Pitts. De fato, elas estabeleceram-se como um tema interdisciplinar com raízes profundas em computação, física, matemática, neurociências e psicologia. Atualmente, as redes neurais artificiais têm sido amplamente exploradas na modelagem e previsão de séries temporais não-lineares; particularmente em séries caóticas estas ainda são úteis, porém com certas restrições para previsão.

Para fixar o vocabulário utilizado até aqui e na seqüência do capítulo, pretende-se, a partir da comparação com o neurônio biológico, identificar os constituintes da célula artificial. Os dendritos são representados por unidades sensoriais de entrada, cujas ligações com o corpo celular artificial são realizadas por meio de elementos denominados pesos, simulando as sinapses. Os estímulos são processados por uma função soma (corpo celular artificial), e o limiar crítico por uma função de ativação (ou função de transferência), que então transmite o pulso por unidades de saída, que representam o axônio.

O neurônio artificial ilustrado na figura 4.2 inclui também um viés aplicado externamente, representado por b_k . O viés tem o efeito de aumentar ou diminuir a entrada líquida na função de ativação, dependendo se ele é positivo ou negativo, respectivamente. Em termos matemáticos, descreve-se o k -ésimo neurônio pelas seguintes equações

$$u_k = \sum_{i=1}^n w_{ki} x_i , \quad (4.1)$$

$$y_k = \varphi(u_k + b_k) , \quad (4.2)$$

onde $x_1, x_2, \dots, x_n$ são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{kn}$ são os pesos sinápticos; u_k é a saída da função soma, representada por Σ_k na figura, devido aos sinais de entrada ponderados pelos pesos sinápticos do neurônio k ; b_k é um viés; φ é a função de ativação; e y_k é o sinal de saída do neurônio artificial. O uso do viés b_k tem o efeito de aplicar uma transformação à saída u_k da função soma, como mostrado por

$$v_k = u_k + b_k . \quad (4.3)$$

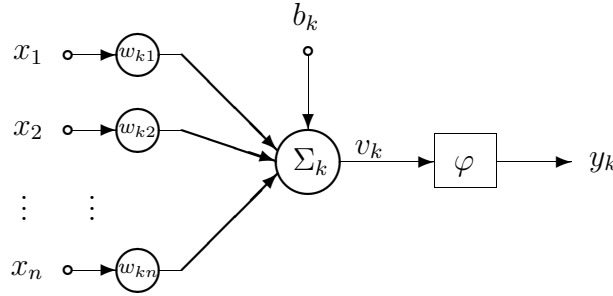


FIGURA 4.2: Esquema dos constituintes do neurônio artificial.

Equivalentemente, reescrevem-se as equações (4.1) e (4.2) respectivamente como

$$v_k = \sum_{i=0}^n w_{ki} x_i, \quad (4.4)$$

$$y_k = \varphi(v_k), \quad (4.5)$$

ou seja, na equação (4.4), adiciona-se uma nova conexão sináptica. A sua entrada é definida como $x_0 = 1$ e o seu peso é $w_{k0} = b_k$.

Redes neurais artificiais com apenas uma camada de pesos sinápticos, como o modelo *perceptron*, são baseadas em uma combinação linear de variáveis de entrada, que são transformadas por uma função de ativação, resultando em redes muito limitadas em relação ao tipo de funções que conseguem representar. Além disso, os neurônios internos são de grande importância na rede, pois provou-se que sem estes torna-se impossível a resolução de problemas não linearmente separáveis [63]. Para permitir um mapeamento mais genérico, consideram-se sucessivas transformações na rede formada por várias camadas de pesos ajustáveis, utilizando um algoritmo de aprendizado. Segundo Bishop [64], duas camadas de pesos sinápticos (um conjunto oculto de neurônios artificiais) são capazes de aproximar qualquer mapeamento dado por uma função contínua. Para a utilização do algoritmo de aprendizado *error backpropagation*, a única restrição imposta é que a rede seja alimentada de modo progressivo, ou *feedforward*, para que não contenha passos cíclicos. Isto assegura que a saída da rede possa ser determinada como uma função explícita dos parâmetros de entrada e dos pesos sinápticos. Tais redes são teoricamente mais fáceis de serem analisadas do que topologias mais genéricas, e podem frequentemente ser implementadas de uma maneira mais eficiente em uma simulação computacional.

Um exemplo de rede multicamada *feedforward* é mostrado na figura 4.3. O diagrama representa uma rede de duas camadas de pesos sinápticos, com n unidades

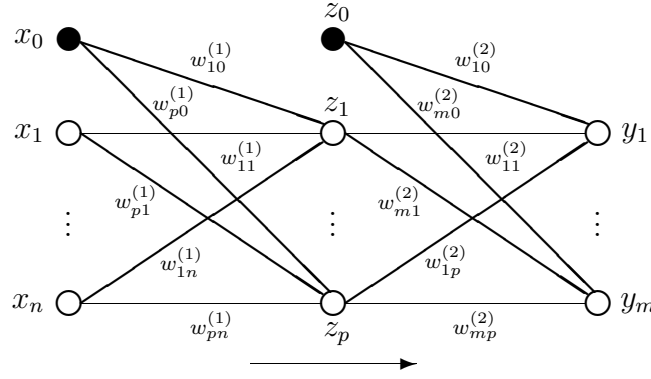


FIGURA 4.3: Exemplo de rede multicamada *feedforward*, onde $\{x_i\}_{i=1}^n$ representam as variáveis de entrada, $\{z_j\}_{j=1}^p$ as unidades ocultas, $\{y_k\}_{k=1}^m$ as unidades de saída, $w_{ji}^{(1)}$ e $w_{kj}^{(2)}$ os pesos sinápticos e $x_0 = z_0 = 1$. Os pesos $w_{10}^{(1)}, \dots, w_{p0}^{(1)}$ e $w_{10}^{(2)}, \dots, w_{m0}^{(2)}$ são vieses. A flecha indica o sentido de propagação do sinal.

de entrada x , p unidades de neurônios artificiais z formando um conjunto oculto, e m unidades de neurônios artificiais y formando um conjunto de saída. Os pesos sinápticos w representam as conexões entre as unidades de processamento.

Os parâmetros n , p e m estão vinculados de acordo com o teorema de Kolmogorov. Este teorema permite demonstrar que qualquer função contínua

$$f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m ,$$

definida no hipercubo unitário n -dimensional e com valores em um espaço de dimensão m , pode ser representada por uma rede neural artificial com n unidades de entrada, $p = 2n + 1$ unidades de neurônios artificiais no conjunto oculto, e m unidades de neurônios artificiais no conjunto de saída. Apesar disso, na prática deve-se favorecer o uso do menor número possível de neurônios, pois isso restringe o fenômeno de *overfitting*.¹⁷

Descreve-se as funções analíticas correspondentes à rede ilustrada na figura 4.3 fazendo os passos a seguir. A saída da j -ésima unidade oculta, que é obtida formando-se uma combinação linear ponderada dos valores de entrada e adicionada

¹⁷Fenômeno em que a rede tem um bom desempenho no conjunto de treinamento, mas que decai substancialmente na validação, isto é, sobre dados que não foram apresentados a ela. Isso significa que a rede “memorizou” o treinamento e que não generaliza bem. O conjunto de treinamento consiste de uma coleção de vetores de entrada, e respectivas saídas, que é utilizada para o aprendizado da rede. O conjunto de validação, que é um conjunto desconhecido do ambiente, serve para avaliar o desempenho da rede.

de um viés, é escrita como

$$v_j = \sum_{i=0}^n w_{ji}^{(1)} x_i , \quad (4.6)$$

onde a ativação da unidade oculta j é então obtida pela transformação da soma linear na equação (4.6), utilizando a função de ativação φ_1 , por meio de

$$z_j = \varphi_1(v_j) . \quad (4.7)$$

As saídas da rede são obtidas transformando as ativações das unidades ocultas através de uma segunda camada de elementos de processamento. Assim, para cada saída k , constrói-se uma combinação linear com as saídas das unidades intermediárias na forma

$$v_k = \sum_{j=0}^p w_{kj}^{(2)} z_j , \quad (4.8)$$

onde a ativação da k -ésima unidade de saída é então obtida transformando esta combinação linear, utilizando uma segunda função de ativação, resultando em

$$y_k = \varphi_2(v_k) . \quad (4.9)$$

A função de ativação φ_2 não precisa necessariamente ser igual à função de ativação φ_1 utilizada anteriormente.

Combinando as quatro equações que descrevem o processo, obtém-se uma expressão explícita para a k -ésima saída em função somente das duas camadas de pesos sinápticos e das entradas, reescrevendo y_k na forma

$$y_k = \varphi_2 \left[\sum_{j=0}^p w_{kj}^{(2)} \varphi_1 \left(\sum_{i=0}^n w_{ji}^{(1)} x_i \right) \right] . \quad (4.10)$$

As funções de ativação φ_1 e φ_2 determinam, respectivamente, o grau de importância das somas ponderadas v_j e v_k , gerando excitações ou inibições nos neurônios artificiais associados às ativações. De acordo com Haykin [58], a ativação $\varphi(v)$ define a saída de um neurônio em termos do campo local induzido v . Existem vários tipos de funções de ativação, dentre elas destacam-se as funções degrau de Heaviside ou *threshold*, logística sigmóide e tangente hiperbólica, definidas respectivamente como

$$\varphi(v) = \begin{cases} 1, & \text{para } v \geq 0 \\ 0, & \text{para } v < 0 \end{cases} , \quad (4.11)$$

$$\varphi(v) = \frac{1}{1 + e^{-v}} , \quad (4.12)$$

$$\varphi(v) = \tanh(v) = \frac{e^v - e^{-v}}{e^v + e^{-v}}. \quad (4.13)$$

Observa-se que a função de ativação tangente hiperbólica é, na verdade, uma transformação linear da função logística sigmóide.¹⁸ Dessa forma, as duas funções são equivalentes, porém geram pesos e vieses diferentes.

Um *perceptron* de múltiplas camadas treinado com algoritmo *error backpropagation*, em geral, pode aprender mais rápido (em termos do número de iterações de treinamento necessárias) quando a função de ativação incorporada no modelo do neurônio da rede for anti-simétrica do que quando for não-simétrica, ou seja, aconselham-se funções da ativação que satisfazem à condição

$$\varphi(-v) = -\varphi(v).$$

A explicação para esta observação é que para minimizar o tempo de aprendizagem deve-se evitar o uso de entradas com médias diferentes de zero [55, 58, 64]. Considerando o vetor de entrada aplicado a um neurônio pertencente ao primeiro conjunto oculto de um *perceptron* de múltiplas camadas, é fácil remover a média de cada elemento antes de aplicá-lo na rede. Mas, para os sinais aplicados aos neurônios dos conjuntos ocultos restantes e do conjunto de saída, utiliza-se uma função de ativação anti-simétrica tal como a função tangente hiperbólica. Com esta escolha, permite-se que a saída de cada neurônio assuma valores tanto positivos como negativos no intervalo $[-1, 1]$, possibilitando uma provável média nula. Se a função de ativação for não-simétrica, como no caso da função logística sigmóide, a saída de cada neurônio está restrita ao intervalo $[0, 1]$ e, assim, introduz uma fonte de viés sistemática para neurônios localizados além do primeiro conjunto oculto da rede. Em LeCun *et al.* [65], apresentam-se evidências empíricas nesse contexto.

4.2 Método *Error Backpropagation*

A propriedade que é de importância primordial para uma rede neural artificial é a sua habilidade de aprender na tentativa de melhorar seu desempenho. Isto ocorre de acordo com alguma medida preestabelecida. Uma rede neural artificial aprende por meio de um processo iterativo de ajustes aplicados aos seus pesos sinápticos e vieses. Idealmente, a rede torna-se mais eficiente após cada iteração do processo de aprendizagem.

¹⁸Mais especificamente, a ativação $\tilde{\varphi}(\tilde{v})$ efetuada pela função tangente hiperbólica é equivalente à ativação $\varphi(v)$ efetuada pela função logística sigmóide se aplicadas as transformações lineares $\tilde{v} = v/2$ e $\tilde{\varphi} = 2\varphi - 1$.

Mais precisamente, a aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural artificial são adaptados por meio de um processo de estimulação do ambiente no qual a rede está inserida [58]. O tipo de aprendizagem é determinado pela maneira que a modificação dos parâmetros ocorre. Um conjunto preestabelecido de regras bem definidas para a solução de um problema de aprendizagem é denominado algoritmo de aprendizagem.

Normalmente, considera-se um professor, sendo então denominada aprendizagem supervisionada. Neste caso, o professor tem um conhecimento prévio sobre o ambiente, com este conhecimento sendo representado por um conjunto de exemplos de entrada-saída. Entretanto, o ambiente é desconhecido pela rede. Supõe-se então que o professor e a rede sejam expostos a um vetor de treinamento retirado do ambiente. Em virtude do seu conhecimento prévio, o professor é capaz de fornecer à rede neural uma resposta desejada para este vetor de treinamento. Esta resposta representa a ação ótima a ser realizada pela rede. Os parâmetros da rede são ajustados sob a influência combinada do vetor de treinamento e do sinal de erro, que é definido como a diferença entre a resposta desejada e a resposta da rede. Este ajuste é realizado iterativamente com o objetivo de fazer a rede neural ter o conhecimento do ambiente disponível ao professor. Desta forma, transfere-se o conhecimento do professor para a rede por meio do treinamento, da forma mais completa possível. Quando esta condição é alcançada, dispensa-se o professor e permite-se então à rede lidar com o ambiente inteiramente por si mesma.

A forma de aprendizagem supervisionada que foi descrita é a aprendizagem por correção de erro. Para ilustrar esta aprendizagem, considera-se o caso simples de um neurônio k , alimentado por um vetor de sinal $z(n)$, produzido por um ou mais conjuntos de neurônios ocultos, que são, por sua vez, acionados por sinais de entrada $x(n)$ aplicados na rede. O argumento n representa o n -ésimo padrão de entrada da rede. O sinal de saída do neurônio k é representado por $y_k(n)$. Este sinal de saída é comparado com uma resposta desejada, ou alvo (*target*), representada por $t_k(n)$. Conseqüentemente, é produzido um sinal de erro, representado por $\epsilon_k(n)$. Por definição, tem-se que

$$\epsilon_k(n) = t_k(n) - y_k(n) . \quad (4.14)$$

O sinal de erro $\epsilon_k(n)$ aciona um mecanismo de controle, cujo propósito é aplicar uma seqüência de ajustes corretivos aos pesos sinápticos do neurônio k . Os ajustes corretivos são projetados para aproximar passo a passo o sinal de saída $y_k(n)$ da resposta desejada $t_k(n)$. Este objetivo é alcançado minimizando-se uma função de

custo $E(n)$, definida em termos do sinal de erro $\epsilon_k(n)$ como

$$E(n) = \left[\frac{\epsilon_k(n)}{\sqrt{2}} \right]^2. \quad (4.15)$$

Com isso, a função de custo $E(n)$ representa o valor “instantâneo da energia” do erro.

Na verdade, $E(n)$ pode ser expressa como uma função diferenciável da variável de saída desde que a função de ativação seja diferenciável. Os ajustes passo a passo dos pesos sinápticos do neurônio k continuam até o sistema atingir um estado estável, ou seja, até os pesos sinápticos estarem essencialmente estabilizados. A função de custo pode então ser definida como uma medida de desempenho para o sistema. Também é possível pensar em termos da soma de erros quadráticos sobre toda a amostra de treinamento, definida como função dos parâmetros livres do sistema. Esta função é escrita por

$$E_{tot} = \sum_{n=1}^N E(n), \quad (4.16)$$

onde N é o número de padrões. Do ponto de vista computacional esta medida de desempenho torna a aplicação do método mais rápida.¹⁹ Para que o sistema melhore seu desempenho e, portanto, aprenda com o professor, a operação da rede deve ocorrer sucessivamente em direção a um ponto mínimo de E_{tot} ; tal mínimo pode ser local ou global. Resolve-se este problema com a aplicação inicial de várias configurações de camadas de pesos sinápticos diferentes na rede, comparando-se logo após a eficiência de cada configuração. Evidentemente o mínimo global nunca será atingido na prática e a rede sempre utiliza algum ótimo local.

Basicamente, o método de aprendizagem por correção de erro consiste de dois passos através da rede: um passo para frente, a propagação, e um passo para trás, a retropropagação. No passo para frente, os N padrões são aplicados como entradas e seus efeitos propagam-se através da rede até o sistema obter a saída. Durante o passo de propagação, os pesos sinápticos da rede são todos fixos. No passo para trás, por outro lado, os pesos sinápticos são todos ajustados de acordo com uma regra de correção de erro da função E_{tot} . Este sinal de erro é propagado para trás através da rede, contra a direção de propagação usual, vindo daí o nome *error backpropagation*.

Apresenta-se agora uma breve descrição do algoritmo. O campo local induzido no neurônio k , é dado por

$$v_k(n) = \sum_{j=0}^p w_{kj} z_j(n),$$

¹⁹No caso particular em que a rede será utilizada no trabalho, isto é, na modelagem de séries temporais determinísticas não-lineares.

onde p é o número total de entradas (excluindo o viés) aplicadas no neurônio k e n indica que o sinal $z_j(n)$ é proveniente da aplicação do n -ésimo padrão de treinamento na rede. O peso sináptico w_{k0} corresponde ao viés, e conseqüentemente, $z_0(n) = 1$. Desta forma, o sinal funcional $y_k(n)$ que aparece na saída do neurônio k é definido por

$$y_k(n) = \varphi[v_k(n)] .$$

O algoritmo *error backpropagation* tem por intenção aplicar uma correção Δw_{kj} ao peso sináptico w_{kj} , de modo que

$$\tilde{w}_{kj} = w_{kj} + \Delta w_{kj} , \quad (4.17)$$

onde a correção do peso sináptico é proporcional ao gradiente $\partial E_{tot}/\partial w_{kj}$, ou seja, $\Delta w_{kj} \propto \sum_{n=1}^N \partial E(n)/\partial w_{kj}$. De acordo com a regra da cadeia, define-se $\partial E(n)/\partial w_{kj}$ como

$$\frac{\partial E(n)}{\partial w_{kj}} = \frac{\partial E(n)}{\partial \epsilon_k(n)} \frac{\partial \epsilon_k(n)}{\partial y_k(n)} \frac{\partial y_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial w_{kj}} . \quad (4.18)$$

Da definição de $E(n)$, tem-se que

$$\frac{\partial E(n)}{\partial \epsilon_k(n)} = \epsilon_k(n) ,$$

da definição de $\epsilon_k(n)$ obtém-se

$$\frac{\partial \epsilon_k(n)}{\partial y_k(n)} = -1 ,$$

e da definição de $y_k(n)$ verifica-se

$$\frac{\partial y_k(n)}{\partial v_k(n)} = \frac{\partial \varphi[v_k(n)]}{\partial v_k(n)} = \varphi'[v_k(n)] ,$$

onde o uso do apóstrofo significa a diferenciação em relação ao argumento. E finalmente, diferenciar $v_k(n)$ em relação a w_{kj} produz

$$\frac{\partial v_k(n)}{\partial w_{kj}} = \sum_{j'=0}^p \left(\frac{\partial w_{kj'}}{\partial w_{kj}} \right) z_{j'}(n) = \sum_{j'=0}^p \delta_{j'j} z_{j'}(n) = z_j(n) .$$

onde $\delta_{j'j}$ é a função delta de Kronecker. Lembrando que a ativação da unidade oculta j independe do peso w_{kj} definido na segunda camada de pesos sinápticos.

Usando as equações obtidas, tem-se

$$\frac{\partial E(n)}{\partial w_{kj}} = -\epsilon_k(n) \varphi'[v_k(n)] z_j(n) , \quad (4.19)$$

e a derivada da função E_{tot} pode ser obtida repetindo-se este processo para cada padrão do conjunto de treinamento. E então, na soma em todos os padrões, obtém-se

$$\frac{\partial E_{tot}}{\partial w_{kj}} = - \sum_{n=1}^N \epsilon_k(n) \varphi'[v_k(n)] z_j(n) . \quad (4.20)$$

A equação (4.19) pode ser reescrita como

$$\frac{\partial E(n)}{\partial w_{kj}} = -\Delta_k(n) z_j(n) , \quad (4.21)$$

onde o gradiente local $\Delta_k(n)$ é definido por

$$\Delta_k(n) = -\frac{\partial E(n)}{\partial v_k(n)} = \epsilon_k(n) \varphi'[v_k(n)] , \quad (4.22)$$

para a unidade de saída k . O gradiente local aponta para as modificações locais necessárias nos pesos sinápticos. Para as unidades ocultas, utilizando-se da regra da cadeia novamente, obtém-se

$$\Delta_j(n) = -\frac{\partial E(n)}{\partial v_j(n)} = - \sum_{k=1}^m \frac{\partial E(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial v_j(n)} . \quad (4.23)$$

Para o caso particular em que a rede é constituída por duas camadas de pesos sinápticos, escreve-se finalmente

$$\Delta_j(n) = \sum_{k=1}^m \left\{ \Delta_k(n) \sum_{j'=0}^p w_{kj'} \varphi'[v_{j'}(n)] \delta_{j'j} \right\} = \varphi'[v_j(n)] \sum_{k=1}^m w_{kj} \Delta_k(n) , \quad (4.24)$$

para

$$\frac{\partial E(n)}{\partial w_{ji}} = -\Delta_j(n) \sum_{i'=0}^n \left(\frac{\partial w_{ji'}}{\partial w_{ji}} \right) x_{i'}(n) = -\Delta_j(n) x_i(n) , \quad (4.25)$$

onde os pesos w_{ji} são definidos na primeira camada de pesos sinápticos e $x_i(n)$ são as entradas do n -ésimo padrão de treinamento aplicado na rede. No ajuste dos pesos relacionados ao neurônio k , definidos neste caso em particular, na segunda camada de pesos sinápticos, impõe-se que a correção Δw_{kj} aplicada ao peso sináptico w_{kj} é definida pela regra delta²⁰ e, assim, obtém-se

$$\Delta w_{kj} = -\eta \sum_{n=1}^N \frac{\partial E(n)}{\partial w_{kj}} = \eta \sum_{n=1}^N \Delta_k(n) z_j(n) , \quad (4.26)$$

²⁰A regra delta, ou regra de Widrow-Hoff, diz que o ajuste do peso sináptico de um neurônio é proporcional ao produto do sinal de erro pelo sinal de entrada da sinapse em questão.

e para as conexões da primeira camada de pesos sinápticos, utilizando-se da regra delta novamente, a correção será dada por

$$\Delta w_{ji} = -\eta \sum_{n=1}^N \frac{\partial E(n)}{\partial w_{ji}} = \eta \sum_{n=1}^N \Delta_j(n) x_i(n) . \quad (4.27)$$

Desta forma, o ajuste dos pesos nas conexões sinápticas somente é feito quando todo o conjunto de treinamento for apresentado à rede. O parâmetro η representa a taxa de aprendizagem do algoritmo *error backpropagation* e o sinal negativo na definição das correções Δw_{kj} e Δw_{ji} indica a descida do gradiente no espaço dos pesos, isto é, busca uma direção para a mudança de pesos que reduza valor de E_{tot} .

Este método fornece uma “aproximação” para a trajetória no espaço dos pesos. Quanto menor for o parâmetro da taxa de aprendizagem η , menor serão as variações dos pesos sinápticos da rede, de uma iteração para outra, e mais suave será a trajetória no espaço dos pesos. Esta melhoria, entretanto, é obtida à custa de uma lentidão na convergência do algoritmo. Por outro lado, se o parâmetro da taxa de aprendizagem η for muito grande, para acelerar o processo, grandes modificações nos pesos sinápticos podem tornar a convergência instável.

O cálculo do gradiente local para cada neurônio do *perceptron* de múltiplas camadas requer o conhecimento da derivada da função de ativação φ associada ao neurônio. Para a derivada existir, necessariamente a função φ deve ser contínua. Basicamente, a diferenciabilidade é a única exigência que a função de ativação deve satisfazer. Dois exemplos de função de ativação são dados pelas equações (4.12) e (4.13). Descreve-se, a seguir, a aplicação destas funções para o cálculo dos gradientes locais.

Para a função logística sigmóide, definida pela equação (4.12), diferenciando $\varphi[v_k(n)]$ em relação a $v_k(n)$ obtém-se

$$\varphi'[v_k(n)] = \frac{e^{-v_k(n)}}{[1 + e^{-v_k(n)}]^2} . \quad (4.28)$$

Sendo $y_k(n) = \varphi[v_k(n)]$, reescreve-se a derivada como

$$\varphi'[v_k(n)] = y_k(n)[1 - y_k(n)] . \quad (4.29)$$

O gradiente local para o neurônio k pertencente ao conjunto de saída é, então, definido por

$$\Delta_k(n) = y_k(n)[1 - y_k(n)][t_k(n) - y_k(n)] , \quad (4.30)$$

onde $t_k(n)$ é a resposta desejada para o n -ésimo padrão de entrada da rede. Por outro lado, para uma unidade j pertencente ao conjunto oculto, escreve-se o gradiente local

como

$$\Delta_j(n) = z_j(n)[1 - z_j(n)] \sum_{k=1}^m w_{kj} \Delta_k(n) . \quad (4.31)$$

Para a função tangente hiperbólica, definida pela equação (4.13), que pode ser obtida por transformações lineares da função logística sigmóide, a derivada em relação a $v_k(n)$ é dada por

$$\varphi'[v_k(n)] = \text{sech}^2[v_k(n)] = 1 - \tanh^2[v_k(n)] . \quad (4.32)$$

Sendo $y_k(n) = \varphi[v_k(n)]$, reescreve-se a derivada como

$$\varphi'[v_k(n)] = [1 - y_k(n)][1 + y_k(n)] . \quad (4.33)$$

O gradiente local para o neurônio k pertencente ao conjunto de saída é, então, definido por

$$\Delta_k(n) = [1 - y_k(n)][1 + y_k(n)][t_k(n) - y_k(n)] , \quad (4.34)$$

e para uma unidade oculta j , escreve-se o gradiente local como

$$\Delta_j(n) = [1 - z_j(n)][1 + z_j(n)] \sum_{k=1}^m w_{kj} \Delta_k(n) . \quad (4.35)$$

Utilizando as equações (4.30) e (4.31) para a função logística sigmóide e, as equações (4.34) e (4.35) para a função tangente hiperbólica, calculam-se facilmente os gradientes locais $\Delta_k(n)$ e $\Delta_j(n)$ utilizados nas correções dos pesos sinápticos durante o treinamento, verificando a eficiência da rede na fase de validação. A aplicação do algoritmo, apresentada no capítulo 5, foi realizada no caso particular em que o conjunto de saída é formado somente por um neurônio artificial. Para maiores detalhes sobre o método *error backpropagation*, aconselham-se [58, 62, 64].

Capítulo 5

Resultados e Conclusões

5.1 Introdução

Neste capítulo, apresentam-se os resultados obtidos na análise realizada pelos testes propostos para detecção de determinismo em séries temporais empíricas. Como observado em capítulos anteriores, para tal feito, faz-se necessário estudos preliminares na caracterização da série. Estes estudos consistem basicamente em testes para análise de estacionariedade e linearidade, bem como de um método para redução de ruído. A transformada em ondeletas, que foi utilizada para realização da filtragem, requer como condição básica que o sinal tenha uma certa regularidade [49, 53, 54]. De modo que a aplicação desta técnica será feita após verificação de caos determinístico, pois nesse contexto a diferenciabilidade da série caótica é bem representada nos coeficientes da transformada; validando-se o método mediante a análise dos resíduos obtidos no processo de filtragem. Dos testes propostos para detecção de determinismo, o baseado na abordagem métrica mostra uma grande sensibilidade à presença de ruído [4, 9, 11]. Portanto, optou-se pela não utilização da dimensão de correlação.²¹

Inicialmente, descreve-se as séries temporais empíricas utilizadas no trabalho e as condições em que estas foram registradas. Na sequência, classifica-se as séries quanto ao mecanismo gerador usando a complexidade de Lempel-Ziv e o diagrama de recorrência, seguido da caracterização das mesmas por meio dos testes propostos de estacionariedade e linearidade. Logo após, mostra-se o processo de redução de ruído por meio da análise de multirresolução, aplicando a transformada em ondeletas discreta. Finalmente, faz-se uma demonstração prática da utilização de redes neurais artificiais em séries temporais devidamente comprovadas caóticas.

²¹A subjetividade na escolha da região de escala do parâmetro ε também contribuiu para esta decisão.

5.2 Atividade Elétrica Cerebral

As séries temporais empíricas utilizadas são referentes à atividade elétrica cerebral do rato. Os registros foram realizados por eletrodos implantados na formação hipocampal do animal. Esta formação abriga um dos principais arquivos de memória e esta é a matéria prima dos sonhos. A memória é encarregada de armazenar e processar dados e fatos relativos ao que se aprende obtidos por meio de canais sensoriais. Por exemplo, sendo a visão o principal canal sensorial humano, é compreensível que a maioria dos sonhos tenham conteúdo visual e que em cegos de nascimento, predominam sonhos táteis e auditivos.

A gênese do sonho começa com a revocação (ou liberação) de múltiplas informações memorizadas que se combinam, formando uma configuração neural específica para cada combinação; a identificação desta configuração pelo processo consciente constitui um sonho, que gera, subseqüentemente, um padrão de atividade onírica, no qual é detectável pela análise dos movimentos que o exprimem [66]. A fase do sono em que ocorrem os sonhos caracteriza-se eletrofisiologicamente por uma dessincronização das ondas cerebrais, isto é, atividade com baixa voltagem e alta frequência, similar aos padrões observados no estado de vigília. O sono sincronizado diz respeito a uma atividade (não onírica) de ondas cerebrais com alta voltagem e baixa frequência, correspondentes aos períodos de sono profundo. Há evidências de que na fase de sono sincronizado ocorre a restauração de funções corporais, enquanto que, as funções do sono dessincronizado estão relacionadas ao reprocessamento de informação e na consolidação da memória.

Em humanos, durante o sono dessincronizado, observa-se a movimentação de vários segmentos do corpo, ocorrência de frequências cardíaca e respiratória semelhantes ao período de vigília, bem como a atividade intermitente de movimentos oculares rápidos. Por esse motivo, o sono dessincronizado também é chamado sono REM (*rapid eyes movement*). Experiências com coelhos, gatos e ratos no período de sono dessincronizado também verificaram a existência de padrões oníricos. Investigações feitas com animais permitem, em comparação com humanos, maior variedade de procedimentos.

Em ratos, o ritmo *theta* hipocampal, cuja frequência oscila entre 4 a 7 Hz, é uma das principais características eletrofisiológicas do estado de vigília e do sono dessincronizado. No experimento, registrou-se ondas *theta* nestes períodos em quatro regiões distintas da formação hipocampal. Estas denominam-se *ca1d*, *ca1e*, *ca3d* e *ca3e*; para análise utilizou-se séries temporais com 44800 pontos cada, medidas simultaneamente nestas regiões do hipocampo no período de sono dessincronizado. A figura 5.1 mostra o animal em preparação para implantação dos eletrodos de



FIGURA 5.1: Preparação do animal para implantação dos eletrodos de medida.



FIGURA 5.2: Registro da atividade eletrofisiológica na formação hipocampal.

medida e, a figura 5.2, mostra os eletrodos implantados durante registro das ondas *theta* no estado de vigília. Para maiores detalhes sobre o experimento, aconselha-se ver em Valle [66].

5.3 Caracterização das Séries Temporais Empíricas

Apresentam-se agora os resultados obtidos na classificação quanto ao mecanismo gerador, assim como os obtidos na aplicação dos testes propostos para caracterização das séries temporais empíricas quanto aos critérios de estacionariedade e linearidade. Em [67], utilizou-se estas técnicas, com exceção do método *cross prediction error*, na análise de séries provenientes da simulação do mercado financeiro por meio da teoria dos jogos, mais precisamente o jogo da minoria. Neste mesmo trabalho também foram utilizadas séries temporais previamente definidas para validação dos testes.

Na tabela 5.1, mostram-se os resultados obtidos para a complexidade de Lempel-Ziv, as especificações da modelagem linear ARIMA, as estatísticas calculadas na verificação de raiz unitária e as decisões tomadas em relação aos testes de linearidade.

TABELA 5.1: Sumário dos resultados.

série temporal	$\mathcal{C}_{LZ}$	ARIMA(p, d, q)	raiz unitária	linearidade
<i>ca1d</i>	0,4288	(4, 0, 4)	$ t_{\hat{\gamma}} = 23,4300$	não-linear
<i>ca1e</i>	0,4612	(1, 0, 4)	$ t_{\hat{\rho}} = 15,2679$	inconclusivo
<i>ca3d</i>	0,4624	(4, 0, 3)	$ t_{\hat{\gamma}} = 26,8270$	não-linear
<i>ca3e</i>	0,4816	(4, 0, 4)	$ t_{\hat{\gamma}} = 25,7273$	não-linear

Observa-se que a complexidade de Lempel-Ziv acabou revelando-se útil na caracterização de determinismo, as quatro séries empíricas utilizadas no trabalho apresentaram complexidades típicas de séries caóticas. Exemplos para comparação são dados pela complexidade da série associada ao mapeamento de Hénon, $\mathcal{C}_{LZ} = 0,5886$ e a complexidade da série associada ao sistema de Rössler, $\mathcal{C}_{LZ} = 0,0572$, apresentada anteriormente no primeiro capítulo juntamente com a complexidade da série associada ao sistema de Lorenz, $\mathcal{C}_{LZ} = 0,0639$. Além disso, observa-se o fato de que as quatro séries empíricas possuem complexidades muito próximas, indicando uma possível similaridade dinâmica de acordo com a construção desta técnica algorítmica. Basicamente, a quantidade de dígitos inseridos e copiados no processo

de reconstrução, caracteriza essa similaridade entre as séries [14, 23].²² A conversão da série temporal em uma sequência de dígitos binários foi feita por meio de uma análise comparativa do valor médio da série temporal; substituíram-se observações superiores e inferiores ao valor médio da série pelos dígitos 1 e 0, respectivamente.

Para seleção dos modelos lineares foram utilizados os critérios de informação de Akaike e Schwarz [40, 41]. A estimativa do mínimo destas informações selecionam os modelos com melhores ajustes na amostra segundo esses critérios. Todas as séries, de acordo com o teste da raiz unitária de Dickey-Fuller, são integradas de ordem zero, ou $\mathcal{I}(0)$, conforme as estatísticas obtidas. Isto significa que, para uma aproximação linear, as séries temporais empíricas são definidas como estacionárias, no sentido de apresentarem média e variância constantes. A importância da verificação de raiz unitária neste estudo, restringiu-se somente ao cálculo das especificações ARIMA utilizadas na aplicação da estatística BDS, cujos resultados são mostrados na tabela 5.2, juntamente com os obtidos pelo método *surrogate data*.

Os três testes de hipóteses são bicaudais com nível de significância de $ns = 5\%$. Na aplicação do teste de Dickey-Fuller, o valor crítico referente à especificação sem intercepto e sem tendência, no nível de significância adotado, foi $t_c = 1,9393$. O limiar crítico utilizado na estatística BDS, determinado a partir da distribuição $N(0,1)$, foi $z_{0,05} = 1,9599$; e para o método *surrogate data*, o limiar crítico foi $t_{0,05} = 2,0244$. Determinou-se o número de graus de liberdade da distribuição t_{K-1} pela condição $K = 2/ns - 1$ (onde K é o número de *surrogates*), proposta em Kantz e Schreiber [4].

Os observáveis não-lineares utilizados na aplicação do método *surrogate data* são definidos como

$$\zeta = \langle (x_{t-1} - 2x_t + x_{t+1})^4 \rangle, \quad (5.1)$$

$$\xi = \langle (x_t - x_{t+1})^3 \rangle. \quad (5.2)$$

A opção por estas duas medidas foi motivada pelo fato que na condição $K \approx 10^3$ as distribuições geradas pelos observáveis $\{\zeta'_p\}_{p=1}^K$ e $\{\xi'_j\}_{j=1}^K$, convergem igualmente para distribuições do tipo $N(0,1)$. Evidências numéricas realizadas nesse contexto indicam que esta é frequentemente uma aproximação razoável.

A não-linearidade foi verificada em três séries empíricas pelos resultados obtidos, sendo estes inconclusivos para a série *cale*. A necessidade de mais técnicas de classificação quanto à este critério, para obtenção de resultados mais confiáveis, fica em evidência por tal motivo e será avaliada em futuras análises. Na hipótese de que

²²Uma hipótese a se verificar é de que estas séries temporais empíricas têm o mesmo vocabulário.

TABELA 5.2: Valores encontrados para a estatística BDS e o método *surrogate data*.

série temporal	ε	$ W_{m,\ell}(\varepsilon) $, $m : 2 - 7$	$ \zeta'_0 $	$ \xi'_0 $
<i>ca1d</i>	0,0391	3,5865; 3,1017; 2,9409 2,6644; 2,9698; 3,3646	6,4595	4,7234
<i>ca1e</i>	0,0734	2,6075; 1,9804; 1,2493 1,8591; 3,3565; 5,6746	1,4646	2,3978
<i>ca3d</i>	0,0362	10,0565; 12,0933; 15,8778 17,3050; 18,9862; 20,0271	3,3780	2,6198
<i>ca3e</i>	0,0335	6,7539; 8,4235; 8,7387 9,1534; 9,6140; 8,9526	4,1774	4,1038

as séries empíricas apresentam dinâmicas similares, baseada nos valores encontrados para a complexidade de Lempel-Ziv, optou-se por classificar a série temporal *ca1e* também como não-linear.

Para o diagrama de recorrência, detectou-se estruturas típicas de séries caóticas na aplicação do método nas quatro séries empíricas. As regiões de *close returns*, que indicam a existência de órbitas periódicas instáveis, foram observadas em quantidade e tamanho menores se comparadas com diagramas obtidos na aplicação de séries geradas numericamente por modelos caóticos. Um exemplo desta situação, para comparação, é apresentado na figura 5.3 para a série associada ao mapeamento de Hénon. Por esse motivo, adotou-se, na tentativa de definir melhor as aplicações do método nas séries temporais empíricas, a expressão “recorrência fraca”. Na figura 5.4, mostra-se o diagrama de recorrência para a série *ca1d*. Diagramas similares foram observados para as outras três séries empíricas.

Finalmente, na aplicação do método *cross prediction error*, as superfícies de erro, que representam a estatística γ_{ij} calculada por meio dos segmentos de treinamento $\mathcal{S}_i^\ell$ e de teste $\mathcal{S}_j^\ell$, indicaram somente na série *ca1d* evidências de não-estacionariedade dinâmica. Observa-se na figura 5.5 que os valores da estatística de erro γ_{ij} foram altos devido basicamente a um segmento, o indexado por $i = 8$. Este segmento, forneceu erros altos tanto na etapa de treinamento quanto na de teste, mostrando-se inadequado para a avaliação da estacionariedade dinâmica de outros segmentos, como observado em pontos distribuídos ao longo de $j = 8$. Os

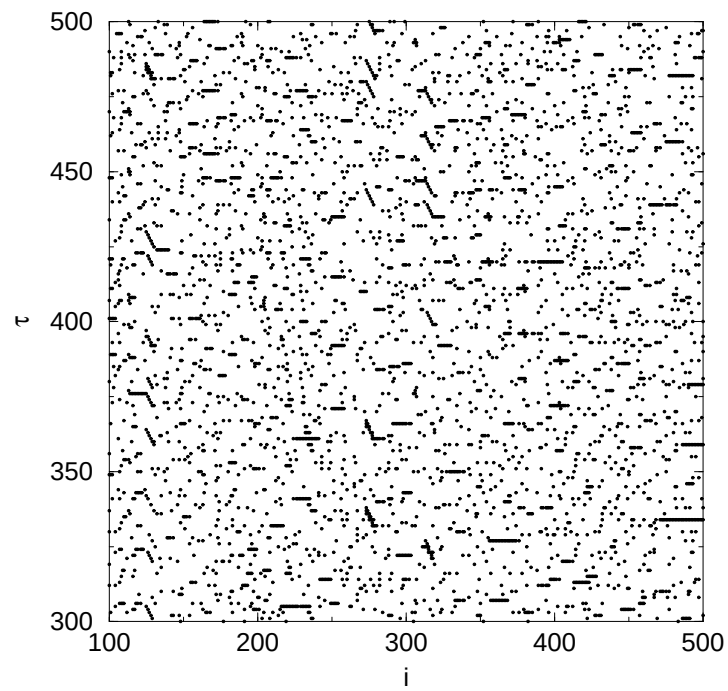
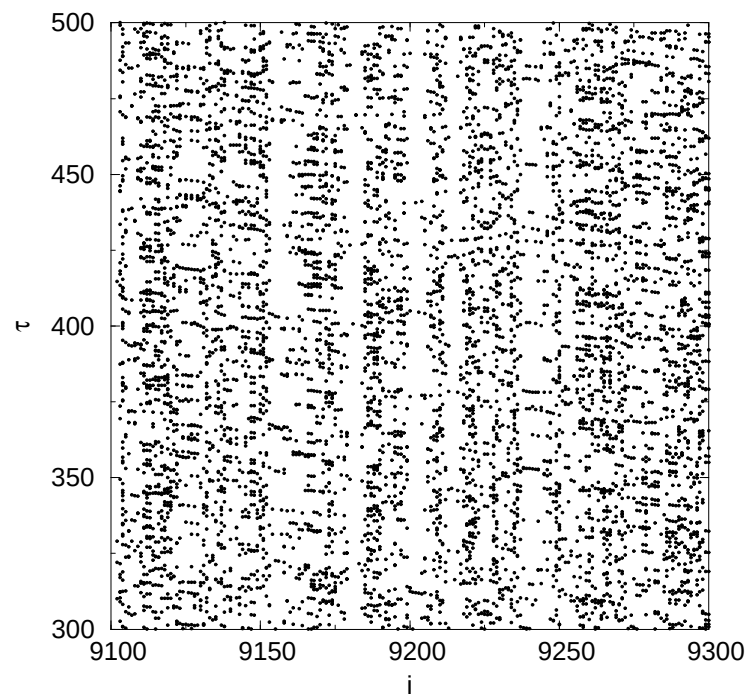


FIGURA 5.3: Diagrama de recorrência para a série associada ao mapa de Hénon.

FIGURA 5.4: Diagrama de recorrência para a série empírica $ca1d$.

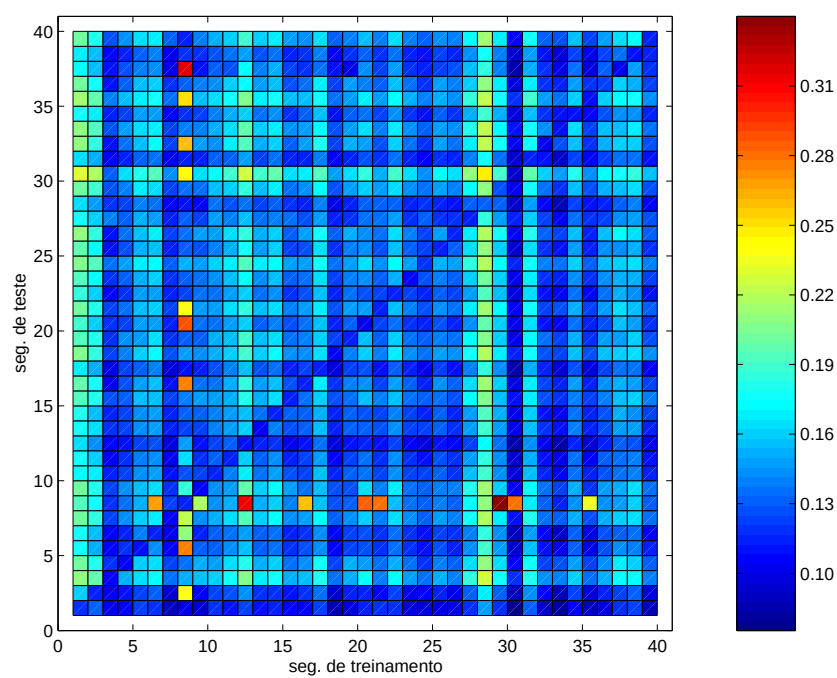
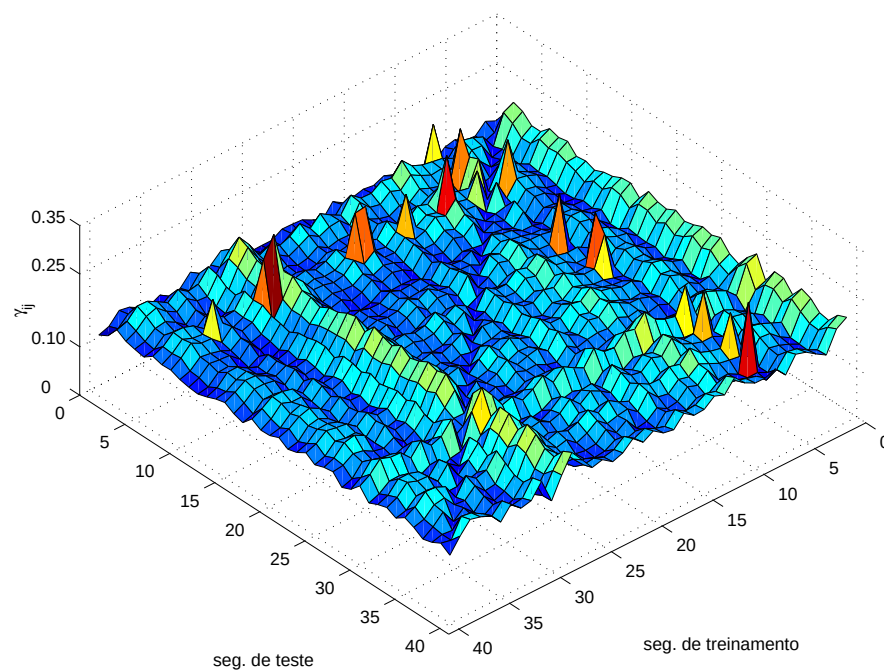


FIGURA 5.5: *Cross prediction error* aplicado nos segmentos obtidos a partir da série *ca1d*. Erros acima de 0,25 são considerados altos.

parâmetros utilizados na aplicação do método foram definidos como $\ell = 1120$, $m = 3$ e $\varepsilon = 0,15$. Para melhorar o desempenho do treinamento das redes neurais artificiais, foram retirados os segmentos $\{\mathcal{S}_k^\ell\}_{k=1}^8$ da série *ca1d*, sendo feita a modelagem com os segmentos restantes, ou seja, os indexados por $k = 9, 10, \dots, 40$; também foram retirados os segmentos $\{\mathcal{S}_k^\ell\}_{k=1}^8$ das outras três séries empíricas, pois a proposta de utilização das redes neurais artificiais no trabalho é de descrever um modelo multivariado não-linear. Devido à fragilidade do estimador que auxilia no cálculo da estatística γ_{ij} , as séries empíricas passaram antes por um processo de filtragem de ruído. Na aplicação do método para as outras três séries empíricas obteve-se superfícies típicas de séries temporais estacionárias, no que diz respeito à escala da estatística de erro γ_{ij} .

5.4 Filtragem de Ruído

Um fator importante na análise de séries empíricas é a constante presença de ruído. Pelo menos, dois tipos de ruído podem ser caracterizados: o ruído adicionado durante o registro (ruído de amostragem) e o ruído dinâmico. O ruído de amostragem é manifestado quando faz-se a leitura de algum observável do sistema. O que se obtém é algo do tipo

$$\tilde{x}_n = x_n + \epsilon_n, \quad x_{n+1} = g(x_n), \quad (5.3)$$

onde ϵ_n representa o ruído estocástico presente no sinal $\tilde{x}_n$ e está descorrelacionado com o sinal x_n . O ruído dinâmico, por sua vez, está presente no processo de realimentação inerente à dinâmica do sistema e apresenta um nível maior de complexidade, com respeito a interpretação e tratamento dos dados registrados. Nesta situação, a dinâmica pode ser definida pelo mapeamento

$$x_{n+1} = g(\tilde{x}_n) = g(x_n + \epsilon_n), \quad (5.4)$$

indicando que o sistema é perturbado por uma variável aleatória (de baixa amplitude) em cada passo de tempo.

No caso de ruído de amostragem, uma órbita aproximadamente “limpa” relacionada à dinâmica do sistema existe e pode ser encontrada efetuando-se uma filtragem. Algo interpretado como ruído dinâmico pode ser uma componente determinística do sistema, de maior dimensionalidade e menor amplitude; mesmo que este não seja o caso, o ruído dinâmico pode ser essencial para a compreensão da dinâmica observada. Por exemplo, caso um ruído gaussiano interaja com a dinâmica, existe uma probabilidade não nula de que o ruído leve a trajetória para fora da órbita.

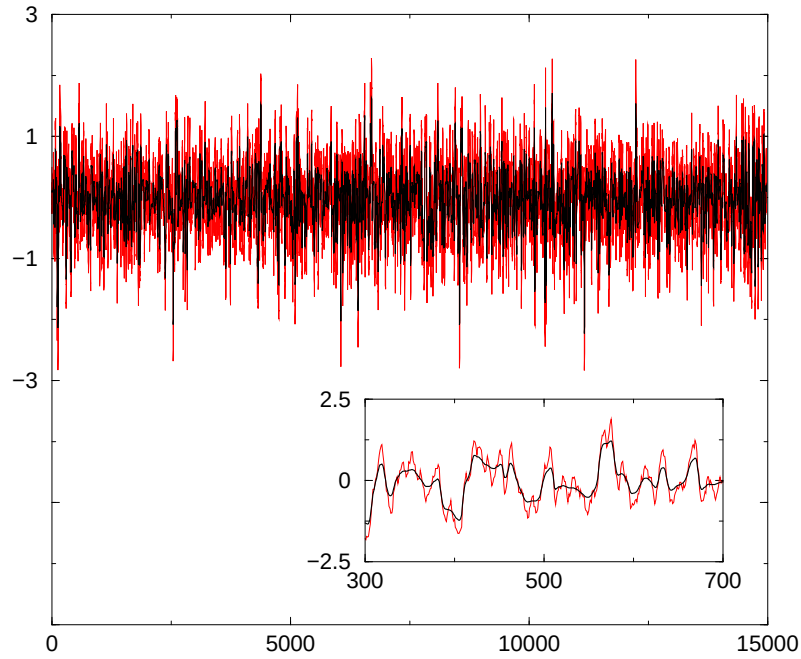


FIGURA 5.6: Série empírica *cale* contaminada por ruído (vermelho) e filtrada (preto) pelo método *de-noising*.

Qualquer tentativa de modelagem deste comportamento, sem levar em consideração o ruído, irá presumidamente falhar. Neste trabalho, não foi abordada a questão do tratamento de ruído dinâmico (*shadowing*), mas é importante mencionar que sua presença leva a consequências mais drásticas do que simplesmente esconder estruturas determinísticas de baixa escala. Para maiores detalhes aconselha-se [4].

A técnica *de-noising*, baseada na aplicação da transformada em ondeletas, tem-se mostrado uma eficiente ferramenta na filtragem de ruído em séries temporais determinísticas [48, 49, 54]. Esta técnica considera que a forma de interação do ruído seja aditiva; como descrito anteriormente no capítulo 3, este método supõe que os coeficientes da transformada que têm valores próximos de zero correspondam a ruído puro aditivo. Foi utilizada como ondeleta-mãe a função *daublet* ${}_4\psi(t)$. A opção por esta ondeleta no trabalho, ao invés de outras ondeletas-mãe, está associada com o grau de regularidade mínimo que a função necessita para representar um sinal determinístico.²³

Na figura 5.6, mostra-se parcialmente a série empírica *cale* antes (vermelho) e depois (preto) da filtragem realizada, no detalhe tem-se uma visão ampliada da

²³Entende-se regularidade, no contexto deste trabalho, como diferenciabilidade. Em particular, ${}_4\psi(t)$ é uma função de classe $C^{1,275}$.

TABELA 5.3: Caracterização dos resíduos obtidos e complexidade de Lempel-Ziv após filtragem realizada.

série temporal	$\mathcal{C}_{LZ}^\varepsilon$	ε	$ W_{m,\ell}(\varepsilon) $, $m : 2 - 4$	$\mathcal{C}_{LZ}^x$
<i>ca1d</i>	0,8463	0,1245	1,6149; 1,1064; 0,7735	0,2853
<i>ca1e</i>	0,8557	0,1328	1,3311; 1,0780; 1,1652	0,2751
<i>ca3d</i>	0,8237	0,1386	1,5128; 0,7420; 0,6722	0,3130
<i>ca3e</i>	0,8789	0,1269	0,9767; 1,4748; 1,2439	0,2973

diferença entre as séries original e filtrada. Percebe-se que a série filtrada oscila menos que a série contaminada por ruído; a escolha do esquema utilizado para seleção dos coeficientes de ondeleta, o *soft threshold*, foi devido a este comportamento, desejado na tentativa de minimizar o fenômeno de *overfitting*, mencionado anteriormente no capítulo 4.

Na tabela 5.3, mostram-se os valores para a complexidade de Lempel-Ziv $\mathcal{C}_{LZ}^\varepsilon$ e estatística BDS, úteis na caracterização dos resíduos. Na mesma tabela, apresentam-se também os valores da complexidade de Lempel-Ziv $\mathcal{C}_{LZ}^x$ para as séries temporais empíricas após a filtragem. Comparando esses valores com os mostrados na tabela 5.1, tem-se uma idéia do quanto este método algorítmico é sensível à presença de ruído, não chegando a comprometer a decisão tomada no que diz respeito à classificação do mecanismo gerador. O diagrama de recorrência também forneceu resultados favoráveis na aplicação dos resíduos obtidos. Para detalhes sobre os efeitos da presença de ruído, bem como de outros métodos de redução, aconselham-se [4, 7, 8, 9].

5.5 Modelagem Não-Linear

Para modelagem das séries temporais empíricas, foram utilizadas redes neurais com apenas um conjunto oculto de neurônios artificiais. Como já mencionado no capítulo 4, esta arquitetura é adequada para modelagem de séries provenientes de mapeamentos dados por funções contínuas [64]. Em todos os neurônios artificiais, pertencentes ao conjunto oculto e na saída da rede, a função de ativação utilizada foi a tangente hiperbólica. Esta opção foi devida à rápida convergência na fase de treinamento em comparação com outras funções de ativação [65]. Utilizou-se como parâmetro da taxa de aprendizagem $\eta = 10^{-4}$.

Para cada série empírica, destinou-se 90% da amostragem total (32256 pontos) ao

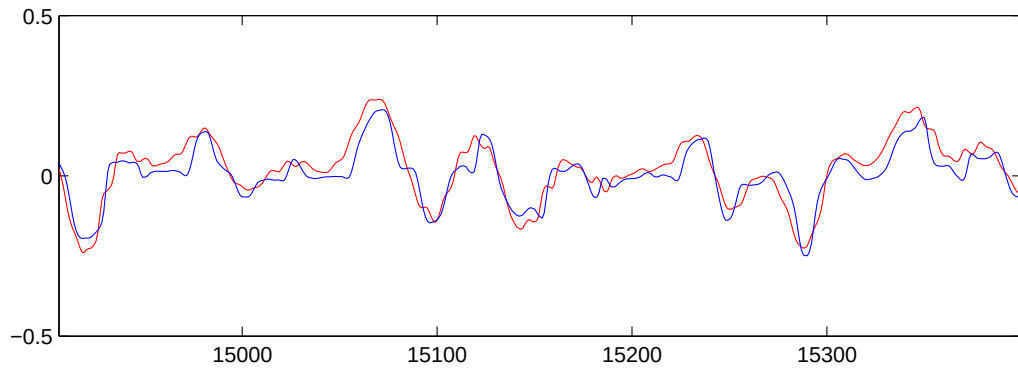


FIGURA 5.7: Série empírica $ca3e$ (azul) e modelagem (vermelho) realizada por rede neural artificial utilizando as entradas $ca3d_{n-\tau}$, $ca3e_{n-\tau}$ e $ca3e_{n-2\tau}$.

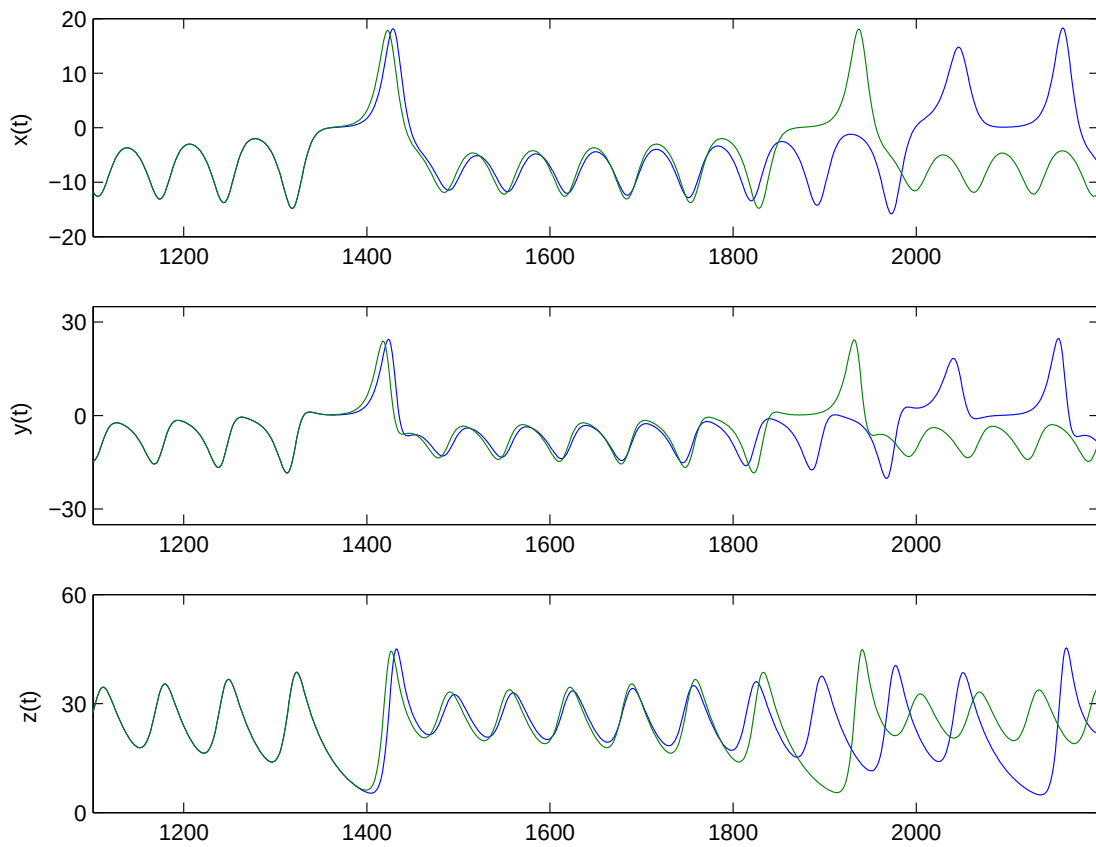


FIGURA 5.8: Soluções do sistema de Lorenz para duas condições iniciais diferentes. Neste exemplo, utilizou-se $\delta x(0) \approx 10^{-3}$.

treinamento e a outra parte restante (3584 pontos) foi utilizada na fase de validação, também denominada *recall*. De acordo com o teorema de Kolmogorov, o número de entradas na rede é dado pela dimensão de imersão, que por sua vez, é determinada por meio do cálculo da dimensão de correlação; lembra-se que nesta etapa as séries já foram filtradas. Para as quatro séries empíricas, a dimensão de imersão tem o valor $m = 3$. Deste modo, ainda utilizando-se do mesmo teorema, tem-se sete neurônios artificiais no conjunto oculto.

Nessas condições, onde para cada série, pretende-se estimar um modelo multivariado não-linear, obtém-se no mínimo 220 configurações diferentes de padrões de entrada do tipo $\{x_{n-\tau}, x_{n-2\tau}, x_{n-3\tau}\}$, por exemplo, que auxiliam na determinação de x_n .²⁴ O parâmetro de atraso τ foi escolhido pelo critério da localização do primeiro zero da função de autocorrelação da série temporal [68]; para as séries empíricas utilizadas no trabalho obteve-se $\tau = 8$. Diferentes critérios podem ser utilizados para interromper o treinamento da rede, neste trabalho considerou-se quando o erro quadrático médio não mais diminuir sistematicamente com as devidas modificações nos pesos sinápticos e vieses. Na figura 5.7 mostra-se a modelagem feita para a série empírica *ca3e*.

Para escolha do padrão de entrada ideal, optou-se pelo que forneceu o menor erro quadrático médio na fase de validação, obtendo o sistema

$$\begin{cases} ca1d_n = f(ca1d_{n-\tau}, ca3d_{n-2\tau}, ca3e_{n-\tau}) \\ ca1e_n = g(ca1e_{n-\tau}, ca3d_{n-\tau}, ca3e_{n-3\tau}) \\ ca3d_n = h(ca3d_{n-\tau}, ca3d_{n-2\tau}, ca3e_{n-2\tau}) \\ ca3e_n = s(ca3d_{n-\tau}, ca3e_{n-\tau}, ca3e_{n-2\tau}) \end{cases}, \quad (5.5)$$

onde f, g, h e s representam funções não-lineares de classe C^1 .

Finalmente, lembra-se que por causa da incerteza (sempre presente) das condições iniciais, na modelagem de séries temporais caóticas, a previsão após um tempo determinado (que estará associado ao expoente característico de Lyapunov do sistema) irá presumidamente divergir em relação à série original. Na figura 5.8 ilustra-se as soluções do sistema de Lorenz para duas condições iniciais próximas.

5.6 Considerações Finais

Foram obtidas fortes evidências de que o mecanismo gerador das ondas *theta* medidas na formação hipocampal do rato é determinístico. Além do mais, a não-

²⁴Neste caso, para um modelo auto-regressivo não-linear ou com entradas dadas por qualquer uma das outras três séries empíricas, o número de configurações diferentes de padrões de entrada é dado por $\binom{p}{m} = \frac{p!}{m!(p-m)!}$, com $p = 12$ e $m = 3$.

linearidade, que seria um primeiro passo na detecção de caos, também foi verificada pelos testes propostos, sendo estes inconclusivos para apenas uma das quatro séries temporais empíricas analisadas. É interessante mencionar que, até o momento, ondas *theta* ainda não tinham sido estudadas quanto ao seu caráter determinístico, sendo pioneiras as evidências de caos aqui verificadas e apresentadas.

Como consequência desta caracterização, propõe-se uma origem para o padrão *theta* registrado, que foi determinada pela análise dos funcionais do sistema (5.5), obtido na modelagem não-paramétrica feita por redes neurais artificiais. Logo, definiu-se as regiões *ca3d* e *ca3e* como sendo as possíveis origens do padrão *theta*. Futuramente, pretende-se modelar este mesmo sistema, aparentemente dinâmico, utilizando-se da aplicação do método *cluster-weighted modelling* [69, 70], pois este também mostrou-se útil na modelagem de séries determinísticas [67].

O presente trabalho descreveu um conjunto de métodos aplicáveis na caracterização e modelagem de séries empíricas, mais especificamente, séries temporais caóticas. Focou-se numa breve incursão em métodos alternativos pouco utilizados na literatura, baseados nas propriedades topológicas de atratores estranhos e na teoria da complexidade de algoritmos, para detecção de caos em sinais experimentais; onde basicamente faz-se a detecção com base em medidas métricas. Este estudo não tem como propósito abordar todas as técnicas disponíveis sobre o assunto, mas sim propor uma metodologia consistente com a teoria do caos, fornecendo uma análise completa quando nela for aplicada uma série temporal dinamicamente caótica. No fluxograma apresentado no apêndice B, esquematiza-se esta metodologia. Por fim, lembra-se que a distinção entre a natureza caótica ou estocástica do sistema em estudo é imprescindível por um motivo mais prático do que filosófico; é necessário que se conheça o mecanismo gerador de um sistema para que se possa utilizar as ferramentas adequadas na extração de informações do mesmo.

Apêndice A

Propriedades das Ondeletas de Daubechies

A.1 Condições sobre os Coeficientes de Filtro

Algumas propriedades úteis em sistemas de ondeletas são satisfeitas quando impõem-se determinadas condições no cálculo dos coeficientes de filtro $\{\ell_k\}$.

Para se definir de forma única cada uma das funções escala, é usual impor que

$$\int_{-\infty}^{+\infty} {}_N\phi(t)dt = 1. \quad (\text{A.1})$$

Esta primeira condição é geralmente conhecida como condição de consistência. Integrando a equação de dilatação tem-se

$$\int_{-\infty}^{+\infty} {}_N\phi(t)dt = \sqrt{2} \sum_{k=0}^{2N-1} \ell_k \int_{-\infty}^{+\infty} {}_N\phi(2t-k)dt = \frac{1}{\sqrt{2}} \sum_{k=0}^{2N-1} \ell_k \int_{-\infty}^{+\infty} {}_N\phi(t')dt',$$

que resulta na condição

$$\sum_{k=0}^{2N-1} \ell_k = \sqrt{2}. \quad (\text{A.2})$$

Quando se impõe que a função escala seja ortonormal a todas as suas translações inteiras, obtém-se a segunda condição. A condição de ortonormalidade é escrita como

$$\begin{aligned} \langle {}_N\phi(t), {}_N\phi(t-k) \rangle &= \int_{-\infty}^{+\infty} {}_N\phi(t) \cdot {}_N\phi^*(t-k)dt = \delta_{0k}, \\ \langle {}_N\phi(t), {}_N\phi(t-k) \rangle &= 2 \sum_{m=0}^{2N-1} \ell_m \sum_{n=0}^{2N-1} \ell_n^* \int_{-\infty}^{+\infty} {}_N\phi(2t-m) \cdot {}_N\phi^*[2(t-k)-n]dt, \\ \langle {}_N\phi(t), {}_N\phi(t-k) \rangle &= \sum_{m=0}^{2N-1} \ell_m \sum_{n=0}^{2N-1} \ell_n^* \int_{-\infty}^{+\infty} {}_N\phi(t'-m) \cdot {}_N\phi^*[t'-(n+2k)]dt', \\ \langle {}_N\phi(t), {}_N\phi(t-k) \rangle &= \sum_{m=0}^{2N-1} \ell_m \sum_{n=0}^{2N-1} \ell_n^* \delta_{m,n+2k}, \end{aligned}$$

que resulta na condição

$$\sum_{n=0}^{2N-1} \ell_n^* \ell_{n+2k} = \delta_{0k} , \quad \text{com } k \in \mathbb{Z} . \quad (\text{A.3})$$

É importante verificar que esta condição, em conjunto com as definições (3.30) e (3.31), implica a ortonormalidade desejada entre a ondeleta-mãe ${}_N\psi(t)$ e todas as suas translações inteiras. Pode-se escrever

$$\begin{aligned} \langle {}_N\psi(t), {}_N\psi(t-k) \rangle &= \int_{-\infty}^{+\infty} {}_N\psi(t) \cdot {}_N\psi^*(t-k) dt = \delta_{0k} , \\ \langle {}_N\psi(t), {}_N\psi(t-k) \rangle &= 2 \sum_{m=0}^{2N-1} h_m \sum_{n=0}^{2N-1} h_n^* \int_{-\infty}^{+\infty} {}_N\phi(2t-m) \cdot {}_N\phi^*[2(t-k)-n] dt , \\ \langle {}_N\psi(t), {}_N\psi(t-k) \rangle &= \sum_{m=0}^{2N-1} (-1)^m \ell_{2N-1-m}^* \sum_{n=0}^{2N-1} (-1)^n \ell_{2N-1-n} \delta_{m,n+2k} , \\ \langle {}_N\psi(t), {}_N\psi(t-k) \rangle &= \sum_{n=0}^{2N-1} (-1)^{n+2k} (-1)^n \ell_{2N-1-(n+2k)}^* \ell_{2N-1-n} , \\ \langle {}_N\psi(t), {}_N\psi(t-k) \rangle &= \sum_{n=0}^{2N-1} (-1)^{2n+2k} \ell_{2N-1-n-2k}^* \ell_{2N-1-n} , \end{aligned}$$

atendendo a condição (A.3).

As condições (A.2) e (A.3) permitem escrever $N+1$ equações relacionando os $2N$ coeficientes ℓ_k presentes na definição de ${}_N\phi(t)$ e ${}_N\psi(t)$. Para determinar de forma única estes coeficientes, torna-se necessário estabelecer uma terceira condição que permite obter as $N-1$ equações que faltam. Esta condição aparece quando se impõe que qualquer polinômio do tipo

$$f(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_{N-1} t^{N-1} ,$$

pode ser representado, de uma forma exata, pela expansão

$$f(t) = \sum_{k \in \mathbb{Z}} a_k {}_N\phi(t-k) .$$

Multiplicando ambos os membros da equação por ${}_N\psi(t)$ e logo após integrando, obtém-se

$$\begin{aligned} \int_{-\infty}^{+\infty} f(t) {}_N\psi(t) dt &= \sum_{k \in \mathbb{Z}} a_k \int_{-\infty}^{+\infty} {}_N\phi(t-k) \cdot {}_N\psi(t) dt = 0 , \\ \int_{-\infty}^{+\infty} f(t) {}_N\psi(t) dt &= \int_{-\infty}^{+\infty} (\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_{N-1} t^{N-1}) {}_N\psi(t) dt = 0 . \end{aligned}$$

Esta igualdade tem que ser válida para todos os valores α_j , para $j = 0, 1, \dots, N-1$. Escolhendo $\alpha_j = 1$ e $\alpha_p = 0$, para $j \neq p$, obtém-se a condição

$$\int_{-\infty}^{+\infty} t^j {}_N\psi(t) dt = 0.$$

Substituindo a definição da ondeleta-mãe, tem-se

$$\int_{-\infty}^{+\infty} t^j {}_N\psi(t) dt = \sum_{k=0}^{2N-1} (-1)^k \ell_{2N-1-k}^* \int_{-\infty}^{+\infty} t^j {}_N\phi(2t-k) dt,$$

fazendo $t' = 2t - k$ tem-se

$$\int_{-\infty}^{+\infty} t^j {}_N\psi(t) dt = \frac{1}{2} \sum_{k=0}^{2N-1} (-1)^k \ell_{2N-1-k}^* \int_{-\infty}^{+\infty} \left(\frac{t'+k}{2} \right)^j {}_N\phi(t') dt'. \quad (\text{A.4})$$

Utilizando a igualdade

$$(a+b)^m = \sum_{m'=0}^m \binom{m}{m'} a^{m'} b^{m-m'}, \quad (\text{A.5})$$

obtém-se para relação (A.4) a expressão

$$\int_{-\infty}^{+\infty} t^j {}_N\psi(t) dt = 2^{-j-1} \sum_{r=0}^j \binom{j}{r} \sum_{k=0}^{2N-1} (-1)^k \ell_{2N-1-k}^* k^{j-r} \int_{-\infty}^{+\infty} t'^r {}_N\phi(t') dt'.$$

As integrais do tipo $\int_{-\infty}^{+\infty} t^r {}_N\phi(t) dt$ são não nulas, uma vez que se pretende garantir que as funções $f(t) = t^r$, com $r = 0, 1, \dots, N-1$, possam ser representadas como combinação linear da função escala e das suas translações inteiras. Consequentemente, escreve-se

$$2^{-j-1} \sum_{r=0}^j \binom{j}{r} \sum_{k=0}^{2N-1} (-1)^k \ell_{2N-1-k}^* k^{j-r} = 0.$$

Escrevendo esta equação para os sucessivos valores de j , obtém-se

$$\sum_{k=0}^{2N-1} (-1)^k \ell_{2N-1-k}^* k^j = 0, \quad (\text{A.6})$$

fazendo a mudança de coordenadas $m = 2N-1-k$ e tomando o complexo conjugado em ambos os lados, tem-se

$$\sum_{m=0}^{2N-1} (-1)^{2N-1-m} \ell_m (2N-1-m)^j = 0. \quad (\text{A.7})$$

TABELA A.1: Valores dos coeficientes de filtro para a *daublet* ${}_4\psi(t)$.

k	ℓ_k	k	ℓ_k
0	0,23037781330869	4	-0,18703481171906
1	0,71484657055304	5	0,03084138183614
2	0,63088076792932	6	0,03288301166690
3	-0,02798376941711	7	-0,01059740178481

Usando novamente a igualdade (A.5) na equação (A.7) e escrevendo os sucessivos valores de j , finalmente obtém-se a condição

$$\sum_{m=0}^{2N-1} (-1)^{2N-1-m} \ell_m (-m)^j = 0. \quad (\text{A.8})$$

A condição (A.8) permite escrever N equações relacionando os coeficientes de filtro. Lembra-se que uma vez estabelecidas as condições (A.2) e (A.3), faltam $N - 1$ equações para o cálculo de tais coeficientes. Haverá, portanto, uma equação redundante.

Para ilustrar a aplicação das condições que permitem obter os coeficientes de filtro, apresenta-se as equações que conduzem à determinação dos diferentes valores de ℓ_k , $k = 0, 1, \dots, 7$, para a família de ondeletas com $N = 4$. A condição (A.2) permite escrever

$$\ell_0 + \ell_1 + \ell_2 + \ell_3 + \ell_4 + \ell_5 + \ell_6 + \ell_7 = \sqrt{2}.$$

De (A.3), com interesse de obter coeficientes reais, tem-se

$$\begin{aligned} \ell_0^2 + \ell_1^2 + \ell_2^2 + \ell_3^2 + \ell_4^2 + \ell_5^2 + \ell_6^2 + \ell_7^2 &= 1 \quad (k = 0), \\ \ell_0 \ell_2 + \ell_1 \ell_3 + \ell_2 \ell_4 + \ell_3 \ell_5 + \ell_4 \ell_6 + \ell_5 \ell_7 &= 0 \quad (k = 1), \\ \ell_0 \ell_4 + \ell_1 \ell_5 + \ell_2 \ell_6 + \ell_3 \ell_7 &= 0 \quad (k = 2), \\ \ell_0 \ell_6 + \ell_1 \ell_7 &= 0 \quad (k = 3). \end{aligned}$$

Finalmente, de (A.8) tem-se

$$\begin{aligned} -\ell_1 + 2\ell_2 - 3\ell_3 + 4\ell_4 - 5\ell_5 + 6\ell_6 - 7\ell_7 &= 0 \quad (j = 1), \\ \ell_1 - 4\ell_2 + 9\ell_3 - 16\ell_4 + 25\ell_5 - 36\ell_6 + 49\ell_7 &= 0 \quad (j = 2), \\ -\ell_1 + 8\ell_2 - 27\ell_3 + 64\ell_4 - 125\ell_5 + 216\ell_6 - 343\ell_7 &= 0 \quad (j = 3). \end{aligned}$$

A resolução deste sistema não-linear de oito equações permitirá recuperar os valores indicados na tabela A.1.

A.2 Geração das Funções Escala

Uma vez calculados os valores dos coeficientes de filtro, é possível gerar a função escala por meio da equação de dilatação (3.29). Não existindo solução analítica para esta equação, torna-se necessário recorrer a um procedimento numérico para obter a função ${}_N\phi(t)$. O método de geração mais eficaz, e conseqüentemente o mais utilizado, foi proposto por Strang [51]. Este método baseia-se na aplicação recursiva da equação de dilatação (3.29).

O método de Strang permite calcular, primeiramente, o valor que a função escala assume em cada um dos inteiros pertencentes ao intervalo em que se encontra definida. Uma vez obtida esta informação, a aplicação sucessiva da equação (3.29) permite obter os valores que ${}_N\phi(t)$ assume em todos os pontos diádicos $t = k \times 2^{-j}$, com $(j, k) \in \mathbb{Z}$.

Lembra-se, do capítulo 3, que o suporte da função escala ${}_N\phi(t)$ é definido como $\text{supp } {}_N\phi(t) = [0, 2N - 1]$. Como conseqüência, verifica-se que ${}_N\phi(m) = 0$ para qualquer $m \notin [0, 2N - 1]$. Escrevendo a equação de dilatação (3.29) para cada um dos inteiros pertencentes ao intervalo $[0, 2N - 1]$, obtém-se

$$\begin{aligned} {}_N\phi(0) &= \sqrt{2}\ell_0 {}_N\phi(0) , \\ {}_N\phi(1) &= \sqrt{2}[\ell_0 {}_N\phi(2) + \ell_1 {}_N\phi(1) + \ell_2 {}_N\phi(0)] , \\ {}_N\phi(2) &= \sqrt{2}[\ell_0 {}_N\phi(4) + \ell_1 {}_N\phi(3) + \ell_2 {}_N\phi(2) + \ell_3 {}_N\phi(1) + \ell_4 {}_N\phi(0)] , \\ &\vdots \\ {}_N\phi(2N - 2) &= \sqrt{2}[\ell_{2N-3} {}_N\phi(2N - 1) + \ell_{2N-2} {}_N\phi(2N - 2) + \ell_{2N-1} {}_N\phi(2N - 3)] , \\ {}_N\phi(2N - 1) &= \sqrt{2}\ell_{2N-1} {}_N\phi(2N - 1) . \end{aligned}$$

Das equações anteriores, e tendo em atenção que $\ell_0 \neq 0$ e $\ell_{2N-1} \neq 0$, conclui-se que ${}_N\phi(0) = {}_N\phi(2N - 1) = 0$.

Para os restantes inteiros, $1 \leq j \leq 2N - 2$, as igualdades anteriores podem ser escritas matricialmente na forma

$$\begin{pmatrix} \ell_1 & \ell_0 & 0 & \dots & 0 & 0 & 0 \\ \ell_3 & \ell_2 & \ell_1 & \dots & 0 & 0 & 0 \\ \ell_5 & \ell_4 & \ell_3 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \ell_{2N-4} & \ell_{2N-5} & \ell_{2N-6} \\ 0 & 0 & 0 & \dots & \ell_{2N-2} & \ell_{2N-3} & \ell_{2N-4} \\ 0 & 0 & 0 & \dots & 0 & \ell_{2N-1} & \ell_{2N-2} \end{pmatrix} \times \begin{pmatrix} {}_N\phi(1) \\ {}_N\phi(2) \\ {}_N\phi(3) \\ \vdots \\ {}_N\phi(2N - 4) \\ {}_N\phi(2N - 3) \\ {}_N\phi(2N - 2) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} {}_N\phi(1) \\ {}_N\phi(2) \\ {}_N\phi(3) \\ \vdots \\ {}_N\phi(2N - 4) \\ {}_N\phi(2N - 3) \\ {}_N\phi(2N - 2) \end{pmatrix} ,$$

ou ainda, de um modo mais compacto

$$\mathcal{L}\Phi = 2^{-1/2}\Phi . \quad (\text{A.9})$$

Os $(2N - 2) \times (2N - 2)$ elementos da matriz $\mathcal{L}$ podem ser calculados por meio da igualdade

$$\mathcal{L}_{ij} = \ell_{2i-j} . \quad (\text{A.10})$$

No vetor Φ encontram-se os valores que a função escala assume nos valores inteiros pertencentes ao intervalo $(0, 2N - 1)$. A igualdade (A.9) permite verificar que o cálculo de tais valores pode ser obtido por meio da determinação do autovetor associado ao autovalor $\lambda = 2^{-1/2}$ da matriz $\mathcal{L}$. Para tal, basta escrever (A.9) na forma

$$[\mathcal{L} - \lambda I_{(2N-2) \times (2N-2)}] \Phi = 0 , \quad (\text{A.11})$$

onde $I_{(2N-2) \times (2N-2)}$ é a matriz identidade $(2N - 2)$ -dimensional. Como em qualquer problema de valores e vetores próprios, a solução da equação acima não é única. Como consequência, torna-se necessário estabelecer uma condição adicional que permita calcular os valores definidos no vetor Φ . É usual impor que

$$\sum_{k=1}^{2N-1} \phi(k) = 1 , \quad (\text{A.12})$$

esta condição é consequência direta da condição de consistência definida na equação (A.1). Os valores que a função escala assume nos inteiros $j = 1, 2, \dots, 2N - 2$ são então calculados aplicando a condição (A.12) à solução obtida a partir de (A.11). Como foi mencionado anteriormente, uma vez calculados estes valores, a aplicação repetida da equação de dilatação (3.29) permite obter o valor da função escala em todos os pontos diádicos ($t = k \times 2^{-j}$) pertencentes ao intervalo de definição. Tem-se então

$$\begin{aligned} {}_N\phi(k/2) &= \sqrt{2} \sum_{m=0}^{2N-1} \ell_m {}_N\phi(k - m) , \\ {}_N\phi(k/4) &= \sqrt{2} \sum_{m=0}^{2N-1} \ell_m {}_N\phi[(k/2) - m] , \\ &\vdots \\ {}_N\phi(k/2^j) &= \sqrt{2} \sum_{m=0}^{2N-1} \ell_m {}_N\phi[(2k/2^j) - m] . \end{aligned}$$

Ilustra-se agora a aplicação do método de Strang na construção da função escala ${}_4\phi(t)$. O suporte desta função é dado por $\sup_{{}_4\phi(t)} = [0, 7]$. O primeiro passo do método de geração consiste na determinação do valor que a função escala assume

TABELA A.2: Valores da função escala ${}_4\phi(j)$ nos inteiros $1 \leq j \leq 6$.

j	${}_4\phi(j)$	j	${}_4\phi(j)$
1	$1,00716998 \times 10^0$	4	$-0,11764359 \times 10^{-1}$
2	$-0,33836960 \times 10^{-1}$	5	$-0,11979577 \times 10^{-2}$
3	$0,39610465 \times 10^{-1}$	6	$0,18829417 \times 10^{-4}$

nos inteiros $j = 1, 2, \dots, 6$ por meio da resolução da equação (A.11) em conjunto com a condição de normalização (A.12).

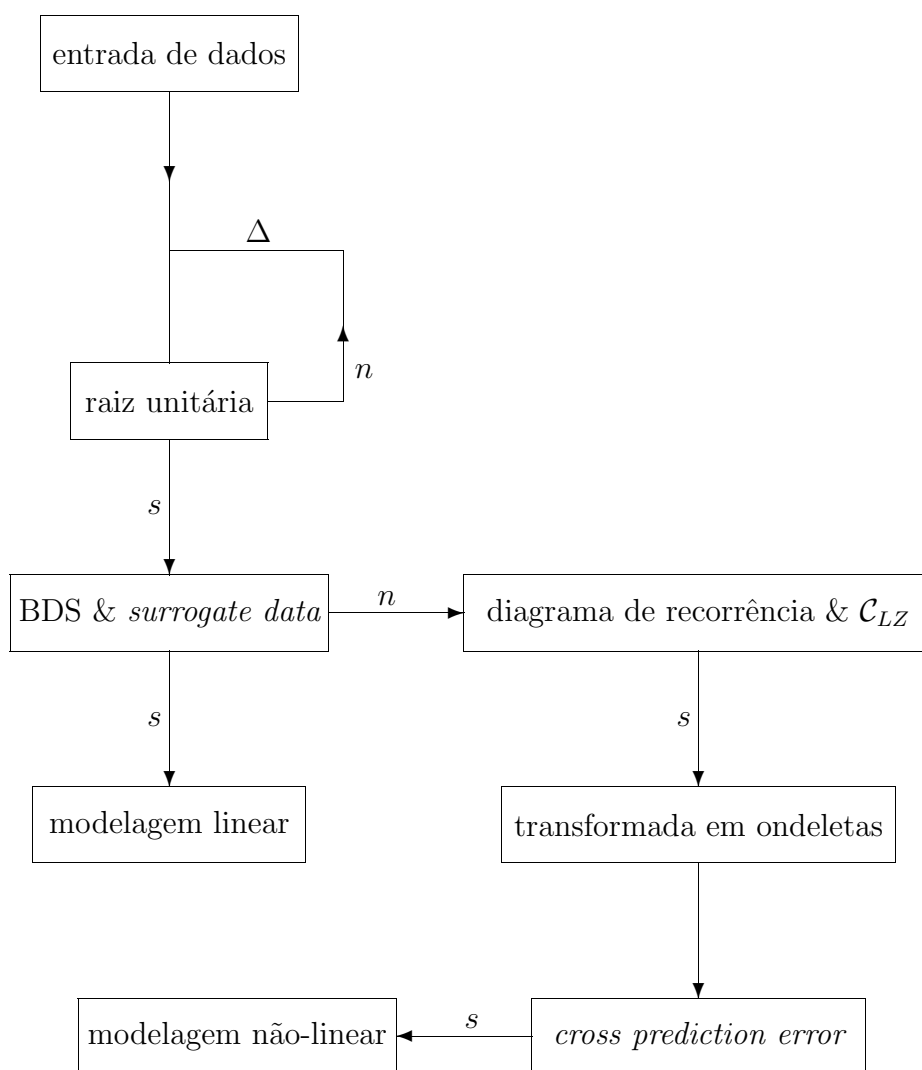
A matriz $\mathcal{L}$, neste caso em particular, é definida como

$$\mathcal{L} = \begin{pmatrix} \ell_1 & \ell_0 & 0 & 0 & 0 & 0 \\ \ell_3 & \ell_2 & \ell_1 & \ell_0 & 0 & 0 \\ \ell_5 & \ell_4 & \ell_3 & \ell_2 & \ell_1 & \ell_0 \\ \ell_7 & \ell_6 & \ell_5 & \ell_4 & \ell_3 & \ell_2 \\ 0 & 0 & \ell_7 & \ell_6 & \ell_5 & \ell_4 \\ 0 & 0 & 0 & 0 & \ell_7 & \ell_6 \end{pmatrix},$$

onde os valores de $\ell_0, \ell_1, \dots, \ell_7$ encontram-se listados na tabela A.1. Finalmente, determinando o autovetor da matriz $\mathcal{L}$ associado ao autovalor $\lambda = 2^{-1/2}$, obtém-se os valores listados na tabela A.2. A aplicação sucessiva da equação (3.29) permite determinar o valor que a função escala assume em cada um dos pontos diádicos no intervalo $[0, 7]$.

Apêndice B

Fluxograma da Metodologia Proposta



Referências Bibliográficas

- [1] W. Hardle, *Applied nonparametric regression* (Cambridge, UK: Cambridge University Press, 1990).
- [2] D.A. Dickey, W.A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root”, *Journal of the American Statistical Association* **74**, 427–431 (1979).
- [3] R. Hegger, H. Kantz, L. Matassini, T. Schreiber, “Coping with nonstationarity by overembedding”, *Physical Review Letters* **84**, 4092–4095 (2000).
- [4] H. Kantz, T. Schreiber, *Nonlinear time series analysis* (Cambridge, UK: Cambridge University Press, 1997).
- [5] T. Schreiber, “Detecting and analysing nonstationarity in a time series using nonlinear cross prediction”, *Physical Review Letters* **78**, 843–846 (1997).
- [6] J.P.M. Heald, J. Stark, “Estimation of noise levels for models of chaotic dynamical systems”, *Physical Review Letters* **84**, 2366–2369 (2000).
- [7] E.J. Kostelich, J.A. Yorke, “Noise reduction: finding the simplest dynamical system consistent with the data”, *Physica* **41D**, 183–196 (1990).
- [8] E.J. Kostelich, T. Schreiber, “Noise reduction in chaotic time series data: a survey of common methods”, *Physical Review* **48E**, 1752–1763 (1993).
- [9] T. Schreiber, H. Kantz, “Observing and predicting chaotic signals: is 2% noise too much?”, In *Predictability of complex dynamical systems*, 43–65, ed. por Y.A. Kravtsov, J.B. Kadtko (Berlin, DE: Springer-Verlag, 1996).
- [10] E. Bradley, R. Mantilla, “Recurrence plots and unstable periodic orbits”, *Chaos: an Interdisciplinary Journal of Nonlinear Science* **12**, 596–600 (2002).
- [11] C.G. Gilmore, “A new test for chaos”, *Journal of Economic Behavior and Organization* **22**, 209–237 (1993).

- [12] P. Grassberger, I. Procaccia, “Characterization of strange attractors”, *Physical Review Letters* **50**, 346–349 (1983).
- [13] V.D. Gusev, L.A. Nemytikova, N.A. Chuzhanova, “On the complexity measures of genetic sequences”, *Bioinformatics* **15**, 994–999 (1999).
- [14] F. Kaspar, H.G. Schuster, “Easily calculable measure for the complexity of spatiotemporal patterns”, *Physical Review* **36A**, 842–848 (1987).
- [15] R. Gilmore, “Topological analysis of chaotic dynamical systems”, *Reviews of Modern Physics* **70**, 1455–1529 (1998).
- [16] G.B. Mindlin, X.-J. Hou, H.G. Solari, R. Gilmore, N.B. Tufillaro, “Classification of strange attractors by integers”, *Physical Review Letters* **64**, 2350–2353 (1990).
- [17] A.N. Kolmogorov, “Three approaches to the quantitative definition of information”, *Problems of Information Transmission* **1**, 1–17 (1965).
- [18] N. Fiedler-Ferrara, C.P.C. Prado, *Caos: uma introdução* (São Paulo, SP: Edgard Blücher, 1994).
- [19] D. Ruelle, F. Takens, “On the nature of turbulence”, *Communications in Mathematical Physics* **20**, 167–192 (1971).
- [20] C. Grebogi, E. Ott, S. Pelikan, J.A. Yorke, “Strange attractors that not chaotic”, *Physica* **13D**, 261–268 (1984).
- [21] J.-P. Eckmann, D. Ruelle, “Ergodic theory of chaos and strange attractors”, *Reviews of Modern Physics* **57**, 617–656 (1985).
- [22] E. Lorenz, “Deterministic nonperiodic flow”, *Journal of Atmospheric Sciences* **20**, 130–141 (1963).
- [23] R. Badii, A. Politi, *Complexity: hierarchical structures and scaling in physics* (Cambridge, UK: Cambridge University Press, 1997).
- [24] J.L. Doob, *Stochastic process* (New York, NY: John Wiley & Sons, 1953).
- [25] J. Lamperti, *Stochastic process* (Berlin, DE: Springer-Verlag, 1977).
- [26] D.L. Cohn, *Measure theory* (Boston, MA: Birkhäuser, 1997).
- [27] S. Lang, *Real and functional analysis* (New York, NY: Springer-Verlag, 1993).

- [28] R. Durrett, *Probability: theory and examples* (Belmont, CA: Duxbury Press, 1995).
- [29] A.N. Shiryaev, *Probability* (New York, NY: Springer-Verlag, 1995).
- [30] F. Takens, “Detecting strange attractors in turbulence”, In *Dynamical systems and turbulence*, 366–381, ed. por D.A. Rand, L.S. Young (Berlin, DE: Springer-Verlag, 1981).
- [31] J.-P. Eckmann, S.O. Kamphorst, D. Ruelle, “Recurrence plots of dynamical systems”, *Europhysics Letters* **4**, 973–977 (1987).
- [32] R. Devaney, Z. Nitecki, “Shift automorphism in the Hénon mapping”, *Communications in Mathematical Physics* **67**, 137–148 (1979).
- [33] O.E. Rössler, “An equation for continuous chaos”, *Physics Letters* **57A**, 397–398 (1976).
- [34] A. Lempel, J. Ziv, “On the complexity of finite sequences”, *IEEE Transactions on Information Theory* **22**, 75–81 (1976).
- [35] J. Ziv, A. Lempel, “Compression of individual sequences by variable rate coding”, *IEEE Transactions on Information Theory* **24**, 530–536 (1978).
- [36] R.J. Solomonoff, “A formal theory of inductive inference, part II”, *Information and Control* **7**, 224–254 (1964).
- [37] W.A. Brock, W. Dechert, B. LeBaron, J.A. Sheinkman, “A test for independence based upon the correlation dimension”, *Econometric Reviews* **15**, 197–235 (1996).
- [38] T. Schreiber, A. Schmitz, “Improved surrogate data for nonlinearity tests”, *Physical Review Letters* **77**, 635–638 (1996).
- [39] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, “Testing for nonlinearity in time series: the method of surrogate data”, *Physica* **58D**, 77–94 (1992).
- [40] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time series analysis: forecasting and control* (Englewood Cliffs, NJ: Prentice Hall, 1994).
- [41] R.S. Pindyck, D.L. Rubinfeld, *Econometric models & econometric forecasts* (New York, NY: MacGraw-Hill, 1991).

- [42] W.A. Fuller, *Introduction to statistical time series* (New York, NY: John Wiley & Sons, 1996).
- [43] M. Hénon, “A two dimensional mapping with a strange attractor”, *Communications in Mathematical Physics* **50**, 69–77 (1976).
- [44] M. Denker, G. Keller, “Rigorous statistical procedures for data from dynamical systems”, *Journal of Statistical Physics* **44**, 67–93 (1986).
- [45] B. LeBaron, “A fast algorithm for the BDS statistic”, *Studies in Nonlinear Dynamics and Econometrics* **2**, 53–59 (1997).
- [46] T. Schreiber, A. Schmitz, “Discrimination power of measures for nonlinearity in a time series”, *Physical Review* **55E**, 5443–5447 (1997).
- [47] I. Daubechies, *Ten lectures on wavelets* (Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992).
- [48] P. Kumar, E. Foufoula-Georgiou, “Wavelet analysis for geophysical applications”, *Reviews of Geophysics* **35**, 385–412 (1997).
- [49] D.B. Percival, A.T. Walden, *Wavelet methods for time series analysis* (Cambridge, UK: Cambridge University Press, 2000).
- [50] I. Daubechies, “Orthonormal bases of compactly supported wavelets”, *Communications on Pure and Applied Mathematics* **41**, 909–996 (1988).
- [51] G. Strang, “Wavelets and Dilation Equations: a Brief Introduction”, *Society for Industrial and Applied Mathematics Review* **32**, 614–627 (1989).
- [52] S. Mallat, “Multiresolution approximations and wavelet orthonormal bases of $\mathcal{L}^2(\mathbb{R})$ ”, *Transactions of the American Mathematical Society* **315**, 69–87 (1989).
- [53] P.A. Morettin, *Ondas e ondaletas: da análise de Fourier à análise de ondaletas* (São Paulo, SP: Editora da Universidade de São Paulo, 1999).
- [54] D.L. Donoho, I.M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage”, *Biometrika* **81**, 425–455 (1994).
- [55] J.A. Freeman, D.M. Skapura, *Neural network: algorithms, applications and programming techniques* (Reading, MA: Addison-Wesley, 1991).

- [56] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, *Biologia molecular da célula* (Porto Alegre, RS: Artes Médicas Sul, 1997).
- [57] W.S. McCulloch, W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, *Bulletin of Mathematical Biophysics* **5**, 115–133 (1943).
- [58] S. Haykin, *Redes neurais: princípios e prática* (Porto Alegre, RS: Bookman Companhia Editora, 2001).
- [59] D.O. Hebb, *The organization of behavior: a neuropsychological theory* (New York, NY: John Wiley & Sons, 1949).
- [60] F. Rosenblatt, “The perceptron: a probabilistic model for information storage and organization in the brain”, *Psychological Review* **65**, 386–408 (1958).
- [61] M.L. Minsky, S.A. Papert, *Perceptrons* (Cambridge, MA: MIT Press, 1969).
- [62] D.E. Rumelhart, G.E. Hinton, R.J. Williams, “Learning representations of backpropagation errors”, *Nature* **323**, 533–536 (1986).
- [63] D.H. Ackley, G.E. Hinton, T.J. Sejnowski, “A learning algorithm for Boltzmann machines”, *Cognitive Science* **9**, 147–169 (1985).
- [64] C.M. Bishop, *Neural network for pattern recognition* (Oxford, UK: Oxford University Press, 1995).
- [65] Y. LeCun, I. Kanter, S.A. Solla, “Second order properties of error surfaces: learning time and generalization”, In *Advances in neural information processing systems*, 918–924, ed. por R. Lippmann, J. Moody, D. Touretzky (San Mateo, CA: Morgan Kaufmann, 1991).
- [66] A.C. Valle, *Análise eletrofisiológica do hipocampo durante o sono dessincronizado do rato* (São Paulo, SP: tese de doutorado, Escola Paulista de Medicina, Universidade Federal de São Paulo, 1995).
- [67] F.F. Ferreira, G. Francisco, B.S. Machado, P. Muruganandam, “Time series analysis for minority game simulations of financial markets”, *Physica* **321A**, 619–632 (2003).
- [68] H.D.I. Abarbanel, *Analysis of observed chaotic data* (New York, NY: Springer-Verlag, 1996).
- [69] N. Gershenfeld, B. Schoner, E. Metois, “Cluster-weighted modelling for time series analysis”, *Nature* **397**, 329–332 (1999).

- [70] N. Gershenfeld, *The nature of mathematical modelling* (Cambridge, UK: Cambridge University Press, 2000).