

UNIVERSIDADE ESTADUAL PAULISTA
"JULIO DE MESQUITA FILHO"
INSTITUTO DE FÍSICA TEÓRICA

MÁRCIO DE MENEZES

REAMOSTRAGEM EM SISTEMAS DINÂMICOS &
ANÁLISE DE REDES DE MAPAS ACOPLADOS

São Paulo
2008



Instituto de Física Teórica
Universidade Estadual Paulista

TESE DE DOUTORAMENTO

IFT-T.010/08

Reamostragem em Sistemas Dinâmicos & Análise de Redes de Mapas Acoplados

Márcio de Menezes

Orientador

Gerson Francisco

Dezembro de 2008



Agradecimentos

Primeiramente quero agradecer ao grande amigo e orientador, o professor Gerson Francisco. Durante todas as etapas, principalmente nas mais difíceis, foi quem sempre me estimulou e me encorajou a prosseguir no trabalho.

Agradeço à minha esposa Keyla e ao meu filho Yuri. A minha ausência, principalmente na etapa final deste trabalho, tornou-se constante. Mesmo assim o apoio dos dois sempre esteve presente.

Agradeço aos amigos que auxiliaram em muitas discussões e ajudaram a eliminar diversas dúvidas. Agradeço principalmente ao amigo Birajara Soares Machado que muito ajudou quando adentrei nesta área.

Agradeço enfim à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) pelo suporte financeiro.

Resumo

Este trabalho trata de dois temas principais: a reamostragem de séries temporais caóticas e a análise de redes de mapas acoplados.

A reamostragem de séries temporais é estudada com o objetivo de encontrar uma incerteza para os invariantes medidos de um sistema dinâmico. Quando um invariante, tal como o expoente de Lyapunov, é obtido a partir de uma série temporal, freqüentemente este valor é calculado sem que seja associada uma medida de incerteza. Isto pode causar problemas, às vezes inviabilizando determinar se um sistema é realmente caótico.

No processo de reamostragem outras séries temporais, que apresentam as mesmas propriedades dinâmicas da série original, são criadas. O processo de reamostragem é baseado nos métodos de previsão de uma série temporal.

Depois que várias séries temporais são obtidas, cada uma delas é utilizada para medir um invariante do sistema, no caso o expoente de Lyapunov. Cada uma das séries apresenta um valor diferente para este expoente, assim obtém-se uma distribuição de valores para tal parâmetro. Com esta distribuição é possível calcular várias estatísticas, como o desvio padrão e alguns percentis para a distribuição.

Nesta tese também é realizado um estudo sobre redes de mapas acoplados. Foram analisadas redes com dimensões: um, dois e três. Para cada um destes casos foram analisadas as propriedades estatísticas, assim como as propriedades dinâmicas.

A partir destas análises, é mostrado que as séries temporais destes mapas apresentam auto-similaridade. Além disso, foi possível verificar que, com o aumento da dimensão, a série temporal torna-se mais correlacionada.

Palavras Chaves: caos; série temporal; reamostragem; modelagem local; rede; mapas acoplados.

Áreas do conhecimento: 1.01.03.04-0; 1.01.04.00-3; 1.05.01.00-2.

Abstract

This work deals with the study of the properties of the system of differential equations of the type

The study of the properties of the system of differential equations of the type (1) is carried out in the framework of the theory of differential equations. It is shown that the system (1) is solvable in the case when the matrix A is constant and the vector b is a function of the independent variable x . It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a constant.

It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a function of the independent variable x . It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a constant.

It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a function of the independent variable x . It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a constant.

It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a function of the independent variable x . It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a constant.

It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a function of the independent variable x . It is also shown that the system (1) is solvable in the case when the matrix A is a function of the independent variable x and the vector b is a constant.

Abstract

This work deals with two main themes: the resampling of time series and the analysis of coupled maps.

The purpose of studying time series resampling is to find uncertainty intervals for the invariants of dynamical systems. When an invariant, such as the Lyapunov exponent, is obtained from the time series, it is frequently the case that no uncertainty interval is computed. This may cause problems, even making it unfeasible to determine if a system is really chaotic.

In the resampling process several time series are created, that share the same dynamical properties of the original series. This resampling process is based on methods of prediction of a time series.

After several time series are obtained, each one is used to measure an invariant of the system, in this case the Lyapunov exponent. Each series presents a distinct value for this exponent, and thus a distribution of values is obtained the parameter. With this distribution it is possible to calculate several statistics, such as the standard deviation and for the distribution.

This thesis also examines a study about coupled maps. It was implemented grids with dimensions 1, 2 and 3. For each case the statistical, as well as dynamical properties were analyzed.

From these analyses it is shown that the time series of these maps show self-similarity. In addition, it was possible to verify that, the series becomes more correlates as the dimension is increased.

1.1	Introduction	17
1.2	Structure of the work	18
1.3	Conclusions	19
2.1	Time series resampling	20
2.2	Time series analysis	21
2.3	Time series prediction	22
2.4	Time series resampling and prediction	23

Índice

Agradecimentos	iii
Resumo	v
Abstract	vii
Introdução	xiii
Reamostragem de séries temporais	xiii
Rede de mapas acoplados	xiv
1 Séries Temporais Não-Lineares	1
1.1 Imersão da série escalar	3
1.2 Estacionariedade	9
2 Teoria da Imersão e do Mergulho	11
2.1 Introdução	11
2.2 Espaço de Fase	11
2.2.1 Campo vetorial	12
2.2.2 A trajetória	13
2.2.3 O atrator	13
2.3 Equivalência entre Espaços de Fase	14
2.3.1 Morfismo	14
2.3.2 Morfismo entre espaços de fase	15
2.3.3 Estruturas preservadas	17
2.4 Mergulho de uma série escalar	18
2.4.1 Coordenadas de derivadas	19
2.4.2 Coordenadas de atraso	20
2.4.3 Dimensão de Imersão	21
2.4.4 Imersão e mergulho	22
2.5 Redução de ruído com imersão filtrada	23

2.5.1	Transformada de Fourier	24
2.5.2	Filtragem e Interpolação por Fourier	25
3	Modelagem Local	33
3.1	Aprendizado de Máquina	33
3.2	Modelagem global $\times$ modelagem local	34
3.3	Modelagem local	37
3.3.1	As modelagens local e global em competições	37
3.3.2	Dados em tempo real	38
3.3.3	Arquitetura do modelo local	38
3.3.4	Escolha dos parâmetros	39
3.4	Modelagem local no espaço de fase	40
4	Arquitetura do modelo local	41
4.1	Exemplos relevantes	42
4.2	Seleção de uma Métrica para a Distância	44
4.3	A construção do modelo	44
4.3.1	Aproximação de grau zero	46
4.3.2	Aproximação linear	47
4.4	Tipos de Modelos Locais	48
4.4.1	Modelos Lineares Locais	49
4.4.2	Modelos de Média Local	50
4.4.3	Modelos Quadráticos Locais	51
4.4.4	Média Local versus Linear Local	51
4.5	Ponderação local para modelos mais suaves	53
4.5.1	Média Local Ponderada	53
4.5.2	Modelos Lineares Locais Ponderados	54
5	Regularização	57
5.1	Introdução	57
5.2	Decomposição em Valor Singular	59
5.2.1	Conceitos de Álgebra Linear	60
5.3	Regressão em Componentes Principais	61
5.4	Ponderação	62
5.4.1	Reduzindo a variância do modelo	63
5.5	Interpretação Qualitativa	64
5.6	Regularização Aplicada a Sistemas Dinâmicos	65
5.7	A escolha da dimensão do espaço da projeção	66
6	Reamostragem	69

ÍNDICE

xi

6.1	Introdução	69
6.1.1	Reamostragem em séries temporais	71
6.2	Sombreamento	73
6.3	A metodologia desenvolvida	75
6.4	Aplicação	78
7	Expoente de Lyapunov	81
7.1	Introdução	81
7.2	Espectro de Lyapunov	82
7.3	Cálculo computacional	83
7.4	O expoente para a série e as reamostragens	85
8	Redes de Mapas Acoplados	89
8.1	Introdução	89
8.2	Rede de mapas acoplados	90
8.2.1	Rede unidimensional	90
8.2.2	Rede bidimensional	91
8.2.3	Rede tridimensional	91
8.2.4	A escolha dos parâmetros e do mapa	92
8.3	Mapa logístico	92
8.4	Séries temporais	93
9	Análise de redes de mapas acoplados	97
9.1	Análise estatística	98
9.1.1	Análise de flutuação sem tendência	98
9.1.2	Densidade espectral de potência	101
9.2	Análise Dinâmica	102
9.2.1	Dimensão Local	102
9.2.2	Dimensão de imersão	104
9.2.3	Estacionaridade e análise de recorrência	105
10	Conclusões	109
10.1	Reamostragem	109
10.1.1	Trabalho futuro	110
10.2	Redes de mapas acoplados	110
10.2.1	Trabalho futuro	110
	Índice Remissivo	113
	Referências	115

Introdução

Este trabalho é dividido em duas partes principais. A primeira parte compreende os capítulos 1 a 7, tratando da reamostragem de séries temporais caóticas. A segunda parte compreende os capítulos 8 e 9, tratando da análise de redes de mapas acoplados.

Reamostragem de séries temporais

Na primeira parte é desenvolvido um algoritmo que permite realizar a reamostragem de séries temporais com baixa dimensão, situação típica em sistemas caóticos. O objetivo das reamostragens é obter uma incerteza para os valores frequentemente calculados a partir de uma série temporal.

Existem estimadores para se calcular vários invariantes que caracterizam sistemas caóticos, entretanto não existe nenhum estimador para o cálculo da incerteza destes valores. Assim a melhor forma de se estimar a incerteza destes valores é através da reamostragem da série temporal.

No capítulo 1 é feita uma breve introdução às séries temporais não-lineares originárias de sistemas determinísticos. No capítulo 2 é apresentada a teoria necessária para tratar destas séries temporais. É mostrado como uma série temporal escalar pode ser imersa num espaço vetorial, chamado de espaço de fase. Além disso, ao final do capítulo 2 é mostrado como pode ser realizada uma imersão filtrada, que é capaz de filtrar os dados à medida que os insere no espaço de fase.

No capítulo 3 discutem-se os modelos locais. Estes modelos são comparados aos modelos globais, mostrando que os modelos locais têm proporcionado melhores resultados. No capítulo 4 é apresentada a estrutura do modelo local. É mostrado que o modelo linear local é uma boa escolha para produzir a evolução dos pontos no espaço de fase. Entretanto, para que o modelo possa produzir bons resultados, no capítulo 5 é introduzido um método de

regularização, que auxilia no comportamento do modelo. O modelo pode apresentar uma variância muito alta quando a série temporal está contaminada por ruído, ou quando o número de exemplos no espaço de fase para produzir uma evolução é muito pequeno.

No capítulo 6 finalmente é mostrado como a modelagem dos dados pode ser utilizada para criar replicações da série temporal original. No capítulo 7 as várias séries reamostradas são utilizadas para se calcular uma distribuição para o expoente de Lyapunov da série original. Para uma comparação, são geradas várias séries temporais integrando várias vezes o mesmo conjunto de equações diferenciais, assim obtém-se uma distribuição para o expoente de Lyapunov. Finalmente as duas distribuições são comparadas (a distribuição produzida pelas várias integrações numéricas e a distribuição devido às reamostragens).

Rede de mapas acoplados

No capítulo 8 desta tese é mostrado o que é uma rede de mapas acoplados e são expostas as redes de dimensão 1, 2 e 3. Ainda neste capítulo, é mostrado como ocorre a interação entre os vizinhos mais próximos numa rede e são escolhidos os parâmetros para estas redes.

No capítulo 9 é feito um estudo estatístico e dinâmico sobre as redes de mapas acoplados. Através desta análise é mostrado como a mudança na dimensão altera as características que são mensuradas.

Finalmente, no capítulo 10 são apresentadas as conclusões para as duas partes do trabalho, ou seja, para a reamostragem de séries temporais e para as redes de mapas acoplados.

Capítulo 1

Séries Temporais Não-Lineares

Uma série temporal é uma sequência de medidas de algum observável ordenada em relação ao tempo. De um modo geral estas medidas feitas são escalares, ou seja, apenas um valor numérico que se altera ao longo do tempo.

O estudo das séries temporais é realizado em diversas áreas do conhecimento, tais como: Meteorologia, Geofísica, Ecologia, Acústica, Economia, etc [1]. Independente da área do conhecimento, este estudo consiste de *análise*, *classificação* e *previsão*. Cada uma destas etapas é importante e elas estão bastante relacionadas entre si.

A primeira etapa no estudo de uma série temporal é a sua *análise*, na qual se busca conhecer as características da série temporal. Nesta etapa, pode-se estar interessado em saber se a série temporal é oriunda de um processo estocástico ou determinístico. Pode-se verificar ainda se a série temporal é linear ou não-linear. Dentre outras verificações que podem e devem ser feitas para se ter um bom conhecimento da série temporal.

Depois de se conhecer as características da série temporal, busca-se fazer a *classificação* da série temporal. Neste ponto, várias séries temporais são agrupadas de acordo com as suas características.

A terceira etapa, que costuma estar presente no estudo das séries temporais, é a *previsão*. Para que seja feita a previsão de uma série temporal, ela deve ter sido primeiramente classificada. É importante observar que existem várias metodologias utilizadas para se fazer previsões baseadas numa série de dados passados. As diferentes metodologias são utilizadas de acordo com as características do conjunto de dados.

Mesmo que o objetivo central seja o de previsão, a classificação é sempre necessária, para que se utilize a metodologia apropriada àquela série tempo-

ral. Observa-se assim que a classificação é extremamente importante para que se façam predições a partir do comportamento passado desta série.

Este trabalho concentra-se na análise e classificação de séries temporais. Apresentando uma forma de classificação com maior acurácia. A previsão das séries não será realizada.

Neste trabalho são tratadas apenas séries temporais não-lineares determinísticas. A caracterização das séries lineares já está bem definida e não será realizada aqui. As séries temporais não-lineares não são tão bem definidas assim. Muitos trabalhos (a maioria deles) apresentam valores obtidos para o expoente de Lyapunov, ou para a dimensão de correlação, a partir de um conjunto de dados, que não apresentam a incerteza associada a estes valores [2, 3, 4].

Vale ressaltar que é muito importante ter algum conhecimento da incerteza associada a qualquer medida. O expoente de Lyapunov, por exemplo, é utilizado para se ter conhecimento da sensibilidade às condições iniciais de um sistema. Assim, não basta medir um expoente de Lyapunov positivo para afirmar que existe este tipo de sensibilidade, é necessário saber que o valor medido é estatisticamente diferente de zero. Este conhecimento somente é possível através da incerteza associada a estes valores.

No estudo de séries temporais não-lineares determinísticas, não existe uma teoria bem definida que apresente uma forma prática de obter a incerteza nos valores estatísticos obtidos de uma amostra. Existem, entretanto, trabalhos que propõem alguns métodos de cálculo do desvio-padrão de parâmetros não-lineares [5, 6, 7, 8].

Uma proposta para que isto seja contornado é através da reamostragem da série temporal original. A reamostragem de uma série temporal nada mais é do que uma outra realização desta série, ou seja, outra possível série gerada pelo mesmo sistema. A forma de se realizar a reamostragem de uma série temporal está intimamente relacionada à previsão de tal série. A metodologia, como ver-se-á a seguir, é muito semelhante à da previsão.

De forma bastante resumida, este trabalho está centrado na parte de análise e classificação das séries temporais. Pois aqui é apresentada uma metodologia para se obter uma incerteza nos valores estatísticos calculados. Com este tipo de análise, que considera também a incerteza, pretende-se realizar uma melhor classificação das séries temporais.

Nesta tese, também é feito um estudo sobre o comportamento de mapeamentos acoplados, utilizando um extenso conjunto de ferramentas para análise de séries temporais. No estudo destas redes de mapas, faz-se uma

análise através de ferramentas estatísticas e através de análises baseadas em sistemas determinísticos não-lineares.

1.1 Imersão da série escalar

Muitas vezes, o sistema dinâmico que originou a série não é completamente entendido, ou seja, não se sabe quais foram as equações que deram origem a uma dada série temporal. Neste caso, as regularidades presentes na própria série podem prover alguma compreensão a respeito do sistema.

Entretanto, para uma boa compreensão da dinâmica através das regularidades da série, esta deve ser composta por dados em grande quantidade e qualidade. Isto torna o tratamento de séries temporais não-lineares é bastante difícil. Bases de dados na ordem de pelo menos 10^3 elementos são usadas normalmente. Além disso, grande parte dos algoritmos de análise não-lineares não podem tolerar uma amplitude de ruído de mais de 2% da amplitude do sinal [9]. Mas sempre existe a possibilidade de filtrar a série de dados antes de se fazer a análise.

As informações sobre o sistema estão presentes na série temporal, mas extraí-las pode ser uma tarefa difícil. As medições escalares são apenas uma projeção (muitas vezes não-linear) das verdadeiras variáveis de estado (não-observadas) de algum sistema. A fim de analisar um sistema baseado unicamente em dados da série temporal, o espaço de estados tem que ser reconstruído. Enquanto o espaço de estados original não pode ser unicamente determinado, um espaço de estados equivalente (no sentido que os dois estão relacionados por um mapeamento suave e inversível) pode ser construído usando o método de *coordenadas de atraso temporal*, onde se constrói vetores a partir de uma série temporal escalar, tomando-se instantes de tempo igualmente espaçados.

Portanto estes vetores reconstruídos serão capazes de reproduzir a dinâmica do sistema em questão. O procedimento para se obter estes vetores é apresentado. Seja uma série temporal com N medições escalares denotadas por $x_1, x_2, \dots, x_N$. Cada *vetor de atraso* desta série temporal é dado por

$$Y_n = (x_n, x_{n+\tau}, \dots, x_{n+(m-2)\tau}, x_{n+(m-1)\tau}),$$

onde τ é um número natural que representa o intervalo de tempo entre duas coordenadas consecutivas. O intervalo de tempo τ entre duas coordenadas pode ser tomada igual a um, mas este valor pode ser variado no intuito de se

obter melhores resultados na análise da série temporal. A dimensão do vetor reconstruído é comumente simbolizada por m e é a dimensão do espaço de fase reconstruído. O índice n varia de 1 a $N - (m - 1)\tau$. Assim, a partir de uma série temporal de tamanho N , são construídos $N - (m - 1)\tau$ vetores.

O vetor y_n é m -dimensional, formado por elementos da série temporal original. O atraso de tempo τ faz com que o vetor y_n seja composto de elementos que não eram necessariamente consecutivos na série temporal. Assim o vetor y_n é composto por elementos que estão a uma distância τ um do outro. Uma maneira olhar o atraso temporal é que ele aumenta a janela de tempo coberta por um vetor, isto faz com que as componentes do vetor estejam menos correlacionadas entre si.

Para demonstrar a reconstrução de um espaço de fase, é utilizada a série temporal de Lorenz [17]. A série foi construída usando as equações propostas por Edward Lorenz para a dinâmica da convecção hidrodinâmica. As equações são:

$$\begin{aligned}\frac{dx}{dt} &= \sigma(y - x) \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= xy - bz\end{aligned}$$

Para esta evolução, os valores dos parâmetros utilizados foram: $\sigma = 10$, $r = 28$ e $b = 8/3$. A unidade de tempo foi considerada igual a 1 segundo.

É importante notar que a taxa de amostragem para a série temporal utilizada foi de 50 Hz, ou seja, um ponto a cada 0,02 segundos.

Na figura 6.1 observa-se a evolução da coordenada x ao longo do tempo. Muitas vezes, quando realiza-se o estudo de uma série temporal, a única informação que se dispõe é uma série escalar como esta. Tal série será utilizada para construir um espaço de fase equivalente ao espaço de fase original do sistema.

O espaço de fase original do sistema proposto por Lorenz pode ser construído a partir dos pontos $(x, y \text{ e } z)$ que são obtidos da evolução das equações diferenciais. Para imprimir este espaço de fase no papel, é necessário fazer uma projeção em duas dimensões. Portanto, pode se tomar simplesmente duas coordenadas do sistema original, por exemplo, as coordenadas x e y , e construir uma projeção do espaço de fase original, assim como é mostrado na figura 1.2. Para facilitar a visualização da dinâmica no espaço de fase,

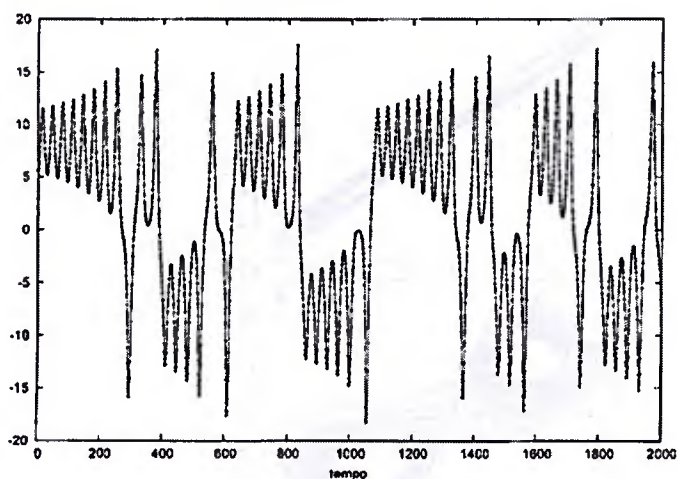


Figura 1.1: Série temporal de Lorenz. A série foi construída usando as equações propostas por Edward Lorenz para a dinâmica da convecção hidrodinâmica.

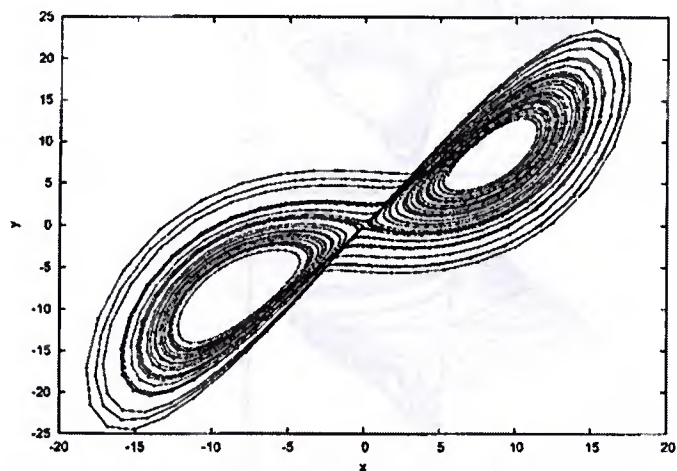


Figura 1.2: Projeção bidimensional do atrator de Lorenz.

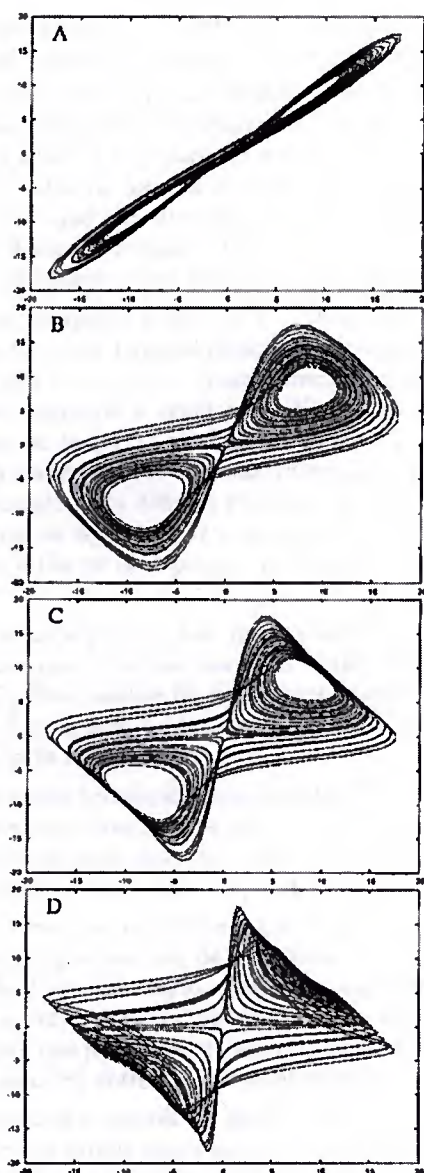


Figura 1.3: O atrator de Lorenz reconstruído para diferentes valores do *atraso temporal*: A) atraso temporal igual a 1; B) atraso temporal igual a 7; C) atraso temporal igual a 10; D) atraso temporal igual a 15. Mais detalhes sobre os diferentes resultados obtidos nas figuras são apresentados no texto.

os pontos temporalmente consecutivos são unidos por uma linha reta. Uma curva suave seria mais indicada para representar a evolução do sistema.

A figura 1.3 mostra o atrator de Lorenz reconstruído em duas dimensões para vários valores do atraso temporal. Comparando a figura 1.3 (espaço de fase reconstruído) com a figura 1.2 (espaço de fase original), observa-se que existe uma semelhança entre os espaços de fase. Esta semelhança é tal que existe um mapeamento que leva um espaço de fase no outro. Este mapeamento simplesmente deforma o espaço de fase, sem alterar as suas propriedades topológicas e dinâmicas, como veremos mais à frente.

Na figura 1.3.A o atraso temporal é igual a 1. Neste caso o atrator encontra-se concentrado em torno da diagonal principal. Isto significa que as duas variáveis (x_i e x_{i+1} - para $0 < i < N - 1$) estão muito correlacionadas. Já na figura 1.3.B o atraso temporal é igual a 7. Neste caso é possível visualizar bem as trajetórias ao longo do atrator. Na figura 1.3.C o atraso temporal é igual a 10. Agora observa-se que algumas trajetórias começam a se concentrar e se sobrepor, ficando mais difícil a identificação das trajetórias individualmente. Finalmente na figura 1.3.D o atraso temporal é igual a 15. Aqui várias trajetórias estão se sobrepondo, dificultando ainda mais a visualização das trajetórias.

De todos os os diagramas de espaço de fase reconstruídos que são apresentados na figura 1.3, vemos que o melhor resultado obtido foi utilizando o atraso temporal igual a 7. Esta análise foi feita simplesmente utilizando uma inspeção visual. Esta é uma metodologia bastante usada para que se descubra um valor razoável para o atraso temporal.

O valor adequado para o atraso temporal depende da taxa de amostragem da série temporal. Uma série com uma taxa de amostragem elevada exigirá um atraso temporal elevado. Já uma série com uma taxa de amostragem baixa requererá um atraso temporal pequeno.

Os parâmetros m e τ devem ser escolhidos para cada série temporal. Infelizmente, não há nenhum algoritmo que dê os valores ótimos para m e τ para qualquer conjunto de dados arbitrário. Diferentes algoritmos obtêm diferentes valores para cada um destes parâmetros. Assim, o pesquisador tem apenas indícios de valores que poderiam ser adequados para a dimensão (m) dos vetores e para o atraso (τ) entre as suas componentes.

Para a escolha da dimensão dos vetores do espaço de fase reconstruído, deve-se observar que um sistema evolui pelo espaço de fase ao longo de uma trajetória suave. Para descrever esta trajetória, é necessário que se reconstrua o espaço de fase com dimensão adequada. A dimensão utilizada para reconstruir o espaço de fase através dos vetores de atraso é chamada de *dimensão de*

imersão (m). Se a dimensão for inferior à mínima necessária (m_0), ocorrerá uma projeção do espaço de fase num espaço de dimensão menor. Desta forma, intersecções entre diferentes trajetórias podem ocorrer. Há, portanto, uma diferença qualitativa entre os valores de m que são demasiado pequenos, e os valores que são suficientemente grandes. Se o valor da dimensão m for menor do que a dimensão adequada (m_0), então haverá variáveis de estado que deveriam estar em diferentes partes do espaço de fase que estarão sobrepostas, criando falsos vizinhos. Entretanto, para $m > m_0$, estes falsos vizinhos não existem. Sendo importante notar que existe apenas um limite inferior para a escolha da dimensão, mas não um limite superior. Assim, procurar por falsos vizinhos é uma boa técnica para encontrar este limite inferior [22].

Não há nenhum resultado semelhante para o atraso temporal τ , mas podemos determinar um intervalo de valores para este parâmetro. Abaixo apresentaremos duas metodologias. De um ponto de vista prático, uma escolha apropriada de τ ajuda na análise não-linear, enquanto uma escolha ruim prejudica [23]. Qualitativamente, os valores pequenos de τ tornam os sucessivos vetores de atraso mais e mais correlacionados, de modo que o espaço de fase do sistema esteja concentrado ao longo da diagonal. Os valores grandes de τ tornam os vetores de atraso virtualmente não-correlacionados, de modo que preencham uma nuvem dispersa no espaço de fase. A melhor estratégia é selecionar uma escala promissora para τ , e então inspecionar visualmente cada espaço de fase reconstruído. Como o ruído na série temporal proíbe o acesso às escalas infinitesimais de comprimento, as maiores estruturas determinísticas possíveis são desejadas. Indicações sobre os valores promissores de τ podem ser obtidas usando o ponto onde a *função de autocorrelação* atinge o valor $1/e$, ou o primeiro mínimo da *informação mútua* [21]. É comum obter valores diferentes para o atraso temporal usando estas duas metodologias. Assim, combinando os dois algoritmos, obtém-se um intervalo para o atraso temporal (τ).

O retrato de fase de um sistema fornece a informação qualitativa sobre um sistema. Um processo verdadeiramente estocástico preencherá uma nuvem no espaço de fase; os sistemas determinísticos com nenhum ruído mostrarão trajetórias suaves. O aparecimento de estruturas no retrato de fase, ou simplesmente furos (espaços vazios), sinaliza a possível presença de uma dinâmica determinística de algum tipo [23]. Uma inspeção do retrato de fase de um sistema é, portanto, uma primeira etapa crítica na análise desse sistema.

A utilização de coordenadas de atraso para reconstruir o espaço de fase surgiu em torno de 1980. O artigo de Packard et al. [18] atribui a idéia a Ruelle. Mas foi Takens [19] quem provou a validade das coordenadas de

atraso. Takens mostrou que, para reconstruir uma variedade compacta e suave de dimensão d , o espaço de fase reconstruído deverá ter dimensão $2d + 1$. Conforme pode ser observado na reconstrução de espaços de fase, este é um limite superior, muitas vezes é possível reconstruir espaço de fase com um número menor de coordenadas.

A prova original feita por Takens aplicava-se a variáveis suaves. Posteriormente Sauer et al. [20] mostra que um atrator fractal com dimensão de correlação d pode ser reconstruído através de vetores m -dimensionais, desde que m seja maior que $2d$. É importante observar que a dimensão d , neste caso, não é necessariamente um número inteiro. No caso de d ser um número natural, temos o mesmo resultado proposto por Takens.

1.2 Estacionariedade

Todos os algoritmos para a análise de série temporal não-linear supõem que a série temporal é originária de um processo estacionário, em uma taxa suficientemente elevada de amostragem [36]. Se esta suposição for violada, então os algoritmos produzirão resultados espúrios. Assim, é preciso uma verificação para a estacionariedade.

No domínio das séries temporais não-lineares, estatísticas lineares tais como correlações não são usadas para testar a não-estacionariedade, porque não podem detectar relações não-lineares. Em um número de sistemas caóticos, as variações dos parâmetros podem não resultar em nenhuma mudança das estatísticas lineares; somente as relações não-lineares são afetadas.

A técnica comumente usada chama-se *gráfico da separação espaço-temporal* [37], a qual relaciona a diferença temporal entre pontos e sua distância espacial. O gráfico da separação espaço-temporal traça curvas de probabilidade constante para que dois pontos estejam dentro de uma distância espacial ϵ , supondo que a diferença temporal é T . Se estas curvas saturarem em um platô (ou em uma oscilação estável), a série temporal é então originada de um processo estacionário em uma taxa adequada de amostragem. Uma outra maneira de olhar o gráfico da separação espaço-temporal é que ele mostra se a escala de tempo das observações feitas em um sistema é suficientemente maior do que a escala de tempo própria do sistema.

Capítulo 2

Teoria da Imersão e do Mergulho

2.1 Introdução

Foi mostrado no capítulo 1 que um espaço de fase (visualmente semelhante ao original) pode ser construído utilizando-se os vetores de atraso. Agora, é necessário conhecer a equivalência existente entre estes dois espaços de fase.

Para um conhecimento das propriedades preservadas neste novo espaço de fase, serão apresentadas primeiramente as características de um espaço de fase. Logo em seguida, será mostrado que um espaço de fase numa variedade pode ser mergulhado em $\mathbb{R}^n$. O mergulho, no caso em questão, é um mapeamento (do espaço de fase original no espaço de fase reconstruído) que preserva algum tipo de estrutura presente no espaço de fase original. Dependendo da estrutura preservada por este mapeamento, ele pode ser chamado de um *mergulho* ou de uma *imersão*.

2.2 Espaço de Fase

Um sistema dinâmico é dado por:

$$\dot{x} = F(x), \quad (2.1)$$

onde $x \in \mathbb{R}^d$ é o estado do sistema e pode ser pensado como um ponto num espaço chamado de *espaço de fase* S . A dimensão de S está associada com o número de *graus de liberdade* do sistema, e é igual a d [10].

Quando não se conhece o espaço de fase original do sistema, e apenas uma série temporal está disponível, pode-se construir um espaço de fase que tenha propriedades semelhantes ao espaço de fase original. Foi mostrado no capítulo 1, de forma empírica, que, através das coordenadas de atraso, pode-se reconstruir um espaço de fase que apresenta alguma semelhança visual com o espaço de fase original. Agora serão mostradas as propriedades matemáticas que são preservadas neste novo espaço de fase.

Neste novo espaço de fase, o estado do sistema será dado pelo vetor $y \in \mathbb{R}^m$. A evolução dos pontos neste novo espaço de fase será dada pela equação:

$$\dot{y} = G(y).$$

Para que seja possível comparar os dois espaços de fase (o do sistema original e o do espaço de fase reconstruído), será necessária uma relação de equivalência entre equações diferenciais [10].

É importante notar que, de uma forma geral, a partir de dados empíricos, o espaço de fase original do sistema é desconhecido. Apenas o espaço de fase reconstruído a partir das coordenadas de atraso é conhecido. Assim, a partir do momento que se consegue estabelecer relações de equivalência entre estes dois espaços de fase, ganha-se um poder preditivo sobre o sistema. Isto ocorre porque o conhecimento sobre o *campo vetorial* $G(\cdot)$ leva a um conhecimento sobre o campo vetorial $F(\cdot)$.

2.2.1 Campo vetorial

Um *campo vetorial* é uma função que associa, a cada ponto do espaço, um vetor, cujas componentes variam, de ponto a ponto, de maneira contínua e diferenciável [11]. Este vetor é chamado de *vetor tangente*.

No caso de interesse, o *campo vetorial* é uma função não-linear diferenciável agindo sobre os pontos do espaço de fase, determinando sua evolução. Como o campo vetorial é diferenciável, ou seja, suave, possui derivada de primeira ordem contínua, podendo ter continuidade nas derivadas de ordem superior.

Um campo vetorial, assim como $F(\cdot)$ ou $G(\cdot)$, é de classe C^1 quando possui derivadas de primeira ordem contínuas. Caso possua derivadas contínuas até ordem r ($r > 1$), então será chamada de uma função de classe C^r .

2.2.2 A trajetória

A equação (2.1) define um *problema de valor inicial*, no sentido de que uma única solução (curva, trajetória, órbita) passa por cada ponto x no espaço de fase. Assim a solução desta equação é dada por $x(t) = \phi_t x_0$, onde x_0 é alguma condição inicial. Aqui ϕ_t representa uma classe de mapeamentos do espaço de fase nele mesmo [10, 11].

A trajetória $\phi_t x_0$ é também chamada de *fluxo* da equação diferencial, onde ϕ_t é o mapeamento do fluxo no instante t . Observa-se que, para cada condição inicial, existe um diferente fluxo, ou seja, uma diferente trajetória. O conjunto de soluções para todos os possíveis valores iniciais pode ser escrito como $\phi_t S$.

2.2.3 O atrator

Em princípio, a dimensão do conjunto $\phi_t S$ poderia ser a mesma que a do próprio espaço S . Entretanto, conforme o sistema evolui, o fluxo pode contrair (é o que ocorre num sistema dissipativo) e, conseqüentemente, a dimensão é reduzida. Essa região, na qual os pontos do espaço de fase convergem assintoticamente, é chamada de *atrator* [10, 12].

Devido à contração do fluxo no espaço de fase é que observamos na série temporal um comportamento transiente seguido de um regime assintótico. Assim o sistema evolui, apresentando um comportamento inicial, até que chegue ao atrator, a partir de onde se dá o regime assintótico. O atrator do sistema é um conjunto com uma dimensão menor que a do espaço S . Desta forma, no atrator, o sistema possui um número de graus de liberdade menor do que aqueles do espaço de fase.

O atrator é chamado de conjunto invariante, pois ele não é alterado pela evolução dinâmica [13]. Além disso, nos casos de interesse, o atrator possui uma estrutura *fractal*, ou seja, a sua *dimensão de capacidade* (ou *dimensão de contagem de caixa*) é diferente da sua dimensão topológica e dada por um número não-inteiro [12].

Segundo Eckman e Ruelle [15], um *atrator* é dito *estranho* quando apresenta *sensibilidade às condições iniciais*. Assim, a estranheza diria respeito à sua dinâmica (sensibilidade às condições iniciais) e não a sua geometria (fractal). Entretanto, segundo Ott [12], quando um atrator é fractal, ou seja, possui dimensão não-inteira, ele é chamado de *atrator estranho*.

2.3 Equivalência entre Espaços de Fase

Conforme foi dito anteriormente, para que seja feita uma comparação entre os dois espaços de fase (o do sistema original e o do espaço de fase reconstruído), é necessário que exista uma relação de equivalência entre estes espaços, assim como uma equivalência entre as equações diferenciais que descrevem as trajetórias em ambos os espaços de fase.

Depois de introduzida esta relação de equivalência, será possível utilizar a informação da dinâmica do espaço de fase reconstruído para compreender a dinâmica do espaço de fase original.

2.3.1 Morfismo

Uma equivalência entre objetos é chamada matematicamente de *morfismo* [16]. *Morfismo* é um mapeamento entre objetos que preserva algum tipo de estrutura destes objetos sobre os quais este mapeamento é aplicado. Dois tipos de morfismos são de importância para o estudo da equivalência dos espaços de fase: *homeomorfismo* e *difeomorfismo*. A seguir será mostrada a estrutura que cada um destes mapeamentos preserva.

Homeomorfismo é um mapeamento que preserva todas as propriedades topológicas de um dado espaço. Um homeomorfismo é descrito por um mapeamento H entre dois espaços topológicos X e Y , contendo as seguintes propriedades [13]:

- o mapeamento H é descrito por uma função *bijetiva*, ou seja, uma função que é simultaneamente *injetiva* e *sobrejetiva*. Sendo injetiva, cada elemento y é imagem de um único elemento do seu domínio x . Sendo sobrejetiva, não existe um elemento y que não seja imagem de pelo menos um elemento do domínio x .
- o mapeamento H é contínuo.
- o mapeamento H é inversível.
- o mapeamento inverso (H^{-1}) é contínuo.

Difeomorfismo é um mapeamento entre duas variedades diferenciáveis, tal que o mapeamento, assim como o mapeamento inverso são funções suaves. Um difeomorfismo de classe C^r é um mapeamento de classe C^r cuja inversa

existe e também é de classe C^r . Um difeomorfismo é descrito por um mapeamento H entre duas variedades X e Y , contendo as seguintes propriedades [11]:

- o mapeamento H é descrito por uma função bijetiva.
- o mapeamento H é diferenciável.
- o mapeamento H é inversível.
- o mapeamento inverso (H^{-1}) é diferenciável.

2.3.2 Morfismo entre espaços de fase

Agora que o significado de equivalência está definido, é preciso entender as propriedades que serão preservadas no espaço de fase reconstruído. Conforme já foi dito, um sistema dinâmico é dado por:

$$\dot{x} = F(x).$$

Mas quando o espaço de fase é reconstruído, obtém-se uma outra variável: y . A equação que descreve a evolução desta nova variável não será necessariamente a mesma, pois as coordenadas desta variável são uma projeção (não necessariamente linear) das coordenadas da variável antiga. Assim, o sistema dinâmico neste espaço de fase reconstruído será descrito pela equação:

$$\dot{y} = G(y).$$

É necessário estabelecer uma relação de equivalência entre estes dois campos vetoriais $F(\cdot)$ e $G(\cdot)$, para que se saiba quais são as propriedades de cada um deles. Com esta relação de equivalência pode-se realizar a análise, ou previsão, da série temporal original, conhecendo-se apenas o espaço de fase reconstruído.

Para o sistema dinâmico original, escreve-se a sua trajetória como $x(t) = \phi_t x_0$, onde ϕ_t é um mapeamento do espaço de fase nele mesmo, que descreve a evolução do sistema. Para as variáveis do espaço de fase reconstruído, pode-se escrever $y(t) = \psi_t y_0$.

É sabido que um sistema dinâmico é completamente determinado por seu estado. Assim, se um sistema encontra-se num estado x_0 , seu estado futuro (x_t) pode ser determinado. Entretanto a variável x não é conhecida, mas apenas a variável y . Como existe um difeomorfismo H que relaciona

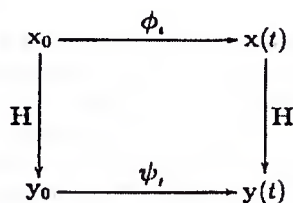


Figura 2.1: Mapeamento (H) do espaço de fase original no espaço de fase reconstruído. No espaço de fase original a variável é $x(t)$, enquanto que a evolução temporal é dada pelo mapeamento ϕ_t . No espaço de fase reconstruído a variável é $y(t)$, enquanto que a evolução temporal é dada pelo mapeamento ψ_t .

ambos os espaços de fase (o original e o reconstruído), pode-se determinar completamente o sistema dinâmico pelo estado y_0 . Conhecendo-se $y_0 = H(x_0)$, pode-se determinar $y_t = H(x_t)$.

Na figura 2.1 observa-se que a partir do mapeamento ψ_t e do difeomorfismo H é possível obter o mapeamento ϕ_t .

A equivalência entre as trajetórias (do espaço de fase original e do espaço de fase reconstruído) pode ser de dois tipos. Uma delas é chamada de *equivalência topológica*, enquanto que a outra é chamada de *equivalência diferenciável*. Para que exista qualquer tipo de equivalência entre os dois fluxos, é necessário que haja um mapeamento que leve uma trajetória na outra. De acordo com as propriedades deste mapeamento, podemos classificar o tipo de equivalência entre os fluxos.

A equivalência topológica entre os espaços de fase (original e reconstruído) é um homeomorfismo. Esta equivalência também é chamada de equivalência C^0 . Ela ocorre quando o mapeamento entre estes espaços é dado por uma função de classe C^0 .

Já a equivalência diferenciável entre os espaços de fase é um difeomorfismo. Esta equivalência também é chamada de equivalência C^r ($r \geq 1$), e ocorre quando o mapeamento entre estes espaços é dado por uma função de classe C^r .

Portanto, dois *campos vetoriais* de classe C^k , $F(\cdot)$ e $G(\cdot)$ são considerados C^r equivalentes ($r \leq k$) se existe um *homeomorfismo* ou *difeomorfismo* H que leva órbitas $\phi_t(x_0)$ de $F(\cdot)$ para órbitas $\psi_t(y_0)$ de $G(\cdot)$ de tal forma que

preserva a orientação [10].

O mapeamento H pode ser considerado como sendo uma mudança de coordenadas inversível, possivelmente não-linear, que, apesar de distorcer o fluxo, fá-lo-á suavemente e sem embaralhar a ordem em que os pontos da trajetória são visitados [10]. Quando $r = 0$, H é um *homeomorfismo*, ou seja, um mapeamento contínuo e descrito por uma função bijetiva. Isto é chamado de *equivalência topológica* ou C^0 . Entretanto se forem satisfeitas as condições de suavidade, então tem-se a *equivalência diferenciável* (C^r , $r \geq 1$), que é mais forte que a anterior.

2.3.3 Estruturas preservadas

Seguem algumas consequências da equivalência topológica [10]:

- $x \in S$ é um ponto de equilíbrio de $F(\cdot)$ se e somente se $H(x)$ for um ponto de equilíbrio de $G(\cdot)$;
- A órbita de x para o campo vetorial $F(\cdot)$ é fechada se e somente se a órbita de $H(x)$ para o campo vetorial $G(\cdot)$ for fechada;
- A aplicação de H sobre a condição x_0 evoluída por ϕ_t é a aplicação de ψ_t sobre $H(x)$:

$$H\phi_t(x_0) = \psi_t(Hx_0).$$

Objetos topológicos definidos pelo fluxo são preservados pela relação de equivalência. A equivalência topológica preserva as propriedades de estabilidade de pontos fixos, mas não distingue entre *nós*, *nós impróprios* ou *focos*. Já no caso da equivalência diferenciável todos estes pontos fixos podem ser caracterizados [10]. Desta forma, membros da mesma classe de equivalência terão, qualitativamente, a mesma dinâmica.

Observa-se, desta forma, que é desejável que H seja um difeomorfismo, para que os pontos fixos possam ser classificados corretamente. A seguir serão mostradas as condições que devem ser satisfeitas para que o mapeamento H seja um difeomorfismo. Satisfazendo algumas condições, as propriedades dos pontos fixos serão preservadas entre os dois espaços de fase.

2.4 Mergulho de uma série escalar

Para o estudo de uma série temporal originária de um sistema dinâmico, descrito por equações diferenciais não-lineares, é desejável conhecer a geometria do espaço de fase original. Como isto não é possível, é necessário saber como proceder na análise da dinâmica quando apenas uma série escalar é conhecida.

Conforme já foi mencionado, pode-se construir um espaço de fase que possua as mesmas propriedades que o espaço de fase original. Este espaço de fase costuma ser chamado simplesmente de *espaço de fase reconstruído*.

A reconstrução do espaço de fase costuma ser feita através de coordenadas de atraso, tal como foi feito no capítulo 1. Entretanto existem outras metodologias para que se construa coordenadas vetoriais a partir de uma série temporal de medidas escalares. A seguir serão apresentadas algumas destas metodologias.

Um sistema dinâmico é dado por

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}),$$

ou por

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n),$$

mas a série temporal tratada é, geralmente, apenas uma sequência de medições escalares (s_n) e será uma função das variáveis dinâmicas do sistema ($\mathbf{x}_n \in \mathcal{S}$). Assim

$$s_n = h(\mathbf{x}_n) + \eta_n,$$

onde h é a *função-mensuração* e η_n é o ruído a que a observação está sujeita.

No caso do sistema dinâmico ser descrito por uma equação diferencial, as variáveis $\mathbf{x}_n$ podem ser interpretadas como uma quantização das variáveis que descrevem este sistema dinâmico. Assim, $\mathbf{x}_n = \mathbf{x}(n \cdot \Delta t)$, onde Δt é a taxa de amostragem da série temporal.

Um sistema dinâmico é totalmente definido através do seu estado. Assim, toda medição é uma função bem definida do estado $\mathbf{x}$. Por isso é que se pode escrever $s = h(\mathbf{x})$ [25].

A relação entre a série temporal observada $\{s_n\}_{n=1}^N$ e as variáveis dinâmicas $\mathbf{x}$ é dada por uma função sobrejetiva. Entretanto, esta função não é injetiva, ou seja, não é verdade que exista um único $\mathbf{x}$ que satisfaça $s = h(\mathbf{x})$. Mas, a partir das medições escalares s , podem ser obtidas medidas vetoriais $\mathbf{y}$, tal que exista um único $\mathbf{x}$ que satisfaça

$$\mathbf{y} = \mathbf{H}(\mathbf{x}),$$

onde H é um difeomorfismo (portanto bijetiva), y é um estado reconstruído, e o espaço onde habitam os vetores y é chamado de *espaço de fase reconstruído*.

Vale ressaltar que x não pode ser determinado a partir de y , mas existe uma relação única entre estes pontos (os do espaço de fase original e os do espaço de fase reconstruído), pois existe uma equivalência entre estes espaços. Assim, não podemos determinar x , porque o difeomorfismo $H(\cdot)$ não pode ser determinado. Mas o atrator definido pelos pontos vetoriais y (do espaço de fase reconstruído) terá as mesmas propriedades do atrator do espaço de fase original (determinado pelas variáveis x).

Será necessária alguma metodologia para a construção dos vetores y a partir de uma série escalar. Duas metodologias serão analisadas: *coordenadas de derivadas* e *coordenadas de atraso*.

2.4.1 Coordenadas de derivadas

Como o sistema dinâmico no espaço de fase reconstruído é descrito por

$$\dot{y} = G(y),$$

pode-se tentar escrever algum tipo de equação diferencial que descreva o espaço de fase reconstruído.

Uma metodologia muito comum, ao se trabalhar numericamente com equações diferenciais, é converter uma equação diferencial de ordem n em um conjunto de n equações diferenciais de primeira ordem [28].

Uma equação diferencial de ordem n , pode ser escrita como:

$$x^{(n)} = f(x^{(n-1)}, x^{(n-2)}, \dots, \dot{x}, x, t).$$

Esta equação pode ser reescrita como um conjunto de n equações diferenciais de primeira ordem, primeiramente define-se

$$x_1(t) \equiv x(t),$$

depois são definidas variáveis que correspondem às derivadas da variável original:

$$\begin{aligned} x_2(t) &\equiv \dot{x}(t) \\ x_3(t) &\equiv \ddot{x}(t) \\ &\vdots \\ x_{n-1}(t) &\equiv x^{(n-2)}(t) \\ x_n(t) &\equiv x^{(n-1)}(t). \end{aligned} \tag{2.2}$$

Reescrevendo como um conjunto de equações diferenciais de primeira ordem:

$$\begin{aligned}\dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= x_3(t) \\ &\vdots \\ \dot{x}_{n-1}(t) &= x_n(t) \\ \dot{x}_n(t) &= f(x_n, x_{n-1}, \dots, x_2, x_1, t).\end{aligned}$$

Semelhantemente, pode-se definir coordenadas de um vetor a partir das derivadas da série temporal [23], ou seja,

$$\begin{aligned}y_1 &= s \\ y_2 &= \dot{s} \\ &\vdots \\ y_{n-1} &= s^{(n-2)} \\ y_n &= s^{(n-1)}.\end{aligned}$$

Todas estas derivadas são obtidas numericamente, a partir das diferenças entre os valores da série temporal. Desta forma, a primeira derivada poderia ser escrita como:

$$\dot{s}(t) = \frac{s(t + \Delta t) - s(t - \Delta t)}{2\Delta t}.$$

Apesar de ser possível construir coordenadas para os pontos do espaço de fase desta forma, esta metodologia é muito sensível ao ruído presente na série temporal. Isto ocorre porque, à medida que utiliza-se derivadas de ordem superior, é feita uma combinação de mais pontos da série temporal (todos sujeitos ao ruído). A primeira derivada, por exemplo, é uma combinação de dois pontos da série temporal, assim a primeira derivada da série apresenta um nível de ruído superior ao da série. O nível de ruído aumenta a cada derivada de ordem superior que é considerada.

A metodologia que costuma ser utilizada para reconstruir o espaço de fase é a utilização de coordenadas de atraso. Esta metodologia, que será apresentada logo a seguir, não está sujeita aos problemas de ampliação do ruído que ocorre com as coordenadas de derivadas.

2.4.2 Coordenadas de atraso

A reconstrução do espaço de fase é geralmente feita através do *método de Takens de mergulho com coordenadas de atraso*. Mais adiante será explicado

outro método, que consegue construir pontos no espaço de fase, realizando simultaneamente a filtragem do ruído.

A partir da série $\{s_i\}_{i=1}^N$, podem ser obtidos os vetores y_i do espaço de fase reconstruído, dados por:

$$y_n = H(x_n) \equiv \begin{pmatrix} h_1(x_n) \\ h_2(x_n) \\ \vdots \\ h_m(x_n) \end{pmatrix} = \begin{pmatrix} h(x_n) \\ h(x_{n-\tau}) \\ \vdots \\ h(x_{n-(m-1)\tau}) \end{pmatrix} = \begin{pmatrix} s_n \\ s_{n-\tau} \\ \vdots \\ s_{n-(m-1)\tau} \end{pmatrix},$$

onde τ é o atraso temporal e m é a dimensão de imersão. O teorema de Takens afirma que se o mergulho for feito com um número suficiente de dimensões, então existirá um difeomorfismo, ou seja, um mapeamento do espaço de fase reconstruído para o espaço de fase original que preserva a informação sobre as derivadas da trajetória original.

Sauer, York e Casdagli [20] provaram importantes generalizações do teorema de Takens do mergulho com atraso temporal. Se a dimensão de capacidade for d , então o número m suficiente da dimensão para o mergulho será o primeiro inteiro maior que $2d$. Para $d = 1$, o suficiente será $m = 3$. É importante dizer que este não é o número necessário de dimensões, mas suficiente.

Para que a correspondência entre x e y seja *bijetiva*, a dimensão m do vetor y deve ser suficientemente grande. A dimensão m do vetor y e, conseqüentemente, do espaço de fase reconstruído deve ser maior do que duas vezes a dimensão de capacidade. Ou seja, maior que duas vezes o número de graus de liberdade ativos do sistema.

2.4.3 Dimensão de Imersão

Para se escrever um vetor y (do espaço de fase reconstruído) em função de um vetor x (do espaço de fase original), a dimensão do vetor y deve ser apropriada. De acordo com o que foi apresentado, a dimensão do espaço de fase reconstruído seria no máximo $2d + 1$, onde d é a dimensão do espaço de fase original, ou a dimensão do atrator original, quando somente a dinâmica do regime assintótico deve ser reconstruída. Entretanto, ao medir uma série temporal, não se sabe a dimensão do espaço de fase original, ou mesmo a dimensão do atrator neste espaço de fase. É preciso saber então como formar os vetores y e qual deve ser sua dimensão.

O atrator formado pelos vetores y_n é equivalente ao atrator no espaço de fase original (desconhecido) se a dimensão m da reconstrução do espaço de fase for suficientemente grande. Assim como foi apresentado no capítulo 1, quando a dimensão do espaço de fase reconstruído não for suficientemente grande, teremos diferentes trajetórias do espaço de fase se intersectando. Somente quando for utilizado um certo valor mínimo para esta dimensão é que o espaço de fase estará reconstruído adequadamente. A demonstração deste resultado foi dada originalmente por Takens [19] (considerando a reconstrução do espaço de fase com dimensão inteira) e posteriormente por Sauer et al. [20] (considerando a reconstrução do atrator com dimensão de contagem de caixa fracionária).

2.4.4 Imersão e mergulho

Um espaço de fase numa variedade pode ser mergulhado em $\mathbb{R}^m$. Este mergulho é dado por um mapeamento que preserva algum tipo de estrutura presente na variedade do espaço de fase original, portanto o mergulho é um morfismo.

Em topologia, o mergulho é um homeomorfismo. Quando este mergulho é suave, sendo descrito por uma função C^r ($r \geq 1$), o mergulho é chamado de *imersão*, sendo um difeomorfismo.

A variedade que descreve o espaço de fase original provavelmente será curva. Assim, para mergulhá-la em $\mathbb{R}^m$, será necessária uma dimensão m elevada, para que não ocorram intersecções das trajetórias neste novo espaço de fase. De acordo com Whitney [24], uma variedade suave d -dimensional pode ser mergulhada em $\mathbb{R}^{2d+1}$. Observe que a condição $m = 2d + 1$ é uma condição suficiente, mas não necessária. Portanto, em muitos casos, a variedade poderá ser imersa em $\mathbb{R}^m$, com $m < 2d + 1$.

O teorema de Whitney não levava em consideração uma série temporal escalar. Este teorema falava simplesmente da imersão de uma variedade. Foi Takens quem generalizou o mergulho para uma série escalar. Ele mostrou que *vetores de atraso* de dimensão m poderiam ser usados para mergulhar o espaço de fase original (desconhecido) de dimensão d em $\mathbb{R}^m$. Para isto, seria necessário que $m = 2d + 1$.

Posteriormente, foi mostrado no artigo *Embedology* [20] que o atrator do espaço de fase original poderia ser mergulhado em $\mathbb{R}^m$, desde que m fosse maior que o dobro da *dimensão de capacidade* do atrator. Isto representa uma generalização em relação ao trabalho de Takens, pois Takens havia considerado o espaço de fase com dimensão inteira. No artigo *Embedology* foi

considerado o atrator com dimensão fractal, podendo descrever também o mergulho de uma variedade com dimensão inteira. No caso da dimensão inteira, obtém-se o mesmo resultado que aquele proposto por Takens.

2.5 Redução de ruído com imersão filtrada

Se m pontos forem suficientes para construir um estado do sistema no espaço de fase, então com um número maior de pontos teríamos apenas informação redundante, desde que a série temporal estivesse isenta de ruído. Entretanto, uma vez que a série temporal disponível está cheia de ruído, pode-se utilizar um número maior de pontos da série para obter um ponto m -dimensional filtrado, ou seja, com taxa de ruído reduzida [20].

Tomando uma janela de tamanho w de pontos consecutivos da série temporal, obtém-se:

$$y_n = B \begin{pmatrix} s_n \\ s_{n+1} \\ \vdots \\ s_{n+(w-1)} \end{pmatrix}, \quad (2.3)$$

onde B é uma matriz $m \times w$ e y_n é um vetor de dimensão m . Os vetores y_n formam um mergulho se o posto da matriz B for suficientemente grande, e se B não colapsar pontos periódicos do atrator que apresentem período inteiro menor que ou igual a w [20].

Surge agora uma nova questão: qual deve ser o tamanho adequado desta janela (w)? Grande parte dos trabalhos utilizam a imersão por coordenadas de atraso. Esta forma, a preocupação nestes artigos está na escolha da dimensão de imersão e do atraso temporal para formar os vetores de atraso. Aqui utiliza-se a imersão filtrada, portanto deve ser primeiramente escolhida uma janela de tamanho adequado. A escolha dos parâmetros m e τ , stá sendo coberta uma janela de tamanho aproximadamente igual a $m \times \tau$, portanto a escolha ótima da janela deve levar em consideração os mesmos fatores do que a escolha de ambos os parâmetros (m e τ).

De acordo com Sauer [27], a janela deve cobrir de 1 a 2 períodos da presentes na série temporal. Nem sempre é fácil medir isto, mas em alguns casos, como na série originária do sistema de equações de Lorenz, é possível encontrar um período médio. No caso de séries com ruído, é importante cobrir mais do que um período de oscilação da série.

Existem várias formas de se realizar a filtragem. Vários algoritmos desenvolvidos para o processamento de sinais digitais podem ser utilizados. A maioria dos algoritmos funciona muito bem com séries lineares, mas não com séries não-lineares. É bom ressaltar aqui que a filtragem não está sendo realizada na série como um todo. Apenas um pequeno intervalo da série está sendo usado para obter um vetor filtrado no espaço de fase. Assim, localmente, a série não-linear pode ser aproximada por uma série linear.

2.5.1 Transformada de Fourier

Uma das ferramentas mais utilizadas no processamento de sinais digitais é a transformada de Fourier [26, 39]. A análise de Fourier é de grande importância no estudo de séries temporais, principalmente das séries temporais lineares. Através da utilização da transformada de Fourier pode-se obter informação sobre uma série temporal. Esta mesma transformada é capaz de filtrar o ruído presente na série.

A transformada discreta de Fourier é dada por:

$$\tilde{x}_k = \sum_{n=1}^w x_n e^{-2\pi i n k / w} \quad (k = 1, 2, \dots, w),$$

e a transformada inversa é dada por:

$$x_n = \frac{1}{w} \sum_{k=1}^w \tilde{x}_k e^{2\pi i n k / w} \quad (n = 1, 2, \dots, w),$$

onde w é o tamanho da janela a ser filtrada. É comum aplicar-se a transformada de Fourier para toda a série. Mas no nosso caso a filtragem será aplicada somente num pequeno trecho da série de tamanho w .

Para obter cada termo da série de Fourier, de acordo com a definição acima, seria necessário algo da ordem de w operações. Portanto, para obter todos os w termos da transformada seriam necessárias w^2 operações. Entretanto, utilizando a transformada rápida de Fourier o número de operações cai para $O(w \log w)$ [26], o que torna o algoritmo muito mais eficiente.

Utilizando a transformada de Fourier, obtém-se as frequências presentes naquele trecho da série temporal. A partir do conhecimento destas frequências é possível remover parte delas, justamente aquelas que estão associadas ao ruído.

O espectro de potências indica se um sistema é periódico ou quase-periódico. Uma série temporal com frequência ω apresenta um δ de Dirac na frequência

ω e outros nos seus harmônicos ($2\omega, 3\omega, \dots$). Um sistema quase-periódico com frequências $\omega_1, \omega_2, \dots, \omega_k$ apresenta deltas nestas posições e também em todas as combinações lineares destas frequências com coeficientes inteiros [39].

Num espectro de frequências experimental não temos δ de Dirac, mas picos com amplitude limitada e largura $2\pi/w$, onde w é o tamanho da janela da série temporal [39].

A análise de Fourier de uma série temporal não-linear gerada por um sistema dinâmico produz um *espectro de potências contínuo*. Um sistema não-linear pode conter picos, mas eles são menores e mais largos. Além disso, este sistema apresenta um espectro de potências semelhante a um ruído de fundo, que, entretanto, não caracteriza dimensão infinita [23].

O espectro de potências é, em geral, muito bom para a análise de sistemas lineares. Entretanto, no estudo de sistemas não-lineares, não se obtém benefícios com o espectro de potências da série temporal, pois neste caso, a dimensão do atrator não está relacionada com o número de frequências independentes no espectro de potências. Entretanto, podemos usar a análise de Fourier localmente, ou seja, a análise é feita em janelas com tamanho bem menor do que o tamanho da série a disposição. Assim como discutido anteriormente, esta janela deve ter tamanho que corresponda a cerca de um ou dois períodos da série temporal.

Apesar da análise de Fourier Discreta poder ser utilizada como um filtro linear, existe um problema que deve ser atentado, para que não seja introduzido um viés no resultado obtido. A descrição é dada a seguir.

2.5.2 Filtragem e Interpolação por Fourier

Na figura 2.2, observa-se um trecho da série temporal com 128 pontos, gerado pelo sistema de equações diferenciais conhecido como Lorenz. A Transformada de Fourier vê a série temporal como uma sequência periódica a cada 128 pontos. Isto é problemático, pois o ponto 1 é distinto e separado do ponto 128, porque foram adquiridos em instantes diferentes. Assim, existe uma descontinuidade entre o ponto 128 e o ponto 1, o que, no domínio da frequência é encarado como uma alta frequência. Portanto, se for aplicado um *filtro passa-baixa*, a série temporal será deformada nas suas extremidades. Isto pode ser observado na versão filtrada da mesma série temporal apresentada na figura 2.3. Nesta filtragem metade das frequências foram retiradas, exatamente aquelas correspondentes às frequências mais altas.

A transformada de Fourier analisa estes 128 pontos como sendo um único

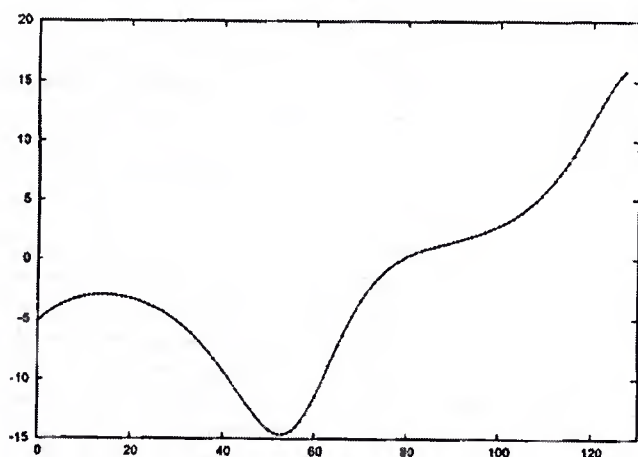


Figura 2.2: Um trecho da série obtida a partir do sistema de equações de Lorenz.

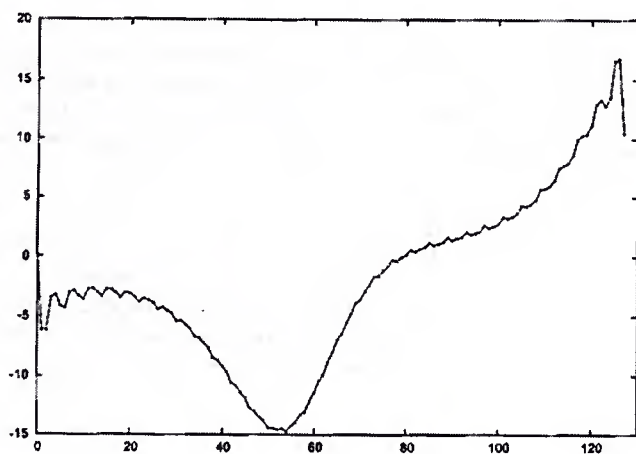


Figura 2.3: A mesma série apresentada na figura anterior, mas agora com a filtragem por Fourier.

período de um sinal periódico infinitamente longo. Isto significa que o lado esquerdo do sinal adquirido está conectado ao lado direito de um sinal duplicado. Do mesmo modo, o lado direito do sinal obtido é conectado ao lado esquerdo de um período idêntico. Isto pode também ser pensado como o lado direito do sinal girando e conectando-se a seu lado esquerdo. Assim a amostra 128 ocorre ao lado da amostra 1, assim como, por exemplo, a amostra 64 ocorre ao lado da amostra 65. O sinal é considerado como sendo circular, ou periódico [26].

Na filtragem por série de Fourier, assumimos que as frequências mais altas correspondem ao ruído. Portanto estas frequências são eliminadas, enquanto que as mais baixas são preservadas. Não se pode garantir que todo o ruído seja removido da série temporal com este procedimento, mas certamente a série temporal será mais suave do que a série com ruído. Quando se trabalha com um sistema de equações diferenciais, o fluxo (trajetória) será suave. Portanto o procedimento de eliminar altas frequências auxilia a obter uma série suave. No caso de um inapeamento do tipo:

$$x_{n+1} = f(x_n),$$

a série resultante não será necessariamente suave, portanto o procedimento de filtragem por transformada de Fourier não deve ser usado neste caso.

A filtragem por Fourier funciona bem para séries contínuas. Séries descontínuas quando são aproximadas por uma série de Fourier apresentam este problema de oscilações, principalmente na região próxima à descontinuidade, como pode ser visto na figura 2.3. Para corrigir este problema, ao lado direito da série copia-se o seu 'espelho' [26]. O espelho da série é a inversão temporal da série. A série acrescentada de seu espelho pode ser observada na figura 2.4.

Agora a série temporal não possui mais um salto, apresentando continuidade inclusive entre o ponto 256 e o ponto 1. Portanto a filtragem por Fourier pode ser aplicada sem que ocorram problemas de oscilações como na figura 2.3. Como o espelho da série foi acrescentado apenas como uma estratégia para eliminar a descontinuidade, este trecho pode ser descartado. Depois de aplicada a filtragem e removido o espelho, obtém-se o resultado mostrado na figura 2.5.

Filtragem por Fourier

Agora a filtragem por Fourier é apresentada de forma resumida. É importante observar que um trecho da série com w pontos apresenta $w/2$ frequências.

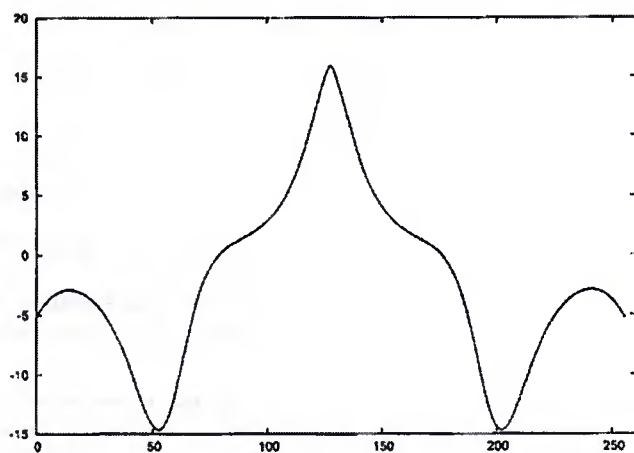


Figura 2.4: A mesma série temporal apresentada na figura 2.2 com sua reversão temporal acrescentada à direita.

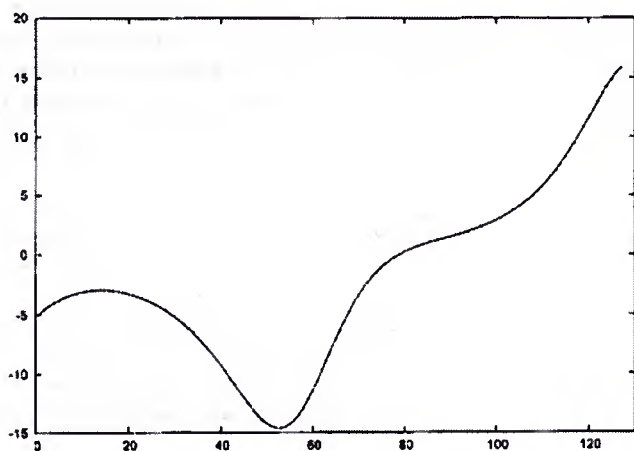


Figura 2.5: Filtragem da série apresentada na figura 2.2 aplicando a técnica de reversão temporal apresentada na figura 2.4 e explicada no texto.

Quando se efetua uma filtragem por um fator t_f , então uma fração $1/t_f$ das frequências inferiores serão mantidas, enquanto as demais serão zeradas.

Para se obter uma filtragem de um trecho de tamanho w , por um fator de filtragem t_f , são realizados os seguintes passos [26]:

1. Faz-se a Transformada de Fourier Discreta de tamanho w , obtendo $w/2$ frequências;
2. Mantém-se as $\frac{w}{2t_f}$ frequências inferiores, zerando-se as demais;
3. Faz-se a Transformada de Fourier Discreta Inversa de tamanho w . Obtendo-se assim a série filtrada.

Para as séries observadas nas figuras anteriores, foi adotado este procedimento que acabou de ser apresentado. Entretanto, na imersão filtrada foi adotado um procedimento um pouco diferente.

Observando o procedimento que foi relatado, observa-se que a série temporal resultante possui menos informação que a série original. Assume-se que a informação que foi perdida (as altas frequências) constituem o ruído da série temporal.

Ao realizar a imersão da série no espaço de fase, deseja-se que qualquer tipo de informação redundante seja eliminada. Assim, é possível representar os vetores num espaço de dimensão menor. Portanto, ao filtrar a janela da série temporal é possível obter um vetor de tamanho reduzido [27].

O procedimento para realizar uma filtragem que reduz a dimensão é o seguinte [32]:

1. Faz-se a Transformada de Fourier Discreta de tamanho w , obtendo $w/2$ frequências;
2. Mantém-se as $\frac{w}{2t_f}$ frequências inferiores, eliminando as demais;
3. Faz-se a Transformada de Fourier Discreta Inversa de tamanho w/t_f . Obtendo-se assim um vetor de tamanho w/t_f .

Quando utiliza-se uma série temporal de tamanho $w = 32$ e realiza-se uma filtragem de Fourier com fator $t_f = 4$, o vetor resultante terá tamanho $w/t_f = 32/4 = 8$. Portanto, com este procedimento a série temporal é filtrada e mergulhada num espaço de fase com dimensão 8.

Este procedimento é chamado por Sauer [27] de *imersão passa-baixa*, pois a série é imersa no espaço de fase através de um processo que elimina as altas-frequências. Esta sequência de etapas está relacionada com a equação 2.3, que define uma imersão filtrada. Observando a equação 2.3, a matriz B pode ser descrita como o produto de três matrizes B_1 , B_2 e B_3 . O significado das 3 matrizes é o seguinte:

- B_1 : é uma matriz $w \times w$ que faz a transformada discreta de Fourier de tamanho w .
- B_2 : é uma matriz $m \times w$ que elimina as frequências mais altas. Apenas as m frequências inferiores são mantidas, onde $m = w/t_f$.
- B_3 : é uma matriz $m \times m$ que faz a transformada de discreta de Fourier inversa de tamanho m .

Interpolação por Fourier

Para uma boa análise de séries temporais, é necessário que ela apresente uma taxa de amostragem suficientemente alta. Ao se analisar o espaço de fase, é necessário que o ponto a ser evoluído possua vários vizinhos, para que possa ser feito um estudo estatístico das características da dinâmica local do espaço de fase.

Em várias análises utilizam-se os vizinhos mais próximos que não sejam temporalmente consecutivos. Portanto, na vizinhança observada, deve ser escolhido apenas o vizinho mais próximo pertencente a cada órbita. Na figura 2.6.a observa-se um ponto a ser evoluído (representado por um x). São observados também os seus vizinhos mais próximos (representados por círculos). Deseja-se escolher um vizinho em cada órbita, mas a escolha não será boa, pois a taxa de amostragem da série não é suficientemente alta para isto. O ideal é que o vizinho em cada uma das órbitas estivesse sobre a reta que é perpendicular às trajetórias. Para melhorar esta situação, é possível interpolar a série temporal e obter (em cada órbita) um vizinho que esteja mais próximo do ponto a ser evoluído. Isto pode ser observado na figura 2.6.b, onde a série temporal foi interpolada, obtendo-se assim mais pontos no espaço de fase. Estes pontos interpolados foram representados por círculos menores, para distingui-los dos pontos que já estavam presentes na série temporal. No exemplo apresentado na figura 2.6 a taxa de interpolação foi de $t_i = 4$.

Portanto a interpolação é muito importante para se obter vizinhos que estejam mais próximos do ponto que se quer estudar e representem melhor a

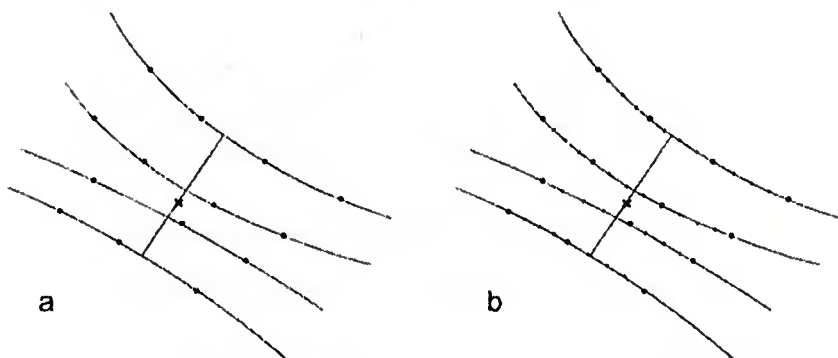


Figura 2.6: Um ponto colocado sobre o espaço de fase (representado por um x) juntamente com os seus vizinhos mais próximos (círculos). Podemos observar que, se a interpolação for utilizada, é possível encontrar um vizinho em cada órbita que esteja mais próximo que os vizinhos observados nesta figura.

dinâmica neste ponto. Existem várias formas de se fazer a interpolação, uma delas é através da transformada de Fourier [26]. Como esta transformada será usada para a filtragem, é interessante utilizá-la também para a interpolação. O mais importante é notar que as duas etapas podem ser combinadas num processo unificado, que acaba reduzindo o tempo computacional [32].

O processo de interpolação por transformada de Fourier, consiste em acrescentar mais frequências à série original. Por exemplo, para dobrar o tamanho da série temporal, interpolando-a, deve-se acrescentar mais frequências à série original, todas elas zeradas [26].

Os passos para se aumentar a taxa de amostragem (de um trecho da série com w pontos) por um fator t_i são [26]:

1. Faz-se a Transformada de Fourier Discreta de tamanho w ;
2. Aumenta-se o tamanho da janela para $w \times t_i$, sendo que os pontos 1 a w permanecem invariantes, e os pontos $w + 1$ até $w \times t_i$ são completados com zeros.
3. Faz-se então a Transformada de Fourier Discreta Inversa de tamanho $w \times t_i$. Obtendo-se assim a série interpolada.

Ao fazer a transformada de Fourier Discreta, são obtidas as frequências presentes na série temporal. Acrescentando frequências superiores (todas elas zeradas) obtém-se a mesma série temporal, mas com uma taxa de amostragem maior. Os pontos interpolados são os mais suaves possíveis no sentido da análise de Fourier [26].

Filtragem e Interpolação combinadas

Para fazer a filtragem e a interpolação simultaneamente, faz-se o seguinte:

1. Faz-se a Transformada de Fourier Discreta numa janela de tamanho w , obtendo $w/2$ frequências;
2. Mantém-se as $\frac{w}{2t_f}$ frequências inferiores, eliminando as demais
3. São acrescentadas frequências superiores, todas elas zeradas;
4. Faz-se a Transformada de Fourier Discreta Inversa de tamanho $w \times t_i/t_f$. Obtendo-se assim uma série de tamanho $w \times t_i/t_f$.

A partir da série filtrada e interpolada, que possui $w \times t_i/t_f$ termos, controem-se os vetores que estarão no espaço de fase. A partir desta série são construídos t_i vetores, cada um deles de tamanho w/t_f . As componentes de cada vetor são formadas tomando elementos da série interpolada espaçados por um intervalo temporal dados por t_i . Nesta etapa o procedimento é igual à construção de vetores de atraso.

Capítulo 3

Modelagem Local

A modelagem de dados é objeto de várias disciplinas, desde estatística linear e regressão não-linear até aprendizado de máquina e sistemas dinâmicos. Na literatura que trata deste problema, uma quantidade enorme de metodologias têm surgido. Entretanto, todos os métodos abrangem uma *arquitetura*, na qual a estrutura do modelo é definida, e a definição dos *parâmetros*, onde o modelo é ajustado aos dados através de um processo de otimização [43].

Neste trabalho descreve-se um tipo de aprendizado de máquina, baseado na teoria de sistemas dinâmicos, que reconhece o espaço de fases e evolui pontos neste espaço, produzindo diversas realizações da série temporal, a partir de diferentes condições iniciais.

3.1 Aprendizado de Máquina

Existem basicamente dois tipos de aprendizado de máquina, o *aprendizado preguiçoso* (*lazy learning*) e o *aprendizado ansioso* (*eager learning*) [42]. O aprendizado é considerado ansioso quando o modelo é construído e disponibilizado para qualquer resposta futura. Ansioso pois estuda todo o conjunto de dados em todos os seus regimes antes que qualquer pergunta seja feita.

Já o aprendizado preguiçoso é aquele no qual se constrói um modelo somente quando se tem uma *pergunta**. Neste caso o modelo é reconstruído para cada nova pergunta que é feita. Este tipo de aprendizado recebe o nome

*No caso de sistemas dinâmicos, a pergunta é um vetor no espaço de fase a ser evoluído. O vetor pergunta é usado como entrada do modelo que dará uma resposta. No nosso caso, a resposta é a evolução do ponto no espaço de fase.

de preguiçoso justamente porque ele não aprende de antemão. Ao invés disso, esta modelagem espera que a pergunta seja feita, para que então comece a estudar os dados para se responder a pergunta original.

Nos modelos ansiosos o conjunto de dados é estudado para se obter um modelo. A base de dados inicial é então descartada, e a descrição é obtida a partir do modelo que responde às perguntas.

Nos modelos preguiçosos os dados nunca são descartados e são as únicas informações disponíveis quando a pergunta é feita. Neste momento ainda não há nenhum modelo. Assim, o modelo preguiçoso adia o processamento dos dados até que haja uma pergunta. Quando isto ocorre, procura-se, na base de dados disponível, exemplos que, de acordo com alguma métrica, sejam mais relevantes para se responder a pergunta. A partir destes exemplos constrói-se um modelo local que possa dar uma resposta apropriada. Neste ponto, a resposta e todas as informações intermediárias são descartadas e, para cada nova pergunta, todo o processo será realizado novamente.

O modelo preguiçoso costuma ser local, porque o modelo é construído usando-se apenas uma pequena vizinhança de pontos. Em contraste, os modelos ansiosos são globais, pois a sua construção utiliza todo o conjunto de dados, que são então descartados.

3.2 Modelagem global $\times$ modelagem local

Métodos preguiçosos geralmente usam aproximações locais, que produzem um comportamento altamente adaptativo, que geralmente não é encontrado em algoritmos ansiosos. É possível criar um algoritmo que utilize todos os dados disponíveis para um aprendizado preguiçoso; ou seja, um *modelo preguiçoso global*. Assim poder-se-ia criar uma metodologia que a cada nova pergunta estudasse todo o conjunto de dados, fazendo uma ponderação (dando uma importância maior para os exemplos mais relevantes àquela pergunta). Entretanto, esta metodologia não é viável, pois o gasto computacional seria impraticável. Portanto um modelo preguiçoso é sempre local.

Num modelo ansioso, ou seja, numa modelagem global um único modelo é usado para ajustar todos os dados de treinamento, todo o espaço com seus vários regimes. E, a partir de então, pode-se perguntar à vontade que as respostas serão dadas. Primeiramente, os exemplos são compilados num modelo, então a base de dados inicial é descartada e a descrição obtida é usada para responder a pergunta feita. Para modelos ansiosos costuma-se utilizar: árvores de decisão, redes neurais, funções de base radial, etc [41, 35].

A seguir é apresentada uma lista comparativa das vantagens e desvantagens dos modelos locais e dos modelos globais.

Controle da Não-Linearidade

Uma vantagem dos modelos locais é que eles dão controle direto da quantidade de não-linearidade através do parâmetro k , o número de vizinhos para construção do modelo. Conforme k aumenta, os modelos locais tornam-se menos locais e mais suaves. No limite em que k torna-se igual ao número de pontos do conjunto de dados o modelo local torna-se um modelo global.

Esta característica dos modelos locais tem sido usada como uma medida de não-linearidade. Casdagli propôs avaliar o desempenho dos modelos locais para vários valores de k . Em processos estocásticos, o valor de k deveria ser grande, próximo de todo o conjunto de dados. Em processos determinísticos de baixa dimensão, o valor de k deveria ser baixo, de forma que apenas a dinâmica local fosse observada pelo modelo [50].

Complexidade Adaptativa

Outra vantagem dos modelos locais é que o grau de não-linearidade adapta-se à distribuição do conjunto de dados. Em regiões onde os pontos estão bastante espaçados, a região coberta pela vizinhança é automaticamente aumentada, desde que o número de vizinhos seja fixado. Este uso eficiente dos dados não é possível com a maioria dos modelos globais.

Preprocessamento

Como o modelo local não é construído até que o vetor pergunta esteja disponível, os modelos locais não têm a necessidade de praticamente nenhum preprocessamento. Entretanto, técnicas de modelagem local mais avançadas (mergulho passa baixa, projeção num subespaço) requerem algum preprocessamento para selecionar os parâmetros do modelo local.

Numa rede neural, por exemplo, os valores de entrada devem estar dentro de uma escala. Em geral, entre -1 e 1, ou entre 0 e 1. Às vezes, recomenda-se uma escala num intervalo inferior a estes [35].

Custo Computacional para Calcular as Saídas

Algoritmos locais costumam ter um custo computacional mais baixo do que algoritmos globais durante a fase de treinamento. (Os algoritmos locais geralmente não possuem fase de treinamento.) Entretanto, os algoritmos locais têm necessidade de grande espaço para memória, além de um alto custo computacional durante a resposta à pergunta.

Modelos globais geralmente podem produzir as saídas muito mais rapidamente que modelos locais, pois os modelos locais têm que encontrar os vizinhos mais próximos e construir o modelo local para cada vetor pergunta.

Felizmente, o tempo para gerar as saídas do modelo local não é algo impraticável. Isto principalmente porque existem algoritmos que podem reduzir o tempo para encontrar os vizinhos mais próximos.

O tempo necessário para a construção do modelo local é menor do que aquele para se encontrar os vizinhos, mesmo que seja utilizado um algoritmo rápido de busca de vizinhos. Isto porque modelos locais são contruídos usando-se apenas um pequeno subconjunto do conjunto de dados e porque a construção do modelo local não é iterativa. Nos modelos globais, a construção dos modelos costuma ser iterativa, assim como ocorre com as redes neurais.

Como se deve fazer uma busca no conjunto de dados para se encontrar os vizinhos mais próximos, o tempo requerido para a saída dos modelos locais cresce conforme o tamanho do conjunto de dados aumenta. Usando uma busca por força bruta, o tempo da localização dos vizinhos aumenta com $O(N^2)$ onde N é o número de pontos do conjunto de dados. Usando-se árvores, para a maioria dos conjuntos de dados obtidas de processos reais, o tempo da localização dos vizinhos aumenta com $O(N \log N)$. Já o algoritmo proposto por Schreiber leva um tempo $O(N)$ [40]. Isto reduz substancialmente o custo de se usar modelos locais num conjunto de dados grande, tornando-os viáveis para a maioria das aplicações.

Seleção Limitada das Medidas de Erro

Uma desvantagem dos modelos locais é que a otimização dos pontos vizinhos está vinculada a minimização da soma dos erros quadráticos, ou equivalentemente, o erro quadrático médio. Isto porque não há nenhuma solução com forma fechada para o modelo ótimo para outras medidas de erro. A utilização de outras medidas de erro implicaria na utilização de um algoritmo iterativo

para a minimização do erro, o que acarretaria num custo computacional extremamente alto.

Modelos globais geralmente utilizam um algoritmo de otimização iterativo, às vezes chamado de programação não-linear, para adaptar os parâmetros do modelo. A maioria destes algoritmos requer que o gradiente do erro seja contínuo com relação aos parâmetros do modelo, o que, por sua vez, requer que a medida do erro seja contínua. Esta família de funções erro inclui a soma ponderada dos erros quadráticos usada por modelos locais e muitas outras funções erro [35].

3.3 Modelagem local

A modelagem local tem apresentado bons resultados na literatura que trata de séries temporais não-lineares determinísticas. Através desta modelagem é possível tratar diferentes regiões do espaço de fase com diferentes modelos. Isto costuma resultar em algoritmos com melhor performance.

Para apoiar o uso da modelagem local, é importante salientar que a modelagem local tem obtido bons resultados em competições internacionais, quando comparada à modelagem global.

3.3.1 As modelagens local e global em competições

Na competição realizada no Santa Fe Institute em 1991 [29], foram apresentadas várias séries temporais para que pesquisadores enviassem suas previsões. A série que recebeu o maior número de previsões foi o "Conjunto de Dados A". As duas previsões mais acuradas da competição foram realizadas por Sauer [32] e por Wan [31], e tratavam desta mesma série. A previsão de Wan foi feita com uma rede neural, um modelo global utilizado como um aproximador de funções. Já o método de Sauer foi baseado na representação da geometria do espaço de fases do sistema que produziu a série temporal; uma modelagem local (preguiçosa).

O "Conjunto de Dados A" consistia de uma série de 1000 pontos, onde pedia-se que fossem previstos os próximos 100. Os melhores resultados foram obtidos por Wan [31]. Entretanto, não se pode dizer que ele foi o vencedor pois, quando os algoritmos de Sauer [32] e Wan [31] foram extrapolados para que previssem não apenas 100 pontos, mas 600, os resultados foram diferentes. Observou-se então, conforme dito pelos organizadores da competição,

que o algoritmo criado por Sauer capturava as características da dinâmica de forma muito melhor que o algoritmo de Wan. Também segundo os organizadores da competição, o algoritmo de Sauer era capaz de prever os dados com maior confiança num horizonte de tempo maior [30]. Portanto o algoritmo que usava uma modelagem local foi o que obteve melhores resultados.

Uma segunda competição deve ser mencionada. Esta ocorreu na Bélgica, no *International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling: Theory and Applications* [33]. A organização desta competição disponibilizou uma série temporal com 2000 pontos. Aos competidores foi pedido que previssem os próximos 200. O campeão desta competição foi McNamara [34] utilizando um modelo local.

De acordo com estas competições, observa-se que a modelagem local costuma dar resultados superiores à modelagem global. Portanto ela deve ser a escolha natural ao se realizar a previsão de uma série temporal. O mesmo deve ser feito ao se realizar a reconstrução das séries temporais. Portanto a modelagem local foi adotada nesta tese.

3.3.2 Dados em tempo real

O modelo preguiçoso conduz a algoritmos que podem ter melhoria de performance com o tempo. Isto ocorre porque pode ser utilizado mesmo quando os dados não estão todos disponíveis no início da análise, mas são coletados em tempo real. Neste caso, um novo exemplo requer apenas que se atualize a base de dados, diferente do modelo ansioso que necessitaria que toda a modelagem fosse refeita.

Além disso, o aprendizado preguiçoso não sofre de interferência, ou seja, adquirindo-se novos exemplos sobre um regime não degrada a performance da modelagem de outros. Entretanto, existe o problema da grande necessidade de memória para armazenamento de informação; e também o problema de velocidade, por ter que criar um novo modelo para cada pergunta que é feita. Mas estes problemas têm sido superados devido à constante evolução do *hardware*.

3.3.3 Arquitetura do modelo local

Para responder a cada pergunta, é necessária uma modelagem local que, como dito anteriormente, pode ser vista como composta de uma *arquitetura* e a identificação de seus *parâmetros*.

No aprendizado preguiçoso é preciso encontrar os exemplos mais relevantes no conjunto de dados e obter a evolução de cada exemplo. A partir destes dados (de entrada e de saída) um modelo local é construído e ajustado [41].

A arquitetura do modelo local abrange as seguintes etapas [43]:

- a seleção de uma métrica para se obter os vizinhos relevantes;
- a escolha de uma função ponderação para os vizinhos;
- a seleção da estrutura do modelo local. Por exemplo, a escolha entre polinômios de vários graus;
- a escolha do número de exemplos (vizinhos) necessários para a modelagem local.

Em cada etapa ocorre a escolha de algum parâmetro. De uma forma geral, os pesquisadores simplesmente escolhem um valor para alguns destes parâmetros sem que seja feita uma otimização para eles. Certas escolhas, como a função que será usada para ponderar os exemplos mais relevantes, costumam ser feitas sem que se busque uma otimização dos resultados.

É importante salientar que toda a estrutura, assim como os parâmetros, deveriam ser escolhidos de forma a otimizar a modelagem.

3.3.4 Escolha dos parâmetros

Depois que a arquitetura do modelo foi definida, é necessário que os parâmetros sejam otimizados. Mas sempre é necessário que o usuário escolha um valor inicial para os parâmetros. A partir daí, estes parâmetros podem ser variados para que se busque uma minimização dos erros.

A acurácia dos modelos não-lineares geralmente é sensível ao valor destes parâmetros definidos pelo usuário e, portanto, à habilidade e ao conhecimento do usuário.

Vale lembrar que não existe, por exemplo, um método único para se escolher o atraso τ para se formar os vetores do espaço de fases reconstruído. Sendo que cada método pode nos dar um valor diferente para este mesmo parâmetro. A escolha inadequada para este, ou outro parâmetro pode prejudicar toda a modelagem.

3.4 Modelagem local no espaço de fase

Os modelos locais requerem uma série de dados com muitos exemplos das variáveis observadas, a partir das quais é criado um modelo com variáveis de entrada e de saída. É necessário que se tenha muitos dados para que seja possível descrever a variedade onde está o atrator no espaço de fases. Assim, o modelo pode ser construído e usado para estimar a saída para um dado ponto no espaço de fases, que não seja parte da série de dados original. Este ponto, que se busca encontrar a evolução, é chamado de pergunta (q - *query*); e a saída do modelo local, $\hat{y}(q)$, é chamada de resposta.

A saída do modelo local é calculada usando-se três estágios consecutivos. Assim [15, 32]:

1. No primeiro estágio a pergunta formulada, ou seja, o ponto a ser evoluído (q) é colocado no espaço de fases e os k vizinhos mais próximos são localizados.
2. No segundo estágio constrói-se um modelo simples usando somente os k vizinhos mais próximos e suas respectivas evoluções. De uma forma geral, os vizinhos costumam ser ponderados de acordo com a distância. Àqueles mais próximos é dado um peso maior, enquanto que aos mais distantes é dado um peso menor.
3. No terceiro estágio, quando o modelo está construído na vizinhança da pergunta (q), a pergunta é utilizada como entrada deste modelo local, para estimar a saída do processo, $\hat{y}(q)$.

No modelo local, são feitas várias modelagens nas diversas regiões do espaço de fases, onde encontram-se os dados originais. Assim, podem haver diferentes modelos descrevendo regimes distintos, resultando numa descrição mais acurada, mesmo quando são usadas aproximações simples (modelos lineares, por exemplo) [41]. Cada modelo é obtido após conhecer a pergunta; pois, assim, são encontrados os pontos vizinhos e é então formulado um modelo que dê a resposta.

Alguns exemplos de modelos locais são [41]:

- *Média ponderada*: este é o modelo mais simples, usando como resposta a média ponderada da resposta dos pontos vizinhos mais próximos;
- *Regressão ponderada local*: ajusta uma superfície aos pontos mais próximos usando uma regressão baseada na distância. O tipo mais comum de regressão é a linear.

Capítulo 4

Arquitetura do modelo local

Ao utilizar um modelo local, parece vantajoso mantê-lo simples, mantendo o critério de treinamento igualmente simples. Assim, costumam ser utilizados modelos de média ponderada, ou modelos que são lineares nos seus parâmetros.

Para o treinamento do modelo, que visa obter os parâmetros que otimizam a modelagem, é comum a utilização do critério de mínimos quadrados. Este critério tem a grande vantagem de possuir um baixo custo computacional.

O aprendizado preguiçoso descrito através de uma modelagem local, assim, os exemplos mais relevantes devem ser encontrados para que o modelo seja descrito corretamente. Para que os exemplos mais relevantes sejam encontrados, é necessário que seja definida uma métrica que avalie quais são os exemplos mais relevantes.

Depois que os exemplos mais relevantes são encontrados, é possível dar uma ponderação a cada um, de acordo com o seu grau de relevância. Então um modelo local (usando os exemplos mais relevantes) é usado para descrever a dinâmica local.

No estudo de séries temporais não-lineares determinísticas, o objetivo da modelagem é encontrar a evolução dos pontos neste espaço. De uma forma geral, não é possível descrever todo o espaço de fase, mas apenas o atrator.

No espaço de fase, é possível saber a evolução de cada ponto, pois os mesmos estão temporalmente ordenados. Assim, para cada exemplo, que é uma entrada do modelo, sabe-se a sua evolução, que é a saída do modelo.

A partir destes dados de entrada e saída, é construído um modelo local. Existem diversas possibilidades para o modelo local que será adotado. A

saída do modelo pode ser simplesmente a média da evolução de cada exemplo relevante. Outra possibilidade é que, a partir dos dados de entrada e saída, seja criado um modelo linear. Este modelo é usado então para prever a evolução de novos pontos, que não estão na base de dados original.

O modelo de média ponderada e o modelo linear são os mais comuns na modelagem local, entretanto algumas vezes são utilizados modelos polinômiais ou até mesmo outros modelos para descrever a evolução dos pontos no espaço de fase.

Depois que o modelo foi escolhido, os vizinhos mais próximos são usados para treinar o modelo. Nesta fase os parâmetros são ajustados, de tal forma que o erro de predição seja minimizado usando como conjunto de treinamento os vizinhos mais próximos.

A arquitetura do modelo local abrange quatro pontos, que são:

- a determinação do tamanho da vizinhança a ser analisada;
- a escolha de uma métrica para a distância;
- o tipo de modelo local usado na modelagem (linear, quadrático, ...);
- a forma da função-ponderação que determina o peso dado a cada vizinho.

Apresenta-se nas próximas seções o detalhamento de cada uma destas escolhas.

4.1 Exemplos relevantes

Na modelagem local, para se responder uma pergunta, é necessário observar os exemplos mais relevantes. A partir destes exemplos, é realizada uma regressão que ajusta o valor dos parâmetros do modelo. Em séries temporais não-lineares, um espaço de fase é construído a partir da série temporal. No espaço de fase, os pontos próximos possuem evolução próximas. Assim, os exemplos mais relevantes são os vizinhos mais próximos.

Na literatura de séries temporais não-lineares determinísticas existem duas formas comuns de se escolher uma vizinhança no espaço de fases. Uma delas é a escolha de uma vizinhança de tamanho fixo. Neste caso, o tamanho da vizinhança deve ser superior ao nível de ruído da série temporal. Se o

tamanho da vizinhança for menor do que o nível de ruído da série temporal, a seleção de vizinhos estará sujeita a uma grande flutuação estatística, causando uma grande flutuação nas predições. Se o tamanho da vizinhança for muito grande, será criado um viés, pois uma aproximação linear para o atrator não será boa [23]. Assim deve haver um comprometimento entre a variância e o viés.

Uma outra forma de se escolher uma vizinhança é através de um número fixo de vizinhos para cada pergunta. Neste caso deve ser observado o comportamento da série temporal no espaço de fase. Devem ser observadas principalmente as regiões com menor número de trajetórias. Desta forma, o tamanho da vizinhança variará localmente. Em regiões onde o número de vizinhos é pequeno, o tamanho da vizinhança deverá ser maior, tornando a predição mais viesada. Mas este procedimento evita o problema de ter vetores-pergunta com poucos vizinhos, ou até mesmo sem vizinhos [41, 32].

Ao se trabalhar com previsão no espaço de fase, deve-se utilizar a informação das trajetórias mais próximas, ao invés de simplesmente utilizar os vizinhos mais próximos. Assim, em cada trajetória próxima, apenas o vizinho mais próximo pertencente àquela trajetória será escolhido [32, 34]. Quando vários pontos de uma mesma trajetória são utilizados na modelagem, pouca informação é acrescentada por estes vizinhos pertencentes à mesma trajetória.

Portanto, na modelagem local serão utilizados os k vizinhos mais próximos, sendo que cada vizinho deve estar numa diferente trajetória. Para garantir que os vizinhos estão em órbitas diferentes, verifica-se o índice de cada vizinho. Se dois (ou mais) vizinhos apresentarem índices consecutivos, então apenas uma deles deve ser considerado como vizinho. Portanto, daqueles que apresentam índices consecutivos, somente o que estiver mais próximo será tomado como vizinho, os demais serão descartados. Desta forma, somente um ponto em cada trajetória, o mais próximo, será considerado na modelagem.

Como somente um vizinho em cada trajetória deve ser escolhido, é importante que este vizinho esteja no ponto da trajetória que mais se aproxima da pergunta. Por este motivo, a interpolação torna-se um procedimento relevante, principalmente quando a taxa de amostragem da série temporal é baixa. Este procedimento de interpolação já foi detalhado na seção 2.5.2.

4.2 Seleção de uma Métrica para a Distância

Para que os vizinhos mais próximos sejam encontrados, é necessário que seja definida uma métrica para a distância. É comum a utilização da métrica euclidiana como a medida de distância entre dois pontos:

$$D_E(\mathbf{q}, \mathbf{x}) \equiv \left(\sum_{i=1}^m (q_i - x_i)^2 \right)^{1/2},$$

onde $\mathbf{q}$ é o vetor pergunta e $\mathbf{x}$ é um ponto da série de dados. Os pontos são m -dimensionais.

Muitas outras medidas poderiam ser usadas. Por exemplo, qualquer p -norma poderia ter sido usada, como definida a seguir

$$L_p(\mathbf{q}, \mathbf{x}) \equiv \left(\sum_{i=1}^m |q_i - x_i|^p \right)^{1/p},$$

onde o parâmetro p da norma é escolhido pelo usuário. Além da norma euclidiana, $L_\infty(\cdot, \cdot)$ e $L_1(\cdot, \cdot)$ também são escolhas comuns.

Muito pouca pesquisa foi feita para determinar que medidas de distância tornam os modelos locais mais exatos, embora muitas diferentes funções de distância tenham sido propostas na literatura [41].

Para garantir o bom comportamento do modelo local, uma metodologia chamada de regularização será adotada. Esta metodologia é baseada na minimização do erro quadrático. Para que haja uma coerência entre o método de regularização e a métrica (para seleção dos vizinhos mais próximos) será utilizada a métrica euclidiana neste trabalho.

4.3 A construção do modelo

Um sistema dinâmico descrito por uma equação diferencial é escrito como:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}).$$

A partir desta equação diferencial, é possível obter um mapeamento, assim:

$$\mathbf{x}(t) = \phi_t(\mathbf{x}_0).$$

Se este mapeamento for reescrito para que a evolução do sistema ocorra simplesmente num intervalo de tempo Δt , obtém-se:

$$x_{\Delta t} = \phi_{\Delta t}(x_0).$$

Agora é possível generalizar esta equação, para que ela represente um mapeamento para a obtenção (de forma iterativa) do estado do sistema em qualquer instante de tempo $n\Delta t$ ($n \in \mathbb{N}$):

$$x_{(n+1)\Delta t} = \phi_{\Delta t}(x_{n\Delta t}).$$

Para simplificar a notação, escreve-se esta equação da seguinte forma:

$$x_{n+1} = f(x_n).$$

Assim $f(\cdot)$ é um mapeamento do espaço de fase nele mesmo, que evolui o sistema por um intervalo de tempo Δt .

Como não são conhecidas as variáveis originais do sistema dinâmico, mas apenas os vetores de atraso, a equação fica:

$$y_{n+1} = g(y_n).$$

Como são utilizadas coordenadas de atraso, não é necessário prever o vetor y_{n+1} . Pode-se simplesmente prever uma coordenada deste vetor, pois, quando se prevê esta componente, é possível criar o vetor y_{n+1} com $m-1$ escalares que já estavam disponíveis mais 1 escalar que foi previsto. Portanto:

$$y_{n+1} = g(y_n).$$

Vale lembrar que a função $g(\cdot)$ não é conhecida, assim a informação sobre esta função será obtida a partir dos k vizinhos mais próximos. Portanto, para encontrar a evolução de um vetor pergunta (q) é necessário recorrer aos exemplos (vizinhos mais próximos). Encontram-se os exemplos mais relevantes no espaço de fase. Pontos que se encontram próximos no espaço de fase possuem uma evolução próxima. Assim, os exemplos mais relevantes são os vizinhos mais próximos. Portanto, a partir dos k vizinhos mais próximos e das suas evoluções será possível estimar a função $g(\cdot)$, que evolui os pontos do espaço de fase.

Para simplificar a notação, será necessário fazer algumas mudanças. Os vetores de entrada da modelagem, assim como os escalares de saída já carregam índices. Será necessário acrescentar mais índices, pois haverá k vizinhos para cada vetor pergunta (q). Para que a notação permaneça simples, os

k pontos vizinhos que serão usados como entrada do modelo serão denotados por $\mathbf{x}_i$ ($i = 1, \dots, k$). Além disso, escalares que são as evoluções dos vetores $\mathbf{x}_i$ serão denotados por y_i ($i = 1, \dots, k$).

A equação que descreve o modelo (para os k vizinhos) pode ser escrita como:

$$y_i = f(\mathbf{x}_i), \quad i = 1, \dots, k.$$

Para o vetor pergunta escreve-se:

$$y = f(\mathbf{q}). \quad (4.1)$$

Como o mapeamento no espaço de fase é uma função contínua e suave, é possível aproximá-lo por uma expansão em série de Taylor. Pode ser feita a aproximação do grau desejado: 0, 1, 2, 3, ...

A equação 4.1 pode ser aproximada por uma expansão em série de Taylor. Esta aproximação será feita em torno do ponto médio da vizinhança do vetor pergunta ($\bar{\mathbf{x}}_q$). Portanto a aproximação de grau zero é dada por:

$$y = f(\mathbf{q}) = f(\bar{\mathbf{x}}_q) + \mathcal{O}(\|\mathbf{q} - \bar{\mathbf{x}}_q\|).$$

Já a aproximação de grau 1 é dada por:

$$y = f(\mathbf{q}) = f(\bar{\mathbf{x}}_q) + \nabla g(\bar{\mathbf{x}}_q) \cdot (\mathbf{q} - \bar{\mathbf{x}}_q) + \mathcal{O}(\|\mathbf{q} - \bar{\mathbf{x}}_q\|^2),$$

onde

$$\bar{\mathbf{x}}_q = \frac{1}{k} \sum_{i=1}^k \mathbf{x}_i.$$

Uma aproximação de grau n ($n > 1$) costuma ser mais próxima da função original, pois utiliza-se até a derivada de ordem n , obtendo assim um polinômio de ordem n . O erro na aproximação é $\mathcal{O}(\|\mathbf{q} - \bar{\mathbf{x}}_q\|^n)$. Quando $\mathbf{q}$ está próximo de $\bar{\mathbf{x}}_q$, o erro deve ser pequeno.

4.3.1 Aproximação de grau zero

A aproximação de grau zero é dada por:

$$y \approx f(\bar{\mathbf{x}}_q).$$

Como a função $f(\cdot)$ não é conhecida, os k vizinhos mais próximos serão usados para se obter uma aproximação para a função. Assim, é definido um estimador para a resposta, sendo este dado por:

$$\hat{y}(\mathbf{q}) = \bar{y}.$$

4.3.2 Aproximação linear

Os modelos lineares são muito comuns devido à sua simplicidade e ao conjunto bem desenvolvido de ferramentas disponíveis da estatística.

Segundo Eckman e Ruelle [15], uma forma de aproximar a dinâmica do sistema que produz a série temporal é com uma aproximação linear que poderia ser determinada a partir dos pontos no espaço de fases por um ajuste de mínimos quadrados. O objetivo deles era de calcular o expoente de Lyapunov. Farmer e Sidorowich [46] foram os primeiros a utilizar esta técnica para a predição de séries temporais não-lineares.

Assume-se que o mapeamento no espaço de fase é uma função suave, ou seja, diferenciável. Assim o mapeamento tangente local será uma aproximação razoável para o mapeamento.

A aproximação linear da equação 4.1 é dada por:

$$y \approx f(\bar{x}_q) + \nabla f(\bar{x}_q) \cdot (q - \bar{x}_q).$$

Como a função $f(\cdot)$ não é conhecida, a aproximação local da função é feita através do conhecimento dos seus vizinhos e suas imagens. Define-se, portanto, um estimador para a resposta:

$$\hat{y}(q) = \alpha \cdot q + \beta$$

onde $\alpha = \nabla f(\bar{x}_q)$ e $\beta = f(\bar{x}_q) - \nabla f(\bar{x}_q) \cdot \bar{x}_q$. Observa-se que α é o vetor de coeficientes angulares do modelo e β é o coeficiente linear.

Para que se determine os parâmetros α e β , é necessário fazer a regressão utilizando os k vizinhos mais próximos. Portanto:

$$\hat{y}_i = \alpha \cdot x_i + \beta, \quad (i = 1, 2, \dots, k).$$

Escrevendo as k equações como apenas uma equação vetorial:

$$\hat{y} = X\alpha + \beta$$

onde X é definida como uma matriz que contém cada um dos k vizinhos mais próximos e $\hat{y}$ é o estimador para as saídas (evoluções) de cada um dos k vizinhos:

$$\hat{y} \equiv \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_k \end{bmatrix}, \quad X \equiv \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_k^T \end{bmatrix} \quad \text{e} \quad \beta \equiv \begin{bmatrix} \beta \\ \beta \\ \vdots \\ \beta \end{bmatrix},$$

onde o ponto vetorial $\mathbf{x}_i$ é o i -ésimo vizinho mais próximo da pergunta $\mathbf{q}$ e y_i é o valor alvo (evolução) para $\mathbf{x}_i$.

Para obter o valor dos parâmetros α e β é necessário estimá-los utilizando alguma metodologia. A mais comum é minimizando o erro quadrático médio.

Para facilitar a obtenção dos parâmetros pode-se realizar a centralização dos dados, assim:

$$\mathbf{A} \equiv \begin{bmatrix} \mathbf{x}_1^T - \bar{\mathbf{x}}_q^T \\ \mathbf{x}_2^T - \bar{\mathbf{x}}_q^T \\ \vdots \\ \mathbf{x}_k^T - \bar{\mathbf{x}}_q^T \end{bmatrix}, \quad \hat{\mathbf{y}} \equiv \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_k - \bar{y} \end{bmatrix},$$

onde $\bar{\mathbf{x}}_q$ é a média dos k vizinhos mais próximos ao vetor $\mathbf{q}$ e $\bar{y}$ é a média das k evoluções.

Realizando a centralização, a equação linear torna-se simplesmente:

$$\hat{\mathbf{y}} = \mathbf{A}\alpha.$$

Assim, a equação para cada um dos vizinhos será escrita como:

$$\hat{y}_i = \bar{y} + \alpha \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_q).$$

Logo, a resposta para a pergunta $\mathbf{q}$ será:

$$\hat{y}(\mathbf{q}) = \bar{y} + \alpha \cdot (\mathbf{q} - \bar{\mathbf{x}}_q).$$

A seguir será mostrada uma estimativa para o vetor de coeficientes α . Para isto será minimizado o erro quadrático médio.

4.4 Tipos de Modelos Locais

Em princípio, qualquer tipo de modelo local poderia ser construído uma vez que os k vizinhos mais próximos foram encontrados. Entretanto, a construção de um modelo não-linear (dados os k pontos) é incomum por, pelo menos, duas razões. Primeiramente, o número de pontos vizinhos (k) é muito pequeno e pode mesmo ser menor do que a dimensão m dos vizinhos. Isto é como resolver um sistema com mais variáveis do que equações. A maioria dos métodos de modelagem não-lineares não são viáveis sob estas circunstâncias

porque requerem que o número de exemplos (vizinhos) seja muito maior que a dimensão do vetor da entrada.

Além disso, o processo de construir o modelo deve ser repetido para cada pergunta. Muitos modelos não-lineares não podem ser usados localmente porque o tempo necessário para construir e avaliar o modelo não-linear seria enorme. Por exemplo, embora as redes neurais possam ser avaliadas rapidamente, requerem geralmente muitos dados de treinamento inicial. No contexto da modelagem local isto teria que ser repetido para cada pergunta e excederia os limites de tempo práticos para a maioria das aplicações.

4.4.1 Modelos Lineares Locais

Pelas razões descritas na seção precedente, é comum a utilização de modelos que podem ser construídos e avaliados muito rapidamente. Os dois tipos mais comuns de modelos locais são [41]:

- modelos de média local;
- modelos lineares locais.

Estes são os dois tipos mais simples de modelos.

Modelos de média local e modelos lineares locais podem ser descritos usando-se a mesma estrutura matemática. Para verificar, considere a saída do modelo linear local para um vetor de entrada q . Este modelo linear para uma dada vizinhança pode ser escrito como

$$\hat{y}(q) = \bar{y} + \alpha \cdot (q - \bar{x}_q),$$

onde $\hat{y}$ é a saída do modelo, $\bar{y}$ é o valor médio da evolução dos k vizinhos, q é o vetor da pergunta, α é o vetor de coeficientes lineares do modelo e $\bar{x}_q$ é a média dos k vizinhos.

Para encontrar o valor dos coeficientes, é minimizada a soma dos erros quadráticos [41]:

$$\begin{aligned} P(\alpha) &\equiv \sum_{i=1}^k (y_i - \hat{y}_i)^2 \\ &= (y - A\alpha)^T (y - A\alpha) \\ &= y^T y - y^T A\alpha - \alpha^T A^T y + \alpha^T A^T A \alpha. \end{aligned}$$

Sendo que a centralização também foi considerada para a variável y , que é dada por:

$$y \equiv \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_k - \bar{y} \end{bmatrix}.$$

Como a soma dos erros quadráticos é quadrática em α , existe somente um extremo que é encontrado igualando o gradiente de $P(\alpha)$ a zero. O gradiente é dado por

$$\nabla_{\alpha} P(\alpha) = -2A^T y + 2A^T A \alpha.$$

Igualando esta equação a zero e resolvendo para α , obtém-se

$$\begin{aligned} \alpha &= (A^T A)^{-1} A^T y \\ &= A^{\dagger} y, \end{aligned} \quad (4.2)$$

onde $A^{\dagger}$ é a pseudo-inversa de Moore-Penrose da matriz A , dada por

$$A^{\dagger} \equiv (A^T A)^{-1} A^T.$$

Se a matriz A for quadrada e com determinante diferente de zero, $A^{\dagger}$ será a inversa da matriz A . Mas, em geral, este não é o caso.

Para mostrar que o extremo encontrado na resolução de $\nabla_{\alpha} P(\cdot) = 0$ é o mínimo global, ao invés de um máximo, deve-se recorrer à matriz Hessiana

$$\nabla_{\alpha}^2 P(\alpha) = 2A^T A,$$

que é positiva definida para qualquer matriz A com colunas independentes. Assim, o valor para α (dado na equação (4.2)) minimiza $P(\cdot)$ globalmente.

4.4.2 Modelos de Média Local

A estrutura desenvolvida na seção precedente pode também ser usada para descrever modelos de média local. Isto é feito eliminando-se as entradas vetoriais (vizinhos) da matriz de A . Neste caso, o vetor de coeficientes (α) é dado por

$$\alpha = 0,$$

Neste caso, a saída estimada, $\hat{y}$, é igual à média amostral da evolução dos vizinhos.

4.4.3 Modelos Quadráticos Locais

Alguns pesquisadores usaram também modelos quadráticos locais, mas há muitas desvantagens deste tipo de modelo, assim é geralmente evitado [41]. Uma desvantagem é que o número de parâmetros aumenta quadraticamente com a dimensão dos pontos no espaço de fase, $\mathcal{O}(m^2)$. Assim, o tamanho da vizinhança deveria também aumentar com $\mathcal{O}(m^2)$. Portanto os modelos quadráticos têm uma performance ruim nas regiões onde há poucos pontos.

4.4.4 Média Local versus Linear Local

Modelos de média local e modelos lineares locais têm seus pontos positivos e negativos, cada um deles apresenta as suas vantagens. Esta seção apresenta algumas das vantagens de cada um destes dois tipos de modelo local.

Complexidade Computacional

O custo computacional da média local é $\mathcal{O}(k)$ onde k é o número de vizinhos usado na construção do modelo. O custo dos modelos lineares locais é $\mathcal{O}(km^2) + \mathcal{O}(m^3)$ [38], que é consideravelmente maior que média local.

O custo destes dois modelos não é muito grande, pois o custo computacional de encontrar os k vizinhos mais próximos é muito maior que o custo da construção do modelo, especialmente para grandes conjuntos de dados.

Sensibilidade à Regularização

O modelo linear é dado por: $\hat{y} = \bar{y} + \alpha \cdot (q - \bar{x}_q)$, onde $\alpha = (A^T A)^{-1} A^T y$. Se o produto matricial $A^T A$ apresentar ao menos um autovalor muito pequeno, então a sua inversa não será muito bem comportada. O produto matricial $A^T A$ apresenta autovalores nulos quando o número de vizinhos é menor que a dimensão dos pontos no espaço de fase. Os autovalores assumem valores pequenos quando os vizinhos são aproximadamente colineares. Quando existem pontos colineares, diminui-se o número de equações, enquanto o número de incógnitas permanece inalterado.

A regularização é um meio de compensar pelo mal-comportamento no produto matricial $A^T A$. Em geral o estimador $\hat{y}$ é muito sensível ao método de regularização [49].

Este problema não ocorre com modelos de média local, que não têm nenhuma entrada local e não requerem nenhuma forma de regularização. Isto também resulta na habilidade dos modelos de média local em usar um único vizinho.

Regularização será discutida em detalhes no capítulo 5.

Estabilidade

Como a média de um conjunto de números não pode ser maior do que o maior número do conjunto nem menor do que o menor número do conjunto, as saídas de modelos de média local têm a garantia de estarem limitadas dentro da escala de valores de alvo dos vizinhos mais próximos.

Os modelos lineares locais não têm nenhuma garantia de tal tipo e podem gerar saídas muito distantes da escala da série de dados. Isto ocorre geralmente quando os vizinhos na série de dados estão muito distantes do ponto da pergunta, que ocorre frequentemente em dimensões elevadas, ou quando o produto matricial $A^T A$ possui autovalores muito pequenos. Estes problemas podem, na maior parte, ser aliviados usando um método eficaz de regularização, mas este não elimina o problema completamente.

Acurácia

Modelos lineares locais geralmente produzem previsões mais acuradas, especialmente quando um método efetivo de regularização é usado e o conjunto de dados contém pouco ruído. Se o conjunto de dados estiver esparsamente distribuído ou contiver muito ruído, modelos de média local serão às vezes mais acurados.

Número de Parâmetros Seleccionados pelo Usuário

Modelos de média local requerem apenas que o usuário selecione o número de vizinhos, k , e a métrica para a distância. Modelos lineares locais requerem que o usuário faça várias escolhas adicionais incluindo o método de regularização, os parâmetros da regularização e quais entradas usar.

4.5 Ponderação local para modelos mais suaves

Modelos Locais são freqüentemente criticados porque sua saída não é uma função suave do vetor de entrada q [18, 36, 45]. Este problema deve-se ao fato de que uma mudança arbitrariamente pequena no vetor pergunta (q) pode causar uma mudança no k -ésimo vizinho mais próximo. Ou seja, um ponto que era vizinho, deixa de ser; e outro ponto que não era vizinho passa a ser. Desta forma, uma pequena variação no vetor pergunta pode causar uma grande variação na resposta.

Esta seção descreve como tornar uma saída de modelos de média local funções suaves das suas variáveis de entrada. Aqui o termo suave significa que a derivada parcial da saída do modelo com respeito a cada entrada do modelo é uma função contínua.

Para que a suavidade na saída do modelo seja obtida, os vizinhos devem ser ponderados de acordo com a sua relevância. Como os pontos mais próximos no espaço de fase produzem resultados mais próximos, os vizinhos devem ser ponderados de acordo com a sua distância. Quanto maior a distância, menor será o peso dado àquele vizinho. Note que a ponderação enfatiza os pontos próximos à pergunta produzindo uma resposta que está mais próxima da evolução dos vizinhos mais próximos, do que no caso não ponderado.

4.5.1 Média Local Ponderada

A modelagem local, tal como descrita na seção anterior, é descontínua. Este problema pode ser resolvido utilizando-se uma função ponderação apropriada que diminui a influência do vizinho mais distante. Por exemplo, considere a média ponderada

$$\hat{y}(q) = \frac{\sum_{i=1}^k y_i w_i^2}{\sum_{i=1}^k w_i^2}, \quad (4.3)$$

onde cada peso* w_i^2 é uma função da distância ao i -ésimo vizinho mais próximo,

$$w_i^2 = W(D_E(q, x_i)).$$

A suavidade de $\hat{y}(q)$ depende apenas da suavidade de $W(\cdot)$ e $D_E(\cdot, \cdot)$. Como a distância euclidiana, $D_E(\cdot, \cdot)$, é uma função suave, a suavidade de $\hat{y}(q)$ depende apenas da função ponderação, $W(\cdot)$.

*Os pesos poderiam ser escritos como w_i , ao invés de w_i^2 . Mas a utilização desta notação, facilitará as contas da próxima seção.

A função ponderação deve ser não-negativa, monotonicamente decrescente e suave [41]. Existem diversas escolhas possíveis para esta função. Uma boa escolha é uma função que torne-se zero apenas quando atinge o vizinho $k + 1$. Assim qualquer função da forma [41]:

$$W(d_i) \equiv \begin{cases} \left(1 - \frac{d_i^n}{d_{k+1}^n}\right)^n & \text{se } d_i \leq d_{k+1} \\ 0 & \text{caso contrário.} \end{cases}$$

seria uma opção. Onde d_i é a distância ao i -ésimo vizinho mais próximo:

$$d_i \equiv D_E(\mathbf{q}, \mathbf{x}_i).$$

Dentre estas funções, foi escolhida [41]:

$$W(d_i) \equiv \begin{cases} \left(1 - \frac{d_i^2}{d_{k+1}^2}\right)^2 & \text{se } d_i \leq d_{k+1} \\ 0 & \text{caso contrário,} \end{cases} \quad (4.4)$$

A função ponderação é suave, pois tem derivada contínua. Além disso, ela depende apenas da razão entre a distância do i -ésimo vizinho e a distância do $(k + 1)$ -ésimo vizinho.

Uma função ponderação suave garante que $\hat{y}(\mathbf{q})$ seja suave exceto no caso onde todos os $k + 1$ dos vizinhos mais próximos são equidistantes. Isto ocorreria se $k + 1$ vizinhos estivessem localizados numa hiper-esfera e o vetor pergunta estivesse localizado no centro desta esfera. Neste caso, a soma no numerador da equação (4.4) seria igual a zero e o valor para $\hat{y}(\mathbf{q})$ neste ponto dependeria da implementação utilizada no algoritmo. Este é, com certeza, um caso patológico e não ocorre na prática.

Embora a média ponderada resulte numa saída contínua para o modelo, ela também sofre de problema de *overfitting* sobre o conjunto de dados mais do que a média não-ponderada. Isto porque os mais próximos dos k vizinhos terão maior influência numa média ponderada e é necessário usar um número maior de vizinhos para se obter o mesmo grau de acurácia [41].

4.5.2 Modelos Lineares Locais Ponderados

O método de ponderação local pode ser generalizado para a estrutura da modelagem linear descrita na seção 4.4.1. Considere o modelo linear que

minimiza a soma ponderada dos erros quadráticos,

$$\begin{aligned}
 P_W(\alpha) &\equiv \sum_{i=1}^k w_i^2 (y_i - \hat{y}_i)^2 \\
 &= (y - A\alpha)^T W^T W (y - A\alpha) \\
 &= y^T W^T W y - \alpha^T A^T W^T W y - y^T W^T W A \alpha + \alpha^T A^T W^T W A \alpha,
 \end{aligned} \tag{4.5}$$

onde $w \in \mathbb{R}^{k,1}$ e $W \in \mathbb{R}^{k,k}$ é a matriz diagonal de w_i ,

$$\begin{aligned}
 W &= \text{diag}(w_i) \\
 &= \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_k \end{bmatrix}.
 \end{aligned}$$

Como W é simétrica, ou seja, $W^T = W$, o produto $W^T W$ é subsequen-
temente denotado por W^2 .

O vetor de coeficientes do modelo linear que minimiza a soma ponderada dos quadrados pode ser encontrado utilizando-se uma versão ponderada da matriz A e do vetor y ,

$$\begin{aligned}
 A_W &\equiv WA, \\
 y_W &\equiv Wy.
 \end{aligned}$$

Após reescrever a equação (4.5) usando estas novas variáveis,

$$P_W(\alpha) = y_W^T y_W - \alpha^T A_W^T y_W - y_W^T A_W \alpha + \alpha^T A_W^T A_W \alpha,$$

torna-se evidente que o problema de regressão linear ponderada é equivalente ao problema não-ponderado resolvido na seção 4.4.1. A solução que minimiza a soma dos erros quadráticos segue da equação (4.2),

$$\alpha = (A_W^T A_W)^{-1} A_W^T y_W = (A^T W^2 A)^{-1} A^T W^2 y. \tag{4.6}$$

Seguindo o mesmo raciocínio que na seção 4.4.2, se os vetores pergunta
forem removidos da matriz A , o erro quadrático seria

$$P(\alpha, \beta) = (y - \beta 1)^T W^T W (y - \beta 1)$$

onde

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Derivando esta expressão em relação a β , e igualando a derivada a zero, obtém-se

$$\beta = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{W} \mathbf{y}.$$

Desta forma, verifica-se que:

$$\hat{y}(\mathbf{q}) = \beta = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{W} \mathbf{y}.$$

Isto é equivalente a média local ponderada descrita na seção anterior (equação (4.3)).

Capítulo 5

Regularização

Uma modelagem linear local precisa é difícil de ser criada, pois ela deve ser construída com uma pequena vizinhança, com poucos vizinhos. Muitas vezes o número de pontos é aproximadamente igual ao número de graus de liberdade, ou até menor. O que torna difícil estimar os coeficientes lineares com acurácia. O problema aumenta quando existem pontos colineares, pois diminui-se o número de equações, enquanto o número de incógnitas permanece inalterado. Para aliviar este problema existem diferentes métodos de regularização, que enviam os coeficientes do modelo reduzindo a variância da saída, aumentando assim a estabilidade do preditor [41].

5.1 Introdução

No capítulo anterior, foi mostrado que o modelo linear local é dado por:

$$\hat{y} = A\alpha,$$

onde a matriz $A \in \mathbb{R}^{k \times m}$ contém os k vizinhos mais próximos depois de centralizados, e α é o vetor de coeficientes angulares. A matriz A é dada por:

$$A \equiv \begin{bmatrix} \mathbf{x}_1^T - \bar{\mathbf{x}}_q \\ \mathbf{x}_2^T - \bar{\mathbf{x}}_q \\ \vdots \\ \mathbf{x}_k^T - \bar{\mathbf{x}}_q \end{bmatrix}.$$

Se a matriz A for quadrada (e com determinante diferente de zero), então o vetor α será dado por:

$$\alpha = A^{-1}y.$$

onde $y \in \mathbb{R}^k$ é o vetor que contém as evoluções centralizadas dos k vizinhos mais próximos, ou seja,

$$y \equiv \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_k - \bar{y} \end{bmatrix}.$$

É importante notar que, se a matriz A não possuir inversa, a solução não será aquela apresentada acima. A solução será aquela que minimizará o erro quadrático:

$$\alpha = A^\dagger y.$$

onde $A^\dagger$ é a pseudo-inversa de Moore-Penrose da matriz A , dada por $A^\dagger \equiv (A^T A)^{-1} A^T$.

Um dos problemas que podem ocorrer é quando as colunas da matriz A são aproximadamente colineares, então o produto matricial $A^T A$ será aproximadamente singular e alguns dos coeficientes do modelo serão muito grandes, pois são proporcionais à inversa de $A^T A$. Como os coeficientes serão muito grandes, farão com que a saída do modelo divirja bastante. Isto não ocorre se o vetor de entrada do modelo, q , estiver no espaço das colunas da matriz A , mas isto raramente ocorre.

O método de regularização substitui o produto matricial $A^T A$ por uma matriz que é melhor comportada. Apesar deste procedimento resultar numa solução viesada, ele também reduz a variância da saída.

O método de regularização utilizado neste trabalho é chamado de Análise das Componentes Principais ou Regressão em Componentes Principais e baseia-se numa técnica de decomposição de matrizes chamada de Decomposição em Valor Singular. Primeiramente será detalhado o funcionamento desta decomposição, para que em seguida seja explicado como esta decomposição pode ser utilizada para a obtenção de uma matriz bem comportada através da Análise das Componentes Principais [39, 10].

5.2 Decomposição em Valor Singular

A Decomposição em Valor Singular é uma técnica para encontrar a pseudo-inversa de uma matriz singular, ou *numericamente quase singular**. Para conhecer o procedimento detalhadamente veja Van-Loan e Golub[38].

A decomposição em valor singular é baseada num teorema da Álgebra Linear que diz que qualquer matriz A de ordem $M \times N$ pode ser escrita como o produto de uma matriz ortogonal $M \times M$ U , uma matriz diagonal $N \times N$ Σ e a transposta de uma matriz ortogonal $N \times N$ V [38].

$$\begin{pmatrix} A \end{pmatrix} = \begin{pmatrix} U \end{pmatrix} \cdot \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \dots & \\ & & & \sigma_N \end{pmatrix} \cdot \begin{pmatrix} V^T \end{pmatrix} \quad (5.1)$$

As matrizes U e V são ortogonais, pois suas colunas são ortonormais:

$$\sum_{i=1}^M U_{ik} U_{in} = \delta_{kn} \quad \begin{matrix} 1 \leq k \leq N \\ 1 \leq n \leq N \end{matrix} \quad (5.2)$$

$$\sum_{j=1}^N V_{jk} V_{jn} = \delta_{kn} \quad \begin{matrix} 1 \leq k \leq N \\ 1 \leq n \leq N \end{matrix}$$

Cada uma das quantidades σ_j é chamada de *valor singular* e cada coluna de U e V é conhecida como *vetor singular* à esquerda e à direita, respectivamente.

Note-se que diferentemente da matriz U , a matriz V é quadrada. Portanto, além de colunas ortonormais, ela também possui linhas ortonormais, $V \cdot V^T = 1$.

*Por *numericamente quase singular* entende-se uma matriz que apresenta um dos autovalores aproximadamente igual a zero, esta diferença sendo muito próxima da precisão computacional.

A decomposição em valor singular pode ser feita para M igual a N , ou diferente. Quando $N > M$, os valores singulares σ_j (para $j > M$) serão todos nulos [38].

A equação (5.1) sempre existe, não importando o quão singular é a matriz, e é única. Podem ser feitas permutações simultâneas das colunas de U , dos elementos de Σ e das colunas de V . Para que a decomposição seja única foi adotado o procedimento (comumente usado) de ordenar de forma decrescente os elementos da matriz Σ , ou seja, a decomposição foi feita de tal forma que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$.

5.2.1 Conceitos de Álgebra Linear

Antes de detalhar cada uma destas matrizes, é necessário fazer uma sucinta revisão de alguns termos da Álgebra Linear, que são de suma importância no entendimento deste método.

Considerando a seguinte equação:

$$Ax = b,$$

onde a é uma matriz $M \times N$ que leva o vetor x à sua imagem b . O subespaço dos vetores x tais que $Ax = 0$ é chamado de *núcleo* (ou *kernel*) da matriz A . O número de vetores x linearmente independentes que podem ser encontrados neste subespaço é a *dimensão do núcleo* da matriz A [38].

Os vetores b para os quais existe um x que satisfaça $Ax = b$ constituem-se num espaço vetorial chamado de imagem de A . A dimensão deste espaço é o posto de A . Um teorema da Álgebra Linear afirma que *posto mais dimensão do kernel é igual a $\max(M, N)$* . Observa-se assim que a imagem da matriz A terá obrigatoriamente dimensão menor ou igual a dimensão do vetor x [38].

O que a decomposição em valor singular faz é construir uma base para o kernel e outra para a imagem da matriz. As colunas de U , cujos elementos σ_j são não-nulos, são um conjunto de vetores ortonormais que formam a base da imagem. As colunas de V , cujos elementos σ_j são nulos, formam uma base ortonormal para o kernel. Abaixo segue a demonstração.

Sejam u_j e v_j as colunas de U e V respectivamente. Então a decomposição $A = U\Sigma V^T$ pode ser escrita como:

$$Av_j = \sigma_j u_j.$$

Se $\sigma_j = 0$, então $Av_j = 0$ e v_j está no kernel de A . Entretanto, se $\sigma_j \neq 0$, então u_j está na imagem de A . Desta forma são construídas bases para o kernel e para a imagem [38].

Ao se trabalhar com um sistema sujeito a ruído, geralmente não se obtém $\sigma_j = 0$, mas apenas $\sigma_j \approx \eta$, onde η é o ruído. Desta forma aquilo que foi chamado de kernel, passa a ser chamado de subespaço do ruído. Assim as colunas de V formam uma base para este subespaço.

Quando $\sigma_j > \eta$, o subespaço que havia sido chamado de imagem da matriz A passa a se chamar de subespaço do sinal. As colunas de U formam uma base para este subespaço.

Para se resolver o sistema, é necessário saber se o vetor b está na imagem da matriz A . Se a resposta for afirmativa, então existe uma solução x para a equação $Ax = b$. De fato, existe mais de uma solução quando existem valores singulares nulos, pois qualquer vetor no espaço nulo pode ser acrescentado ao vetor x numa combinação linear [38].

Se b não estiver na imagem da matriz A , então a equação $Ax = b$ não tem solução. Entretanto, pode-se ainda assim obter uma solução não-exata, a mais próxima o possível no sentido de mínimos quadrados.

Escolhe-se o vetor x com menor módulo. Para tal, faz-se $1/\sigma_j = 0$, sempre que σ_j for nulo, ou da ordem de grandeza do ruído presente nos dados. Desta forma a variância do modelo é reduzida, pois a matriz $(A^T A)^{-1}$ passa a ser bem comportada, fazendo com que o vetor α dos coeficientes do modelo assumam o menor valor possível.

A primeira vista pode parecer estranho que, descartando uma combinação linear do conjunto de equações que se pretende resolver, obtém-se uma solução ótima. Entretanto deve ser lembrado que está se desconsiderando justamente equações que estão tão corrompidas pelo ruído que acabam por desviar a solução obtida na direção de um vetor do subespaço do ruído.

5.3 Regressão em Componentes Principais

Suponha que o posto da matriz A seja m , e que se queira aproximá-la por uma matriz B de posto l ($l < m$). Então a melhor aproximação para a matriz B é dada por:

$$B = \sum_{i=1}^l \sigma_i u_i v_i^T,$$

ou seja, apenas as direções de maior variância foram consideradas. Somente os l vetores singulares associados aos maiores valores singulares estão presentes. É importante lembrar que os valores singulares estão organizados assim: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_l \geq \dots \geq \sigma_m \geq 0$. Além disso, $\sigma_i = 0, \forall i > m$.

Conforme foi visto anteriormente, é necessário fazer um processo de regularização para que o vetor α seja bem comportado. Este processo consiste em zerar os valores singulares próximo ao nível de ruído. Agora foi mostrado que a matriz de posto l que melhor aproxima a matriz A é a matriz onde os valores-singulares de índice superior a l são todos nulos, ou seja:

$$\sigma_j = 0, \quad \text{para } j > l.$$

Se os valores singulares superiores a l forem da ordem de grandeza do ruído presentes nos dados, as direções associadas a estes autovalores devem ser desprezadas, pois estão no subespaço do ruído.

No modelo linear local, o vetor dos coeficientes é dado por:

$$\alpha = A^T y = (A^T A)^{-1} A^T y = V \Sigma^{-1} U^T y.$$

Assim, se os valores singulares forem muito pequenos, este resultado será mal comportado, pois a matriz Σ^{-1} contém na sua diagonal principal os elementos $\frac{1}{\sigma_i}$. Portanto, para que a variância do modelo seja reduzida, os valores singulares pequenos (da ordem de grandeza do ruído) devem ser desprezados. Para isto, o termo $\frac{1}{\sigma_i}$ deve ser zerado sempre que σ_i for da ordem de grandeza do ruído.

A decomposição em valores singulares deve ser usada pois o sistema que produziu os dados apresenta um número menor de dimensões do que o número de linhas e o número de colunas da matriz A . Cada matriz A , definida numa determinada vizinhança, tem de ter posto menor do que o número de vizinhos (k) e que o número de dimensões (m) do ponto no espaço de fase, desde que o número de vizinhos tomados não seja muito pequeno. Pois o sistema dinâmico que produz a série temporal deve possuir um número pequeno de dimensões.

5.4 Ponderação

Regressão em componentes principais é um método para lidar com o mal-comportamento do produto matricial, $A^T A$. Esta técnica apóia-se na decomposição em valor singular da matriz A . Para que seja possível obter um

resultado mais geral, pode-se decompor a matriz de vizinhos ponderados pela distância $A_W \equiv WA$. Assim, a decomposição de A_W é dada por

$$A_W = U\Sigma V^T,$$

onde $U \in \mathbb{R}^{k \times m}$ e $V \in \mathbb{R}^{m \times m}$ são ortonormais, $U^T U = I$ e $V^T V = V V^T = I$, e Σ é uma matriz diagonal e positiva semi-definida, $\Sigma = \text{diag}(\sigma)$ onde $\sigma_i \geq \sigma_{i+1} \geq 0$ para todo i . Os elementos do vetor σ são chamados de valores singulares da matriz A_W . Usando esta decomposição, a solução de mínimos quadrados ponderados pode ser escrita como

$$\begin{aligned}\alpha &= (A_W^T A_W)^{-1} A_W^T y_W, \\ &= V \Sigma^{-1} U^T y_W.\end{aligned}$$

A notação usada aqui é a mesma da apresentada na seção 4.5.2. A saída do modelo linear para um vetor pergunta q é dado por,

$$\begin{aligned}\hat{y}(q) &= \bar{y} + (q - \bar{x})^T \alpha \\ &= \bar{y} + (q - \bar{x})^T V \Sigma^{-1} U^T y_W \\ &= \bar{y} + \sum_{i=1}^{\min(k,m)} ((q - \bar{x})^T v_i) \left(\frac{1}{\sigma_i} \right) (u_i^T y_W),\end{aligned}$$

onde v_i denota a i -ésima coluna de V , e u_i denota a i -ésima coluna de U . Os parênteses foram introduzidos para enfatizar que cada elemento da soma é um produto de três escalares.

5.4.1 Reduzindo a variância do modelo

Se o número de vizinhos for pequeno, ou se eles estiverem ocupando apenas um subespaço do espaço considerado, um ou mais dos valores singulares, σ_i , será aproximadamente zero. Como no modelo local (que encontra a evolução dos pontos) cada elemento da soma é multiplicado por $\frac{1}{\sigma_i}$, os elementos com valor singular pequeno podem fazer com que a saída do modelo seja extremamente alta; ou que, equivalentemente, a variância do modelo cresça. Para conter este efeito, o procedimento freqüentemente adotado é o de truncar a soma,

$$\hat{y}(q) = \bar{y} + \sum_{i=1}^l ((q - \bar{x})^T v_i) \left(\frac{1}{\sigma_i} \right) (u_i^T y_W),$$

onde $l < \min(k, m)$, onde k é o número de vizinhos e m é a dimensão dos pontos no espaço de fase. Esta técnica, chamada de regressão em componentes principais, tem o efeito de suavizar a saída do modelo.

É importante observar que os maiores valores singulares (σ_i) estão associados às direções (no espaço de fase) que possuem maior variância, enquanto que os menores valores singulares estão associados às direções (no espaço de fase) que possuem menor variância. Assim os maiores valores singulares estão associados ao sinal, enquanto que os menores valores singulares estão associados ao ruído.

O número de componentes no somatório, l , é um parâmetro definido pelo usuário. Note que se o usuário especificar $l = 0$, a saída do modelo é igual à média amostral dos valores alvo para os k pontos vizinhos armazenados no vetor y , ou seja, $\hat{y}(q) = \bar{y}$. No outro extremo, onde $l = \min(k, m)$, a saída torna-se equivalente a solução de mínimos quadrados dada por $\hat{y}(q) = (q - \bar{x})^T (A^T A)^{-1} A^T y$. Isto representa um compromisso entre usar média ponderada ($l = 0$) e mínimos quadrados ($l = \min(k, m)$).

5.5 Interpretação Qualitativa

A decomposição em valor singular possibilita a representação dos dados com um menor número de variáveis, e a detecção de padrões em dados ruidosos. Esta decomposição produz duas bases ortonormais, uma definida pelos vetores singulares à esquerda e outra definida pelos vetores singulares à direita. Apesar da redução da dimensionalidade produzida pela decomposição, é necessário acessar os auto-vizinhos que são vetores m -dimensionais. Entretanto o número de auto-vizinhos necessários para a descrição local é igual a l , que é o posto da matriz A depois de desprezado o ruído.

Como a matriz A foi centralizada, ou seja, suas linhas foram centralizadas, o produto matricial $A^T A = \sum_i a_i a_i^T$ é proporcional à matriz de covariância das variáveis x_i (a matriz de covariância dos vizinhos). O produto matricial $A^T A$ é igual a $V \Sigma^2 V^T$. Assim, os auto-vizinhos correspondem às direções de principal variância dos vizinhos, enquanto que o quadrado dos valores singulares (σ_i^2) correspondem às variâncias das direções principais. A magnitude relativa de cada valor singular é um indicativo da sua importância na explicação dos dados.

Os pontos do espaço de fases que são vizinhos da pergunta q formam um plano (ou hiperplano) de dimensão l . Como foram escolhidos k vizinhos, e cada um deles com m coordenadas ($m, k > l$), os vetores auto-vizinhos v_i

($i = 1, \dots, k$) devem formar uma base para este plano (hiperplano). Cada auto-vizinho v_i é uma coluna da matriz V .

Devido ao ruído, não se obtém $A\alpha = 0$, ou seja, a dimensão do kernel é nula. Mas é possível obter $A\alpha = \eta$, onde η é o ruído. Os vetores x que satisfazem esta equação estão no subespaço do ruído. As colunas da matriz U formam uma base para este subespaço.

A imagem da matriz A costuma ser chamada de *subespaço do sinal*. As colunas da matriz V (os vetores v_i que podem ser chamados de auto-vizinhos) formam uma base para este subespaço.

É importante ressaltar que o número de dimensões de um sistema está sempre relacionado ao número de graus de liberdade do mesmo.

5.6 Regularização Aplicada a Sistemas Dinâmicos

A projeção sobre a direção dominante é uma projeção sobre os vetores que representam a maior variância dos dados, correspondendo assim a um menor nível de ruído.

O número de autovetores significativos espelha o número de dimensões. Isto implica no conhecimento do nível de ruído. Todos os autovalores inferiores ao ruído refletem direções ruidosas.

Deve-se lembrar que a *variedade* que contém o atrator não é, em geral, um espaço linear. Uma hiper-superfície pode ser deformada, de modo a apresentar um número maior de dimensões quando escrita numa base linear. Assim a regressão em componentes principais, quando aplicada ao atrator, pode dar resultados não-ótimos.

Assim, ao invés de aplicar a regressão em componentes principais globalmente, deve-se implementá-la localmente. Pois localmente o atrator pode ser projetado numa base cartesiana, com um erro desprezível, desde que a vizinhança seja suficientemente pequena, para que a curvatura da variedade possa ser desprezada.

É difícil determinar o quão pequena deve ser uma vizinhança. Para um dado atrator, o tamanho ideal da vizinhança pode variar, pois, em algumas regiões, o atrator pode apresentar uma curvatura maior, enquanto que em outras, o atrator pode ser mais linear.

Entretanto, a vizinhança não pode ser muito pequena, pois isto levaria a obter um número muito pequeno de vizinhos. Desta forma, o resultado obtido seria viesado. Viesado por ter poucos vizinhos, e todos eles com ruído. Para se reduzir o ruído, é preciso aumentar o tamanho da vizinhança e aproximar uma variedade não-linear por uma linear, o que conduz a um erro sistemático.

Neste ponto é importante lembrar como é feita a reamostragem de uma série temporal. Em suma, para obter a previsão de uma série temporal ou a sua reamostragem, é necessário realizar a evolução de um ponto do espaço de fase. Neste trabalho, para que os pontos sejam evoluídos foi utilizada uma modelagem local, que tem apresentado melhores resultados do que a modelagem global.

Para se utilizar o modelo local, deve ser usado um método de regularização, que reduz a variância na sua saída. A regularização utilizada projeta os pontos de uma região do espaço de fase num subespaço de menor dimensão.

Para que se reduza o ruído presente na série temporal, é criado um espaço com dimensão superior à dimensão que seria necessária ao espaço de fase reconstruído. Posteriormente, estes vetores são projetados num subespaço de dimensão menor, desprezando as dimensões associadas ao ruído.

5.7 A escolha da dimensão do espaço da projeção

Uma forma de se analisar a significância de cada componente e decidir quais delas serão utilizadas na regularização é através da variância relativa acumulada. Ou seja, esta proposta requer que o modelo regularizado explique uma fração da variância dos dados originais. Se fosse escolhido 90%, escolher-se-ia l que satisfizesse

$$\frac{\sum_{i=1}^l \sigma_i}{\sum_{i=1}^{\min(m,k)} \sigma_i} \geq 0.90.$$

Assim, através da regressão em componentes principais, é possível encontrar a verdadeira dimensão da variedade, que é o posto l da matriz. Se $l < k$, a análise de componentes principais produz uma representação dos vizinhos utilizando um menor número de variáveis.

Uma outra possibilidade é a escolha de um valor fixo para os valores singulares não nulos. Por exemplo, o atrator de Lorenz pode ser representado localmente num plano com duas dimensões. Portanto, todos os vizinhos mais próximos de uma pergunta estariam neste plano. Entretanto, deve-se escolher apenas um vizinho em cada trajetória. Isto faz com que a dimensionalidade seja reduzida em uma unidade. Portanto todos os vizinhos estarão sobre uma linha unidimensional, observe isto na figura 2.6. Nesta figura vemos que foi utilizada a interpolação para garantir que os vizinhos mais próximos estivessem sobre uma linha com dimensão 1.

Sauer [32] utilizou esta característica do atrator no seu algoritmo de predição. Portanto os vizinhos eram projetados sobre a direção de maior variância, sendo esta a direção associada ao maior valor singular da decomposição. Portanto o parâmetro l foi tomado igual a 1, porque todos os vizinhos estavam aproximadamente distribuídos ao longo de uma linha. A dispersão que ocorria na distribuição dos pontos ao longo desta linha foi considerado como ruído.

O mesmo procedimento que aquele adotado por Sauer [32] foi feito aqui. Portanto uma primeira filtragem foi realizada por meio da inersão filtrada, quando uma janela de tamanho w foi tomada e foram excluídas as frequências mais altas para se obter um vetor de tamanho m . Depois foi realizada uma segunda filtragem, quando os vetores m -dimensionais foram projetados num subespaço de dimensão menor.

Conforme foi afirmado anteriormente, a janela pode apresentar tamanho superior àquela necessária para a inersão no espaço de fase, pois, devido à filtragem por Fourier, a dimensão será reduzida. Agora ocorre o mesmo, os pontos do espaço de fase serão projetados num subespaço de dimensão menor. Assim a dimensão do espaço de fase pode ser maior do que a que seria realmente necessária. Se a série não possuísse ruído, apenas informação redundante seria acrescentada. No caso de uma série com ruído, algumas dimensões apresentarão uma dispersão simplesmente devido ao ruído. Portanto a informação adicional que está no espaço de fase será utilizada na filtragem por meio da projeção no subespaço do sinal.

Capítulo 6

Reamostragem

6.1 Introdução

Ao se medir algum invariante de um sistema (tal como expoente de Lyapunov, dimensão de capacidade, entre outros) é importante dispor de uma medida de incerteza associada ao valor obtido. Quando se obtém, por exemplo, um expoente de Lyapunov de 0,01, é imprescindível saber se este valor é estatisticamente diferente de zero, para poder afirmar se o sistema sob análise é caótico. Isto é possível somente quando se tem uma medida de incerteza associada ao valor calculado.

A maioria dos trabalhos publicados (que extraem informação a partir de séries temporais determinísticas não-lineares) não apresenta uma incerteza para os valores calculados [2, 3, 4]. Isto não é um erro específico destes trabalhos, mas sim um problema que ocorre com a teoria de sistemas caóticos, que possui alguns estimadores para as quantidades invariantes (expoente de Lyapunov, dimensão, etc), mas não possui um estimador para a incerteza das medidas obtidas.

Um grande problema para estimar a incerteza, é o fato de haver apenas uma realização do sistema a ser avaliado. De uma forma geral, dispõe-se de apenas uma série temporal. Assim consegue-se apenas um valor para a quantidade que está sendo medida, sem que se tenha idéia da sua incerteza.

Se estivessem disponíveis várias realizações do mesmo sistema, ou seja, se várias séries temporais tivessem sido geradas pelo mesmo sistema, seria possível calcular os invariantes para cada uma das séries. A partir de então, ter-se-ia uma distribuição para este invariante, sendo possível calcular a sua média e a sua incerteza (desvio-padrão, por exemplo).

Para tentar contornar este problema, algumas metodologias têm sido desenvolvidas, mas algumas delas já foram contestadas, sendo mostrado que seus resultados não são apropriados.

De uma forma geral, estas metodologias estão baseadas em algum tipo de reamostragem, uma forma de gerar novas bases de dados a partir de uma base original. Reamostragem é uma técnica computacional utilizada para, a partir de uma amostra, gerar novas amostras. A reamostragem baseia-se na simulação, sendo esta simulação baseada na amostra, e não numa população hipotética.

No processo de reamostragem, são criadas várias amostras a partir da original. Depois que as amostras foram criadas, o processo é bastante simples. Deve-se calcular a estimativa de interesse para cada uma das amostras geradas artificialmente. Neste momento, dispõe-se de uma distribuição para a estimativa de interesse, tornando possível o cálculo de um valor médio e intervalos de confiança. Um teste de hipótese pode ser realizado, além do teste de normalidade para a distribuição. Várias estatísticas podem ser calculadas, pois a distribuição está disponível.

Dois tipos muito comuns de reamostragem para lidar com bases de dados são o *jackknife* e o *bootstrap*. No *jackknife*, uma nova amostra é criada simplesmente deixando de fora um elemento da amostra original. Assim, para uma amostra com n dados, é possível criar n novas amostras, para cada uma delas deixa-se de fora um dos n dados originais [?].

No *bootstrap*, os dados para a nova amostra também são obtidos da amostra original, mas agora é feita amostragem com reposição. Seleciona-se, portanto, um dado do conjunto de dados original, mas este dado não é removido desta amostra, podendo ser selecionado novamente. De forma aleatória, seleciona-se dados da amostra original, formando uma nova base de dados [?].

No método *jackknife*, o número máximo de amostras que podem ser criadas é n . Já no *bootstrap* não ocorre este limite. A única preocupação poderia ser com relação ao gerador de números aleatórios, pois, quando for atingido o seu limite máximo, poderiam ser geradas amostras idênticas às que já foram criadas. Mas isto não costuma ser um problema, pois os geradores de números aleatórios atualmente são capazes de gerar uma sequência de 2^{32} (4 294 967 296) números distintos.

No contexto de sistemas dinâmicos, Gençay [5] desenvolveu uma metodologia para obter uma distribuição de valores para o expoente de Lyapunov máximo. Para entender esta metodologia, é preciso saber como este expoente é

calculado. Dado um sistema dinâmico, descrito por:

$$y_{n+1} = g(y_n) = g^{n+1}(y_0),$$

a sua versão linearizada é dada por:

$$y_{n+1} = J_{y_n} J_{y_{n-1}} \cdots J_{y_1} J_{y_0}(y_0),$$

onde J_{y_i} é a matriz jacobiana no ponto y_i . O expoente de Lyapunov máximo será calculado a partir do maior autovalor da matriz:

$$J_{y_n} J_{y_{n-1}} \cdots J_{y_1} J_{y_0}.$$

É importante notar que a ordem nas matrizes é importante, pois o produto matricial não comuta, exceto no caso unidimensional.

Gençay [5] propôs selecionar aleatoriamente os pontos do espaço de fase, e a matriz jacobiana é calculada para cada um destes pontos. Depois disso as matrizes são multiplicadas, para que se possa encontrar o expoente de Lyapunov máximo.

No artigo de Ziehmann et al. [51] foi mostrado que o algoritmo de Gençay [5] produziu resultados inadequados. Eles afirmam que um dos problemas é que o produto matricial não comuta, fazendo com que o processo de aleatorização não seja apropriado. Outro problema é que os dados são dependentes, portanto a sua ordem não deve ser alterada. Neste artigo, também é mostrado um caso em que a distribuição obtida é diferente da verdadeira distribuição.

Uma outra possibilidade no processo de reamostragem, consiste em criar outras séries temporais, de maneira artificial, que possuam as mesmas propriedades que a série original. Mas o problema que surge é como realizar este procedimento.

6.1.1 Reamostragem em séries temporais

A reamostragem de séries temporais é mais difícil do que a reamostragem de uma base de dados que não possui correlação temporal. Isto ocorre porque, se for utilizado um método tradicional de reamostragem, toda a informação sequencial é perdida. Várias alternativas surgiram para lidar com este problema. Entre os métodos existentes, um deles consiste em selecionar janelas e realizar a reamostragem dentro de janelas. Outro consiste em realizar um ajuste e subtrair este ajuste da série temporal. A reamostragem é feita usando os resíduos do ajuste como amostra. Depois que uma nova amostra é criada, a função ajustada aos dados originais é acrescentada à reamostragem.

Uma forma, que utiliza com maior precisão a informação temporal presente na série, é feita através de uma interpolação da série original. A partir de uma série temporal com intervalo de amostragem Δt , cria-se uma série interpolada com intervalo de amostragem δt , para isto utiliza-se um fator de interpolação:

$$k = \frac{\delta t}{\Delta t}.$$

Entre o ponto 1 e o ponto 2 da série temporal original, haverá $k - 1$ novos pontos criados pela interpolação. Um destes pontos é selecionado aleatoriamente para se gerar uma nova série. A partir de então, é incluído na nova série cada um dos pontos que se encontra a um intervalo Δt na série interpolada. Este método foi implementado em sistemas dinâmicos por Giannerini e Rosa [54, 55, 56].

Apesar deste método utilizar melhor as propriedades temporais presentes na série, ele pressupõe que todas as evoluções da série terão trajetórias semelhantes. Quando a trajetória passar próximo a um ponto de instabilidade, a sua evolução será a mesma que na série temporal original, para o respectivo instante de tempo. Assim, este procedimento não parece ser adequado.

O primeiro trabalho a utilizar a informação do espaço de fases para realizar a reamostragem da série temporal foi realizado por Lall e Sharma [57] em 1996. Neste trabalho, são utilizados os vizinhos mais próximos. Para cada ponto no espaço de fase, são encontrados os seus vizinhos mais próximos. A evolução que se busca será igual a evolução de um dos vizinhos. É atribuída uma probabilidade a cada um dos vizinhos e, através de uma escolha aleatória, a evolução que se busca será igual a evolução daquele vizinho que foi escolhido. Esta metodologia parece funcionar bem com séries hidrológicas.

Em séries temporais caóticas, o uso de vizinhos mais próximos no espaço de fase foi realizado por Golub e Sandri [6, 7]. A metodologia desenvolvida por estes pesquisadores é a que mais se aproxima do trabalho aqui desenvolvido. Para recriar uma série temporal, primeiramente um ponto do espaço de fase original é escolhido aleatoriamente. Para evoluir este ponto é utilizada a informação dos vizinhos mais próximos. Todos os vizinhos mais próximos, assim como a evolução de cada um deles, são usados como treinamento de uma rede neural.

A grande diferença entre o trabalho feito por Golub e Sandri [6, 7] e aquele que é desenvolvido aqui é a metodologia utilizada para fazer a aproximação de funções. Ao invés da utilização de uma rede neural, aqui é utilizada a aproximação linear local. Para comparação, vamos verificar o número de vizinhos

necessários para treinar a rede neural e o número de vizinhos necessários para a aproximação linear local. Com redes neurais, foram necessários 80 vizinhos para realizar a reamostragem da série temporal gerada pelo sistema de equações diferenciais de Lorenz [7]. Utilizando a abordagem aqui desenvolvida, foram necessários somente os 13 vizinhos mais próximos. Além disso, a minimização do erro é iterativa na rede neural, enquanto que existe uma forma fechada quando se utiliza aproximações lineares.

Devido à grande diferença no número de vizinhos necessários para fazer a aproximação local, assim como devido à forma como o erro é minimizado, o modelo aqui desenvolvido, possui um custo computacional muito menor que aquele que usa a rede neural. Portanto, usando o modelo linear, o tempo necessário para a reamostragem é muito menor.

6.2 Sombreamento

Uma propriedade muito importante a ser considerada no estudo da reamostragem de séries temporais caóticas é o sombreamento. Esta propriedade será descrita nesta seção.

A evolução de uma sistema dinâmico é sensível à condição inicial. Portanto, o erro que é introduzido num estado faz com que a órbita seja totalmente alterada depois de um intervalo de tempo não muito grande, dado pelo maior expoente de Lyapunov.

Quando uma trajetória é evoluída num computador, a incerteza devido à precisão finita do computador é propagada. Isto faz com que a trajetória não seja a trajetória real. Assim ocorre aquilo que é chamado de *ruído dinâmico*. Existem dois tipos de ruído, o *ruído de medição* e o *ruído dinâmico*.

Para definir o ruído de medição, vamos considerar que a dinâmica é dada por:

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n).$$

Assim a variável, tal como medida pelo observador, será dada por:

$$\mathbf{y}_i = \mathbf{x}_i + \boldsymbol{\eta}_i,$$

onde $\boldsymbol{\eta}_i$ é o vetor ruído.

O ruído dinâmico modifica a evolução da seguinte forma:

$$\mathbf{x}_{n+1} = f(\mathbf{x}_n) + \boldsymbol{\eta}_i$$

Observa-se que, no caso do ruído dinâmico, o estado do sistema será evoluído a partir do estado com ruído. Isto faz com que a trajetória se afaste cada vez mais da trajetória real.

Para analisar o que pode acontecer devido ao ruído dinâmico, considere um exemplo simples utilizando o mapa logístico:

$$x_{n+1} = 4x_n(1 - x_n).$$

Considere uma pseudo-órbita que começa com o ponto $x_0 = 0,5$.

Assumindo que seja cometido um erro muito pequeno e o próximo ponto obtido seja $x_1 = 1 + \delta$. A partir de então a iteração é calculada sem erro. O próximo ponto da série será $x_2 = 4(1 + \delta)(-\delta) < 0$, então esta pseudo-órbita diverge para $-\infty$, e não existe nenhuma órbita verdadeira que esteja próxima desta pseudo-órbita.

Devido ao ruído dinâmico, a predição de uma série temporal caótica (por um longo período de tempo) é inviável. A previsão pode ser feita somente para um intervalo de tempo pequeno. Entretanto, como será discutido, pode ser possível (a partir da evolução de um estado) encontrar uma trajetória que esteja próxima de uma trajetória real do sistema. Para que isto ocorra, é necessário que o sistema possua a propriedade chamada de *sombreamento* [52, 53].

De acordo com a teoria do sombreamento, embora o verdadeiro sistema não dê origem a série que dispomos (quando é iniciada do mesmo ponto), existe uma trajetória real do sistema (começando de um estado inicial ligeiramente perturbado) que permanece próxima ao modelo, por um longo período de tempo. Esta propriedade é de grande importância na reamostragem de uma série temporal.

Para utilizar um algoritmo de previsão, deve-se conhecer as propriedades de sombreamento, para que assim seja possível encontrar uma trajetória que esteja próxima da trajetória real. Ao se realizar a reamostragem de uma série temporal, uma nova realização desta série é feita. Esta série reamostrada deve possuir as mesmas propriedades dinâmicas da série temporal real. Se o algoritmo que faz a evolução da série temporal não for capaz de lidar com o ruído presente nas iterações, é possível que a série reamostrada seja totalmente diferente da série original.

Sistemas hiperbólicos*, ou quase hiperbólicos, são localmente sensíveis às condições iniciais, mas globalmente insensíveis. Assim, existem trajetórias

*Um sistema é hiperbólico quando cada ponto do espaço de fase pode ser fatorado em direções onde o movimento está expandindo exponencialmente, ou contraindo exponencialmente [53].

(chamadas de trajetórias sombreadas) muito próximas à pseudo-trajetória gerada pelo computador [52].

Farmer e Sidorowich [53] indicaram que a condição de hiperbolicidade não é necessária, mostrando casos onde o sistema não é hiperbólico, mas (utilizando baixo nível de ruído) conseguem órbitas sombreadas.

Para formalizar o conceito de sombreamento, pode-se considerar que o erro de um passo é dado por

$$d(f(x_i), x_{i+1}) \leq \delta,$$

enquanto que a *distância de sombreamento* é dada por:

$$d(f^n(x_0), x_n) \leq \epsilon,$$

onde $d(\cdot, \cdot)$ é uma medida de distância.

Assim pode-se definir o sombreamento a partir do erro de um passo. Uma δ -pseudo-órbita é uma sequência de pontos $\{x_0, x_1, \dots\}$, cujo erro de um passo é:

$$d(f(x_i), f(x_{i+1})) \leq \delta, \quad \forall i \geq 0.$$

Diz-se que f tem a propriedade de sombreamento se toda δ -pseudo-órbita pode ser ϵ -sombreada por uma órbita verdadeira, isto é, existe uma condição inicial x_0 tal que $d(f^n(x_0), x_n) \leq \epsilon, \forall n \geq 0$.

Às vezes uma órbita numérica é ϵ -sombreada por uma órbita verdadeira apenas durante um número finito de passos. Dizemos que uma pseudo-órbita $\{x_i\}_{i=1}^N$ tem um *pulso aleatório* (*glitch*) no n -ésimo ponto ($n < N$) se $\{x_i\}_{i=1}^n$ pode ser sombreada, mas $\{x_i\}_{i=1}^{n+1}$ não pode [7].

6.3 A metodologia desenvolvida

Para realizar a reamostragem de uma série temporal, foi desenvolvido nesta tese um algoritmo baseado em algoritmos de previsão. Na previsão de uma série temporal, busca-se a evolução de uma série existente. Para a reamostragem, foi desenvolvido um algoritmo que cria um ponto fictício no espaço de fase e o evolui. A metodologia de evoluir os pontos no espaço de fase é idêntica à existente nos algoritmos de previsão.

A seguir detalhamos a metodologia numa sequência de 13 etapas. Além disso, para a maioria dos itens é indicada a seção desta tese que detalha o procedimento utilizado. Quando uma seção não foi associada ao item, isto

significa que nenhuma seção trata explicitamente do assunto. Estes itens (sem uma seção associada) tratam de procedimentos considerados triviais, portanto não foram detalhados em nenhum outro trecho deste trabalho.

1. A partir da série temporal original de tamanho N , são obtidas $N - w + 1$ janelas de tamanho w (vide seção 2.5).
2. Uma taxa de filtragem t_f é definida. Isto significa que no processo de filtragem somente as t_f frequências mais baixas serão utilizadas. As demais serão tratadas como ruído e serão removidas (vide seção 2.5.2).
3. Uma taxa de interpolação t_i é definida. Isto é feito para que sejam criados mais pontos em cada órbita. Desta forma, será possível encontrar (em cada trajetória) um ponto que esteja mais próximo do vetor pergunta (vide seção 2.5.2).
4. Para criar vetores filtrados no espaço de fase, cada uma das janelas obtidas passará por um processo de imersão passa-baixa. Este procedimento consiste em realizar a transformada de Fourier discreta de tamanho w , depois as t_f frequências mais altas são excluídas e então uma transformada inversa de tamanho $w \times t_i/t_f$ é realizada. Implementando este procedimento numa janela de tamanho w , obtém-se t_i vetores filtrados de tamanho $m = w/t_f$. O número total de vetores obtidos no espaço de fase será: $(N - w + 1) \times t_i$ (vide seção 2.5.2).
5. A série temporal original é interpolada pelo mesmo fator de interpolação t_i . Observe que agora existem $(N - w + 1) \times t_i$ vetores. Para que haja uma evolução correspondendo a cada um dos vetores, a série temporal deve ser interpolada, passando a ter tamanho $N \times t_i$.
6. Uma condição inicial no espaço de fase é escolhida. Esta condição inicial é um ponto no espaço de fase obtido pela imersão da série temporal original. Para que o ponto inicial não seja idêntico a algum ponto da série original, é feita uma pequena perturbação nesta condição inicial. Este ponto perturbado será o vetor pergunta.
7. Os k vizinhos mais próximos ao vetor pergunta são encontrados (vide seção 4.1).
8. Os k vizinhos mais próximos (pontos no espaço de fase m -dimensional), em conjunto com as suas evoluções (escalares), são utilizados para construir o modelo linear (vide seção 4.3).

9. É efetuada a regularização do modelo linear para reduzir a sua variância. Para isto os vizinhos mais próximos serão projetados num espaço l -dimensional (vide capítulo 5).
10. A condição inicial é evoluída pelo modelo linear, obtendo um escalar. Este escalar passa a fazer parte da série reamostrada.
11. A partir do vetor pergunta m -dimensional, realiza-se um processo para encontrar o trecho da série temporal que corresponderia a este vetor. Para isto, realiza-se uma transformada de Fourier de tamanho m , depois realiza-se uma transformada inversa de tamanho w , onde as componentes necessárias para completar o vetor serão todas nulas (vide seção 2.5.2).
12. Utiliza-se as $w - 1$ últimas componentes da janela, mais o valor que foi predito para obter uma nova janela de tamanho w . Esta janela é filtrada para se obter um ponto no espaço de fase m -dimensional. Este ponto é o novo vetor pergunta.
13. Todo o processo (desde o item 7) é repetido até que se obtenha uma série com N pontos.

É importante observar que as primeiras iterações foram desprezadas, no intuito de eliminar qualquer possível transiente.

No processo de gerar uma pequena perturbação para a condição inicial, os k vizinhos mais próximos são encontrados. A perturbação é feita na direção do hiperplano onde os k vizinhos estão distribuídos. A perturbação deve necessariamente ser menor do que a distância do k -ésimo vizinho. Isto garante que a perturbação seja pequena.

Caso a perturbação não fosse realizada, o número máximo de reamostragens distintas possíveis seria igual ao número de pontos no espaço de fase. Quando fosse escolhida a mesma condição inicial, a mesma série seria gerada. Com a perturbação aleatória, a probabilidade de haver duas séries idênticas é praticamente nula.

No algoritmo desenvolvido, a metodologia utilizada foi um pouco diferente daquela apresentada acima. Uma pequena modificação foi feita para que fosse combinada a predição iterada de um único passo, com a predição multi-passo. Para cada previsão, foram utilizadas explicitamente as previsões feitas pelos últimos w pontos no espaço de fase. Isto foi feito para que fosse preservada a estrutura temporal da série[43]. Como w já foi escolhido como um parâmetro que corresponde à escala de tempo da série temporal, este

mesmo parâmetro foi utilizado para que fosse preservada a estrutura da série em relação ao tempo. Cada novo ponto da série temporal será a média das w previsões realizadas pelos últimos w vetores no espaço de fase. No começo da reamostragem da série temporal, quando ainda não dispomos de w vetores, todos os vetores presentes são utilizados.

6.4 Aplicação

Para aplicar a metodologia de reamostragem descrita acima, foi escolhida uma série temporal originada pelo conjunto de equações diferenciais de Lorenz.

Utilizando o algoritmo de Runge-Kutta de quarta ordem [39], as equações de Lorenz foram integradas numericamente. Para esta integração foi utilizado um passo de 0,01, mas foi escolhido um a cada 3 pontos para fazer parte da série, assim o intervalo entre os pontos da série é de 0,03. A partir de então, foi gerada uma série temporal com 1 000 pontos. A série está apresentada na figura 6.1.

A partir desta série foram geradas 100 reamostragens, ou seja, 100 novas séries temporais que apresentam as mesmas características da série original. Elas apresentam as mesmas características por possuírem um mesmo atrator.

Das 100 séries geradas, uma delas foi escolhida aleatoriamente. Esta série está apresentada na figura 6.2. Já na figura 6.3 pode ser observado o espaço de fase da série original assim como o da série reamostrada. Observa-se que ambas possuem a mesma estrutura.

A similaridade dos espaços de fase foi observada não apenas nesta série reamostrada apresentada, mas em todas as séries reamostradas, garantindo que o algoritmo desenvolvido é robusto.

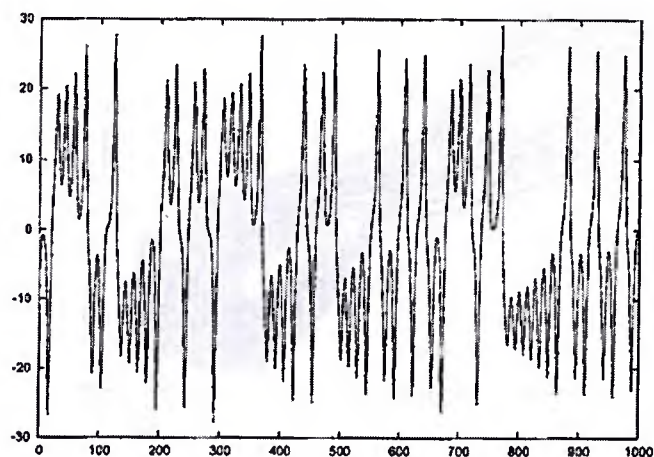


Figura 6.1: Série temporal com 1 000 pontos obtida a partir do sistema de equações de Lorenz.

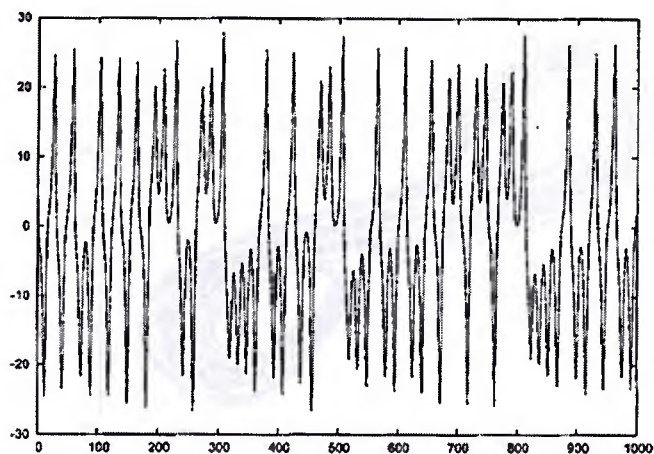


Figura 6.2: Reamostragem da série temporal apresentada na figura 6.1.

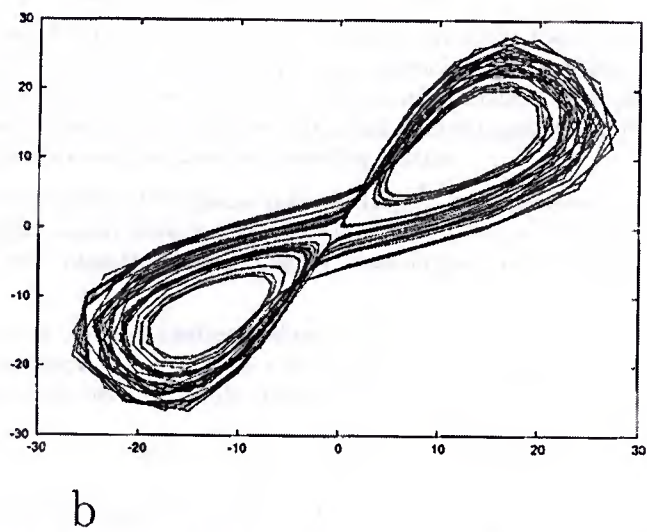
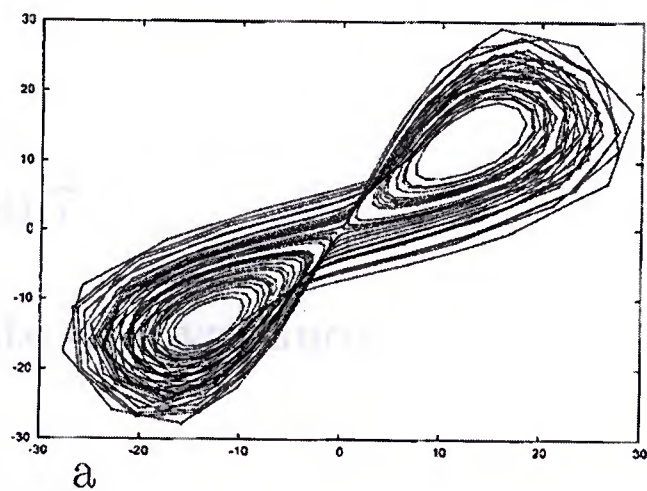


Figura 6.3: Espaço de fase (a) da série temporal original (figura 6.1) e (b) da série reamostrada (figura 6.2).

Capítulo 7

Expoente de Lyapunov

7.1 Introdução

Uma forma de se saber se um sistema é caótico é através do expoente de Lyapunov. Este expoente recebeu tal nome em homenagem ao matemático russo Aleksandr Lyapunov, que defendeu sua tese de doutorado em 1892, cujo título era: "Problema geral da estabilidade do movimento". Com este trabalho Lyapunov iniciou os estudos sobre estabilidade de equações diferenciais.

Ao estudar a estabilidade de um conjunto de equações diferenciais, existe grande interesse em conhecer a taxa média de convergência ou de divergência de duas trajetórias vizinhas. Isto está intimamente relacionado à predição. Sistemas caóticos apresentam divergência exponencial entre trajetórias próximas. Isto caracteriza a sensibilidade às condições iniciais.

Para mensurar se existe divergência entre as trajetórias, utiliza-se o expoente de Lyapunov, o qual mede a taxa média de divergência (ou convergência) exponencial entre duas trajetórias vizinhas. Esta média é realizada sobre todo o atrator.

Considerando um sistema caótico e duas trajetórias infinitesimalmente próximas, cuja separação inicial é dada por δx_0 , depois de um intervalo de tempo t , passarão a ter uma separação dada por:

$$|\delta x(t)| = e^{\lambda t} |\delta x_0|,$$

onde λ é o expoente de Lyapunov. Quanto maior for este expoente, tanto maior será a taxa de divergência exponencial.

Expoentes negativos indicam que as trajetórias se aproximam ao longo do tempo. Por outro lado, expoentes positivos indicam que as trajetórias se afastam. Os sistemas caóticos apresentam divergência exponencial entre as trajetórias próximas, portanto o expoente positivo caracteriza esta propriedade*. É importante ressaltar que o expoente de Lyapunov também pode ser nulo. Neste caso, não pode ser dito que as trajetórias não se alteram com o tempo. A única afirmação a ser feita é que não ocorre uma divergência (ou convergência) exponencial.

O expoente de Lyapunov é utilizado para caracterizar se um sistema é caótico. Além de permitir detectar a caoticidade, o expoente de Lyapunov é de grande utilidade no estudo de séries temporais, pois, medindo a sensibilidade às condições iniciais e a taxa de divergência das trajetórias próximas, pode-se conhecer o horizonte de predição da série.

7.2 Espectro de Lyapunov

Um sistema possui tantos expoentes de Lyapunov quanto a dimensão do espaço de fase. Cada um dos expoentes corresponde a um eixo principal de um elemento de volume. A orientação de cada um dos eixos principais varia com a posição no atrator. Portanto, os expoentes de Lyapunov não correspondem à qualquer orientação particular no espaço de fase.

Os expoentes de Lyapunov descrevem o comportamento do espaço tangente ao espaço de fase. Sendo o sistema dinâmico dado por $x_n = f(x_{n-1})$, é fácil ver que $x_n = f^n(x_0)$.

Seja x_0 a condição inicial e δ_0 uma perturbação; então a evolução da perturbação, que é um vetor tangente, é dada por [12]:

$$\delta_n = J(x_{n-1})\delta_{n-1},$$

onde $J(x_{n-1})$ é a matriz jacobiana calculada no ponto x_{n-1} .

A evolução da perturbação é dada por:

$$\delta_n = J^n(x_0)\delta_0.$$

onde

$$J^n(x_0) = J(x_{n-1}) \cdot J(x_{n-2}) \cdots J(x_0).$$

*Para uma discussão mais técnica sobre o conceito de caos, veja as seguintes referências [58, 59]

A partir da matriz jacobiana podemos definir a seguinte matriz:

$$L(x_0) = ((J^n(x_0))^T J^n(x_0)),$$

que é hermitiana e real, assim os seus autovalores (designados por $\Lambda_i(x_0)$) serão reais e não-negativos.

Utilizando os autovalores $\Lambda_i(x_0)$ é possível definir os expoentes de Lyapunov λ_i , assim:

$$\lambda_i(x_0) = \frac{1}{2n} \ln \Lambda_i(x_0).$$

Para calcular o expoente de Lyapunov, o número de iterações do mapeamento $f(\cdot)$ deve ser muito grande. Os resultados teóricos utilizam $n \rightarrow \infty$, mas nos cálculos a partir de um conjunto de dados, um valor como $n = 1\,000$, por exemplo, é utilizado.

O conjunto de expoentes de Lyapunov de um atrator é conhecido como *espectro de expoentes de Lyapunov*. Como os expoentes estão associados com as principais direções de expansão, ou contração, eles costumam ser ordenados de forma decrescente, portanto:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

Apesar de existir todo um espectro de expoentes de Lyapunov, de uma forma geral apenas o expoente máximo é de interesse. Se o maior expoente for positivo, então o sistema é caótico. Se o maior expoente não for positivo, então o sistema não é caótico. Existe a possibilidade de não se saber com certeza se o expoente é positivo. Um expoente de Lyapunov máximo muito próximo de zero pode não dar muita informação sobre o sistema. Neste caso deve-se conhecer a incerteza associada a este valor para que se possa afirmar se o sistema é caótico.

7.3 Cálculo computacional

Para encontrar o expoente de Lyapunov a partir de uma série temporal, vários algoritmos foram propostos. Os algoritmos que parecem funcionar melhor são os propostos por Rosenstein et al. [3] e por Kantz [60]. Eles são bastante similares e serão descritos a seguir.

Para encontrar o expoente de Lyapunov a partir de uma série temporal, primeiramente o atrator do espaço de fase deve ser reconstruído. Isto

costuma ser feito através das coordenadas de atraso. Como o número de expoentes de Lyapunov é o mesmo que a dimensão do espaço de fase, pode ocorrer um problema aqui, associado ao número de dimensões utilizadas na reconstrução do espaço de fase. Nesta reconstrução pode ser utilizado um número de dimensões superior ao número de dimensões do sistema original. Isto acaba gerando expoentes de Lyapunov espúrios, expoentes extras, que não têm nenhum significado.

Como estamos interessados em obter somente o expoente máximo, não é preciso nos preocuparmos com o problema devido à dimensão do espaço de fase reconstruído, portanto pode-se prosseguir com a obtenção do expoente.

O expoente de Lyapunov máximo é dado por:

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \frac{\delta_n}{\delta_0}.$$

Na prática, o limite do número de pontos tendendo a infinito não pode ser obtido. Assim, alguma aproximação deve ser feita no cálculo deste expoente.

Primeiramente, escolhe-se um ponto qualquer x_{n_0} no espaço de fase e os seus vizinhos mais próximos (x_i , $i = 1, \dots, k$) são encontrados. Os vizinhos mais próximos são encontrados para que se possa acompanhar as trajetórias próximas. Encontra-se a evolução para o ponto de referência ($y_{n_0+\Delta n}$), assim como para os seus vizinhos ($y_{i+\Delta n}$). A evolução de cada um destes pontos pode ser considerada a partir da série temporal, sem que consideremos os pontos no espaço de fase. Portanto a distância entre as trajetórias será considerada como a distância entre os escalares que são a evolução do ponto de referência e a evolução dos vizinhos. Obtém-se então a distância média entre a evolução do ponto de referência e a evolução dos vizinhos. É calculado o logaritmo desta distância média.

Para que o resultado represente uma média do que ocorre ao longo de todo o atrator, o procedimento é repetido variando-se o ponto de referência (n_0). Todos os pontos do espaço de fase devem ser utilizados como pontos de referência. Depois que o resultado é obtido para cada um destes pontos de referência, extrai-se a média de todos estes valores. Esta será a taxa logarítmica de expansão, dada por:

$$S(\Delta n) = \frac{1}{N} \sum_{n_0=1}^N \ln \left(\frac{1}{k} \sum_{i=1}^k |y_{n_0+\Delta n} - y_{i+\Delta n}| \right),$$

onde k é o número de vizinhos.

Foi assumido que a divergência é exponencial. Para verificar isto, a taxa logarítmica de expansão, $S(\Delta n)$, deve ser colocada num gráfico contra o

intervalo de tempo (Δn). Se a expansão for exponencial, este gráfico deve se comportar como uma linha reta, pelo menos para intervalos de tempo pequenos. Caso não se observe uma linha reta para nenhum intervalo, o expoente de Lyapunov não pode ser definido.

Deve ser observado que o espaço de fase é limitado, assim o número de passos (Δn) para o cálculo do expoente de Lyapunov não pode ser muito grande. A divergência exponencial ocorre até que a distância entre as órbitas atinja o tamanho do atrator do espaço de fase. Assim, no gráfico da taxa logarítmica de expansão, $S(\Delta n)$, em função do intervalo de tempo (Δn), observa-se um crescimento seguido de um patamar, indicando uma saturação devido ao tamanho do atrator.

7.4 O expoente de Lyapunov para a série temporal e para as suas reamostragens

Um dos objetivos desta tese é obter reamostragens de séries temporais geradas por sistemas caóticos. A partir destas reamostragens é possível associar uma incerteza para os invariantes obtidos deste sistema. Um dos invariantes de interesse é o expoente de Lyapunov, mas outras grandezas dinâmicas podem ser calculadas (ver detalhes em Menezes e Francisco [61]).

Foi calculado numericamente o expoente de Lyapunov para o sistema de equações diferenciais de Lorenz. Para isto, este conjunto de equações diferenciais foi integrado numericamente para se obter a série temporal. Foram realizadas 100 integrações para que se tenha uma distribuição de valores para este expoente.

É de se esperar que ocorra apenas um valor para o expoente de Lyapunov, independente da série utilizada. Entretanto, devido ao número limitado de pontos na série temporal, ocorre uma diferença entre o valor real deste expoente e o valor medido a partir da série temporal. Assim com um total de 100 séries é possível obter uma distribuição para este expoente. Esta distribuição é mostrada na figura 7.1.

O valor do expoente de Lyapunov do sistema de Lorenz é $\lambda = 1,5$ [3]. O valor médio que foi obtido a partir de um conjunto de 100 séries temporais foi de $\bar{\lambda} = 1,51$. O desvio padrão foi de 0,23.

De uma forma geral, ao se trabalhar com uma série empírica, dispõe-se apenas de uma realização da série temporal. Assim, não é possível obter

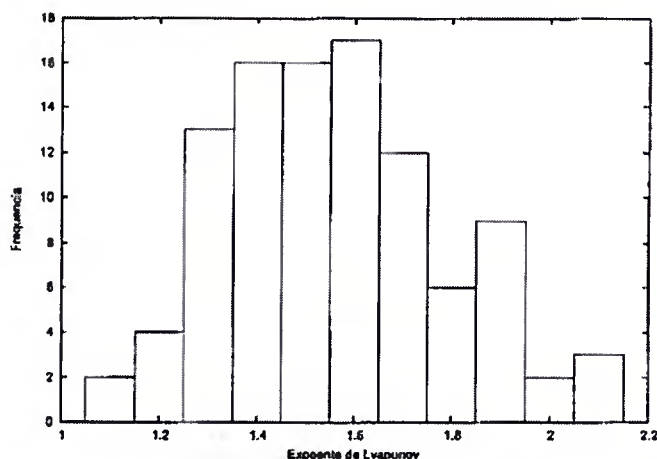


Figura 7.1: Histograma apresentando a distribuição de valores do expoente de Lyapunov para 100 diferentes integrações numéricas das equações de Lorenz.

uma distribuição para um invariante a partir de várias séries reais, contudo é possível obter uma distribuição utilizando as séries reamostradas.

Realizando a reamostragem das séries temporais, os valores obtidos para o expoente de Lyapunov são dados pela distribuição que está apresentada na figura 7.2.

A partir da distribuição de valores do expoente de Lyapunov devido às reamostragens, foi possível encontrar a média e o desvio padrão do expoente. Estes valores são dados por:

$$\bar{\lambda} = 1,56, \quad \sigma = 0,22.$$

Podemos ainda obter intervalos de confiança para o expoente de Lyapunov. Para isto, na tabela 7.1 são mostrados os percentis obtidos a partir das séries reais, assim como os percentis obtidos a partir das séries reamostradas.

Portanto, os valores apresentados na tabela 7.1 para as séries reais servem apenas como uma comparação com os valores obtidos a partir das séries reamostradas. Observamos, que ambas as distribuições possuem intervalos próximos, assim como ocorre com a média e o desvio-padrão de cada uma das distribuições.

A partir destas duas distribuições, observa-se que a técnica de reamostragem desenvolvida neste trabalho pode ser utilizada para criar um intervalo de con-

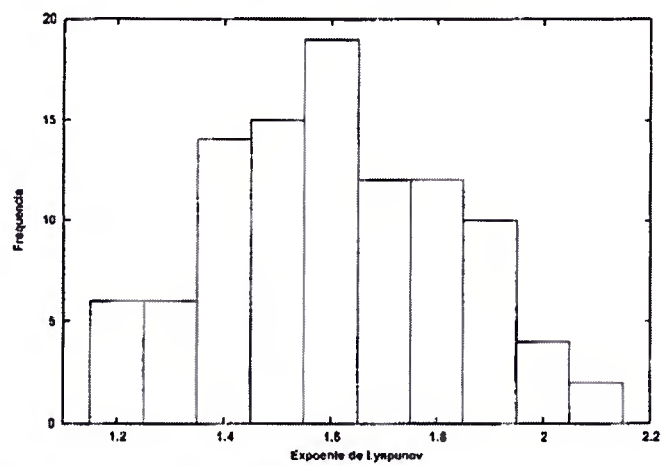


Figura 7.2: Histograma apresentando a distribuição de valores para o expoente de Lyapunov para as 100 reamostragens do sistema de Lorenz.

	5%	10%	50%	90%	95%
Séries Reais	1,17	1,23	1,49	1,81	1,91
Séries Reamostradas	1,19	1,25	1,56	1,86	1,93

Tabela 7.1: Percentis da distribuição do expoente de Lyapunov baseado em 100 realizações do sistema de Lorenz (100 integrações numéricas com diferentes condições iniciais, chamadas de séries reais) e em 100 reamostragens do mesmo sistema (séries reamostradas).

fiança para os invariantes do sistema. O resultado obtido reflete o resultado que se obtém a partir de várias realizações do sistema em estudo.

Capítulo 8

Redes de Mapas Acoplados

8.1 Introdução

O amplo interesse em redes de mapas acoplados na pesquisa científica e nas aplicações é principalmente devido ao seu papel como uma família de sistemas possuindo comportamento universal com forma explícita das interações locais. Aplicações destas redes são encontradas em várias áreas, tais como: dinâmica, turbulência, transições de fase, geofísica, ótica, genética, processos de informação humana, e assim por diante [62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72].

É válido mencionar que Kaneko estudou dinâmica de padrões espaço-temporais em redes de mapas acoplados em 1 e 2 dimensões [73]. Em particular, Kaneko mostrou a existência de vários tipos de padrões similares em redes de mapas unidimensional e bidimensional no regime de acoplamento fraco. Entretanto, no regime de acoplamento forte, a dinâmica das redes em 1 e 2 dimensões diferem [73].

Estas redes de mapas acoplados também exibem a instabilidade do tamanho, onde a dinâmica é completamente dependente do tamanho do sistema [74, 75]. Por exemplo, a rede unidimensional mostra um estado síncrono estável para redes pequenas, enquanto ocorre uma transição para o caos espaço-temporal para redes maiores. Outra característica das redes de mapas acoplados analisadas aqui é a sua auto-similaridade, uma característica presente em diversos fenômenos [77, 78]. Tal propriedade está presente mesmo em abordagens mais sofisticadas da estrutura do espaço-tempo e Física de partículas elementares [79, 80].

Apesar de, em vários estudos, o foco estar em redes unidimensional e bidimensional, aqui obtém-se uma ampla gama de propriedades que caracterizam as séries temporais de redes de mapas acoplados em 1, 2 e 3 dimensões. Assim, uma comparação pode ser obtida, com o objetivo de estudar o comportamento do sistema sob mudança de dimensão. Devido à complexidade deste estudo, o trabalho concentrar-se-á no mapa logístico, mas o trabalho numérico pode ser estendido, usando computadores com mais processamento, para uma evolução do tipo Lorenz, além de outros modelos.

Como mencionado acima, a dependência dinâmica nos parâmetros de controle, particularmente em redes de dimensão 1 e 2 é bem conhecida [73, 81, 82], portanto os valores dos parâmetros foram fixados, permitindo um regime caótico interessante.

8.2 Rede de mapas acoplados

Uma rede de mapas acoplados é uma rede d -dimensional onde cada posição evolui no tempo por meio de um mapa (equação de recorrência). A equação de recorrência apresenta a seguinte forma

$$x_{t+1} = f(x_t).$$

Em geral, as posições do mapa interagem entre si. Assim o estado atual de uma posição da rede depende do estado anterior da própria posição, assim como do estado das demais posições.

A interação costuma ser local, assim cada posição da rede interage somente com os vizinhos mais próximos.

8.2.1 Rede unidimensional

É interessante começar analisando o caso unidimensional. Considerando-se N mapas interagindo entre si, obtém-se

$$x_{t+1}^i = (1 - \epsilon)f(x_t^i) + \frac{\epsilon}{S} \sum_{i' \neq i} f(x_t^{i'}), \quad (i = 1, 2, \dots, N),$$

onde i é o índice que representa cada uma das N posições na rede. O acoplamento entre os vizinhos é dado pelo *parâmetro de acoplamento* ϵ , que assume um valor no intervalo $[0, 1]$. O parâmetro S indica o número de vizinhos que foram considerados na ponderação.

É comum considerar que a interação ocorre somente entre os vizinhos mais próximos, assim [74, 73, 75]

$$x_{i+1}^i = (1 - \epsilon)f(x_i^i) + \frac{\epsilon}{2} [f(x_i^{i-1}) + f(x_i^{i+1})], \quad (i = 1, 2, \dots, N).$$

Cada mapa está acoplado a dois dos seus vizinhos mais próximos.

8.2.2 Rede bidimensional

No caso bidimensional, existe um total de N^2 mapas interagindo. Portanto cada dimensão tem tamanho N , e a expressão é dada por:

$$x_{i+1}^{i,j} = (1 - \epsilon)f(x_i^{i,j}) + \frac{\epsilon}{S} \sum_{i' \neq i, j' \neq j} f(x_i^{i',j'}), \quad (i, j = 1, 2, \dots, N).$$

Considerando que a interação ocorre somente entre os vizinhos mais próximos, obtém-se

$$x_{i+1}^{i,j} = (1 - \epsilon)f(x_i^{i,j}) + \frac{\epsilon}{4} [f(x_i^{i-1,j}) + f(x_i^{i+1,j}) + f(x_i^{i,j-1}) + f(x_i^{i,j+1})], \quad (i, j = 1, 2, \dots, N). \quad (8.1)$$

Cada mapa está acoplado a quatro dos seus vizinhos mais próximos.

8.2.3 Rede tridimensional

Considerando N mapas em cada dimensão, haverá um total de N^3 mapas interagindo. A expressão será dada por:

$$x_{i+1}^{i,j,k} = (1 - \epsilon)f(x_i^{i,j,k}) + \frac{\epsilon}{S} \sum_{i' \neq i, j' \neq j, k' \neq k} f(x_i^{i',j',k'}), \quad (i, j, k = 1, 2, \dots, N).$$

Quando a interação ocorre somente entre os vizinhos mais próximos, obtém-se

$$x_{i+1}^{i,j,k} = (1 - \epsilon)f(x_i^{i,j,k}) + \frac{\epsilon}{6} [f(x_i^{i-1,j,k}) + f(x_i^{i+1,j,k}) + f(x_i^{i,j-1,k}) + f(x_i^{i,j+1,k}) + f(x_i^{i,j,k-1}) + f(x_i^{i,j,k+1})], \quad (i, j, k = 1, 2, \dots, N). \quad (8.2)$$

Cada mapa está acoplado a seis dos seus vizinhos mais próximos.

8.2.4 A escolha dos parâmetros e do mapa

Para facilitar a análise, os parâmetros serão fixos em todo o tratamento da rede de mapas acoplados. O número de mapas em cada dimensão será: $N = 50$. O parâmetro de acoplamento será sempre: $\epsilon = 0,40$. O número de vizinhos é dependente do número de dimensões consideradas:

- para $d = 1$, então $S = 2$;
- para $d = 2$, então $S = 4$;
- para $d = 3$, então $S = 6$.

A escolha do mapa teve de levar em consideração o custo computacional. No caso tridimensional, haverá um total de 125 000 mapas interagindo e evoluindo ao longo do tempo. Devido ao tempo para gerar os dados, foi utilizado o mapa logístico, que é um dos mapas mais simples capaz de gerar comportamento caótico.

8.3 Mapa logístico

O mapa logístico é uma das equações mais elementares que pode gerar um comportamento caótico. Este mapa é simplesmente uma equação de segundo grau, a versão discreta da equação diferencial que descreve a evolução demográfica de uma população.

A versão contínua foi proposta por Pierre François Verhulst, sendo dada por [76]:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{k} \right).$$

De acordo com esta equação, devido aos recursos limitados, a população deveria atingir um regime assintótico constante.

Esta versão contínua não apresenta nenhum comportamento interessante. Se a população tiver um valor inicial diferente de zero, então a mesma convergirá assintoticamente para k .

A versão discreta desta equação é dada por:

$$N_{j+1} = N_j + rN_j \left(1 - \frac{N_j}{k} \right).$$

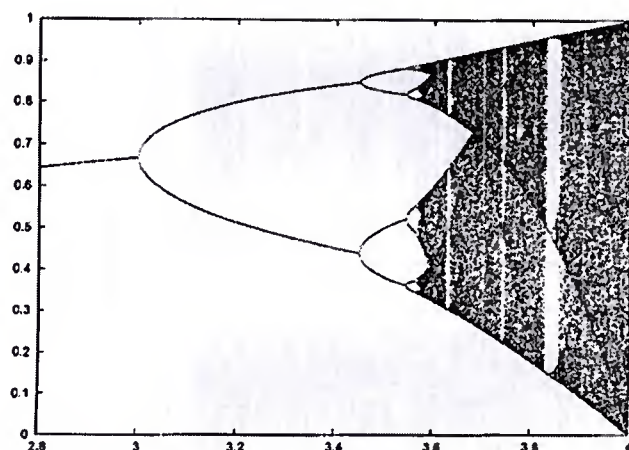


Figura 8.1: Diagrama de bifurcações do mapa logístico.

Através de uma mudança de variável, assim como uma redefinição da constante, é possível reescrever esta equação da seguinte forma:

$$x_{j+1} = \mu x_j (1 - x_j), \quad x \in [0, 1], \quad \mu \in [0, 4].$$

Esta equação possui um comportamento muito mais interessante do que a sua versão contínua.

Observa-se que, para valores pequenos do parâmetro μ , o sistema estabiliza-se num valor fixo. Quando μ é maior do que 3, o sistema oscila entre dois valores eternamente. Observa-se ainda que a medida que o valor do parâmetro μ aumenta, ocorre uma cascata de bifurcações de duplicação de período. Já para μ maior que aproximadamente 3,57 ocorre um comportamento caótico. Este comportamento se prolonga até $\mu = 4$. Para alguns valores específicos do parâmetro μ no intervalo $[3,57, 4]$, observa-se que não ocorre um comportamento caótico, mas apenas uma oscilação entre alguns valores. Para μ maior que 4, a solução diverge para a maioria das condições iniciais.

Para obter um mapa caótico, neste trabalho foi utilizado o valor $\mu = 4$.

8.4 Séries temporais

A série temporal gerada por estes mapas está na figura 8.2. Na série temporal gerada por uma rede de mapas acoplados tridimensional, figura 8.2(c),

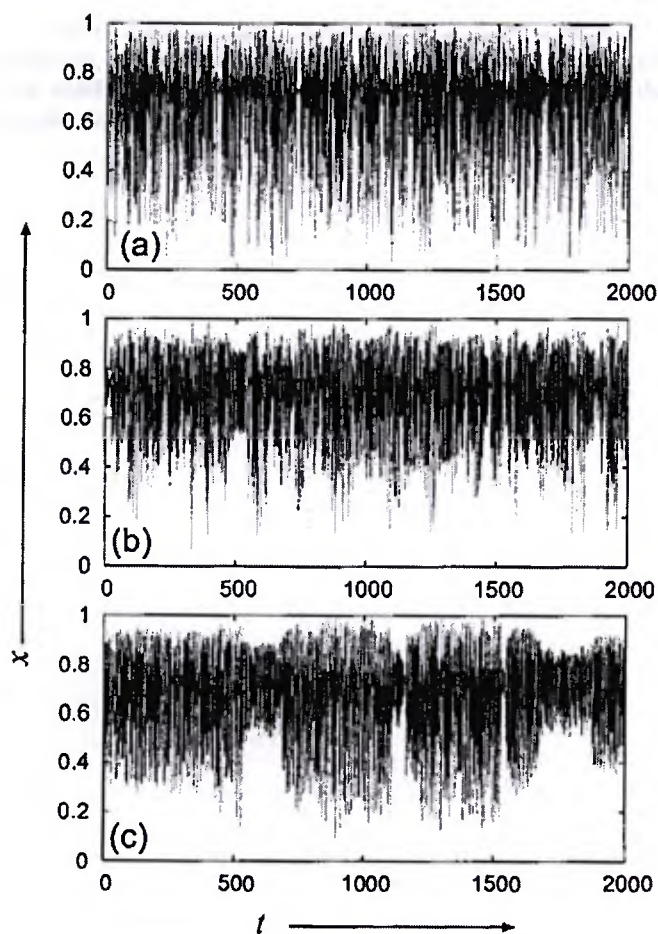


Figura 8.2: Série temporal de um ponto da rede escolhido aleatoriamente para rede de mapas acoplados (a) unidimensional, (b) bidimensional e (c) tridimensional.

observa-se uma característica distinta das demais séries. Ou seja, existe uma variação da amplitude do movimento, resultando em intervalos de tempo em que a variância é alta, seguida de intervalos menores onde a variância é baixa.

Estas séries temporais serão estudadas no próximo capítulo, onde as características estatísticas, dinâmicas e geométricas serão exploradas. Será feita inclusive uma análise do que ocorre com estas medidas em função da mudança na dimensão das redes.

Capítulo 9

Análise de redes de mapas acoplados

Neste capítulo será realizado um estudo sobre as propriedades das séries temporais geradas por redes de mapas acoplados. Serão observadas as características das redes com dimensão 1, 2 e 3.

No presente estudo trabalha-se especificamente com as séries temporais geradas pela evolução de redes de mapas acoplados. Para realizar a análise destas redes, são exploradas as suas propriedades estatísticas, dinâmicas e geométricas; também são obtidas medidas de dimensão e previsibilidade quando a dimensão da rede é aumentada. A robustez destas conclusões foi testada tomando-se vários pontos aleatórios na rede, para cada dimensão, e foi feita a análise da série temporal correspondente. Apesar do fato de as propriedades dos mapas serem conhecidas, ocorrem fenômenos interessantes quando estes mapas estão acoplados. Características diferentes das demais são encontradas em redes tridimensionais.

Diversos sistemas de interesse prático exibem comportamento de escala $1/f$ [83, 84, 85] e aqui é implementada a *análise de flutuação sem tendência* para analisar tal propriedade [86, 87, 89]. São utilizadas caracterizações estatísticas, além de outras, para a rede de mapas acoplados, a fim de estabelecer um conjunto de propriedades que aparecem nas séries temporais. Medidas de dimensão e outros diagnósticos serão usados para detectar diferenças no comportamento destes modelos.

Nesta contribuição são dados vários argumentos que servem para comparar e esclarecer a relação entre as séries temporais das redes de dimensão 1, 2 e 3. Uma nova característica que emerge em modelos tridimensionais,

não presentes em dimensões mais baixas, ocorre no gráfico da flutuação. Uma outra é o aumento da persistência quando a dimensão é aumentada.

9.1 Análise estatística

9.1.1 Análise de flutuação sem tendência

A *análise de flutuação sem tendência* é um método para determinar a auto-similaridade de um sinal. Esta medida foi introduzida por Peng [86, 87] e é uma extensão da análise de flutuação, que é afetada pela não-estacionariedade. Portanto, a análise de flutuação sem tendência pode ser utilizada em séries temporais não-estacionárias.

Auto-similaridade é uma propriedade presente nos objetos fractais, além da dimensão fracionária. A auto-similaridade indica que um objeto, assim como as suas sub-partes, em várias escalas, possuem as mesmas propriedades estatísticas.

O cálculo

Considere uma série temporal $x(i)$, $i = 1, 2, \dots, N$. A partir dela é possível definir uma série auto-similar, que é dada pela integração da série original, ou seja:

$$y(i) = \sum_{j=1}^i [x(j) - \bar{x}],$$

onde $\bar{x}$ é a média, dada por:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x(i).$$

Sempre que a série temporal original $x(i)$ for uma série limitada, ela deve ser integrada, para que as propriedades fractais sejam detectadas [88].

A próxima etapa consiste em verificar se a auto-similaridade está presente na série temporal integrada $y(i)$. Esta série é dividida em l caixas de tamanhos iguais, cada uma delas com tamanho n , onde $l = \text{Int}(N/n)$. Os dados em cada caixa são ajustados por uma função linear através de uma regressão. A função linear, que é escrita como $y_{\text{fit}}(i)$, representa a tendência local. Depois disso, a tendência é removida da série temporal integrada, subtraindo a

tendência local, $y_{\text{fit}}(i)$, em cada caixa. Então a flutuação (que é a raiz da quadrática média do erro) é calculada desta forma:

$$F(n) = \sqrt{\frac{1}{l} \sum_{k=0}^{l-1} \left\{ \sum_{i=kn+1}^{(k+1)n} [y(i) - y_{\text{fit}}(i)]^2 \right\}}.$$

Repetindo o cálculo acima para diferentes tamanhos de caixa, a relação entre a flutuação, $F(n)$, e o tamanho n pode ser estabelecida. Em geral, ocorre uma lei de potência entre $F(n)$ e n , assim $F(n)$ se comporta desta forma: $F(n) \sim n^\alpha$, onde α é chamado de *expoente de escala*, ou *expoente da análise de flutuação sem tendência*, ou *parâmetro de auto-similaridade* [88].

Dependendo do valor de α pode-se classificar a série temporal como sendo correlacionada ou anti-correlacionada. Para uma série temporal não-correlacionada, $\alpha = 0,5$; para uma correlação positiva, $\alpha > 0,5$; e $\alpha < 0,5$ para uma série temporal anti-correlacionada. Uma análise detalhada de diferentes tendências de conjuntos de dados artificiais foi feita por Hu et al. [89].

Como ocorre com vários métodos numéricos, é possível encontrar um valor para α sem que este implique na auto-similaridade da série temporal. A auto-similaridade requer que os pontos (no gráfico log-log) estejam aproximadamente dispostos numa linha ao longo de vários valores do tamanho das caixas n .

Comparação com a autocorrelação

Quando a autocorrelação $C(n)$, separada por n unidades de tempo, decai como

$$C(n) \sim n^{-\gamma}, \quad (9.1)$$

onde $0 < \gamma < 1$, existe uma relação entre o expoente de auto-similaridade α e o expoente de correlação γ , assim [88]:

$$\alpha = 1 - \frac{\gamma}{2}. \quad (9.2)$$

Deve ser observado que $\gamma = 1$ para dados não-correlacionados. O expoente de escala α também está relacionado com a inclinação espectral, $1/f^\beta$, com [88]:

$$\beta = 2\alpha - 1. \quad (9.3)$$

Portanto, $\alpha = 1$ para o espectro de ruído do tipo $1/f$. Pode-se verificar que $\alpha = 1,5$ para movimento browniano (passeio aleatório) onde $\beta = 2$.

Resultados para redes de mapas acoplados

Nesta tese, o expoente α foi calculado para redes de mapas acoplados com uma, duas e três dimensões. Em todos os casos, $F(n)$ segue uma lei de potência, indicando a presença de auto-similaridade. O expoente α encontrado para a rede unidimensional foi de 0,54 que é próximo de 0,5. Isto mostra que a série temporal é não-correlacionada. No caso da rede bidimensional, o expoente $\alpha \sim 0,62$ mostrando uma correlação positiva na série temporal [veja figura 9.1].

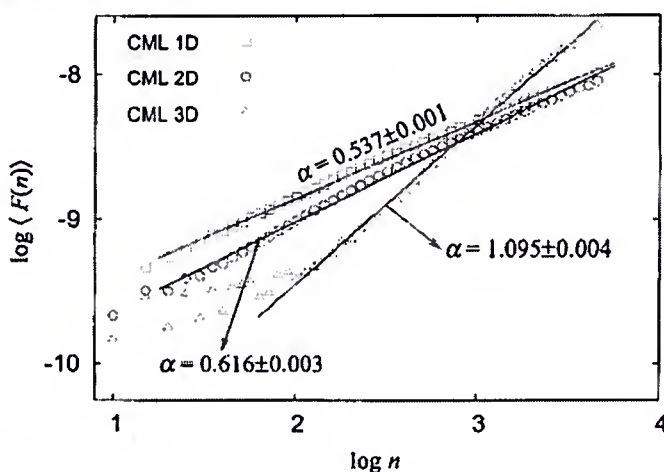


Figura 9.1: A variação de $F(n)$ em função de n para redes de mapas acoplados com dimensão 1, 2 e 3.

A série temporal originada da rede tridimensional exibe um comportamento completamente diferente. Na figura 9.1 é mostrada a variação flutuação, $F(n)$, como função do tamanho da caixa, n . É fácil ver na figura 9.1 que existem dois regimes que se encontram, isto ocorre em torno de $n \sim 10^2$. Antes deste ponto não existe nenhuma lei de escala aparente no gráfico. Observa-se, para n pequeno, que duas aglomerações dominantes estão presentes nos dados.

Para $n > 10^2$, $F(n)$ se comporta como n^α onde $\alpha \approx 1$ indicando uma correlação positiva forte na série temporal. Deve ser notado que, quando $\alpha = 1$, o expoente de auto-correlação é $\gamma = 0$ [veja a equação (9.2)]. Isto

também indica que a série temporal se comporta como um ruído do tipo $1/f^\beta$ [veja a equação (9.5) abaixo].

9.1.2 Densidade espectral de potência

A densidade espectral é freqüentemente usada para caracterizar séries temporais com ruído. No estudo presente utiliza-se a análise de densidade espectral de potência para encontrar a presença de ruído $1/f$, assim como predito pela análise de flutuação sem tendência.

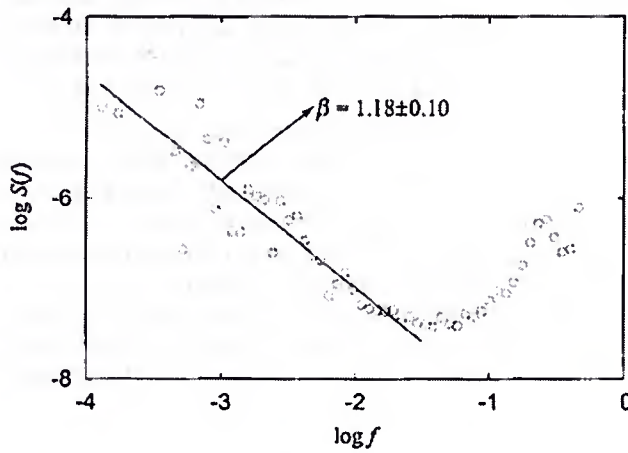


Figura 9.2: Densidade espectral de potência como função da frequência, mostrando espectro de ruído $1/f^\beta$ para a série temporal obtida da rede tridimensional de mapas acoplados.

A densidade espectral de potência para uma série temporal x é definida como:

$$S(f) = |\mathcal{F}(x)|^2, \quad (9.4)$$

onde $\mathcal{F}(x)$ denota a transformada de Fourier discreta da série temporal x . Para confirmar o resultado obtido na análise de flutuação sem tendência, busca-se por uma relação do tipo:

$$S(f) \sim \frac{1}{f^\beta}. \quad (9.5)$$

A figura 9.2 mostra a densidade espectral de potência como função da frequência. Aqui é tomada a média de $S(f)$ sobre uma faixa de frequências Δf . Da figura 9.2 pode ser visto que a densidade espectral de potência (para a rede tridimensional de mapas acoplados) cai quase como $1/f^\beta$, with $\beta \sim 1.18$. Isto é consistente como valor de α para a rede tridimensional na figura 9.1 e equação (9.3).

Feigenbaum [90] mostrou um exemplo de um sistema dinâmico determinístico com o comportamento de ruído $1/f$. Também foi confirmado teoricamente a existência da densidade espectral $1/f$ usando mapas quadráticos simples [91]. Estes fenômenos são fatos conhecidos, que não possuem nenhuma explicação aceitável, apesar da sua universalidade. Portanto, no contexto de sistemas dinâmicos, especialmente em sistemas caóticos, a intermitência é uma das fontes para o ruído $1/f$, e como consequência estes sistemas exibem correlação de longo alcance temporal e espacial [92].

Anteriormente, foi mostrado que o expoente de flutuação para redes unidimensionais de mapas acoplados é próximo de 0,5, similar àquele observado no ruído branco. Para redes bidimensionais, o expoente aumentou sensivelmente. Portanto, existe uma mudança dramática no expoente para redes tridimensionais, com $\alpha \approx 1$ [veja a figura 9.1]. Isto indica a presença de ruído $1/f$. A densidade espectral de potência da série temporal decresce da forma $1/f^\beta$ [veja a figura 9.2]. A estrutura de correlação nos dados, de acordo com a figura 9.1, afetou o espectro de potência, produzindo um desvio em relação à curva $1/f$.

9.2 Análise Dinâmica

9.2.1 Dimensão Local

No estudo de sistemas caóticos espaço-temporais, é interessante conhecer a dimensão local. Assim, para a rede de mapas acoplados, a cada ponto da rede está associada uma dimensão diferente. A dimensão local representa os graus de liberdade existentes naquela região da rede de mapas acoplados. Assim, quanto maior a dimensão local, menor o grau de previsibilidade naquela região. Portanto o conhecimento da dimensão local é de grande importância no entendimento e na caracterização do sistema em estudo.

Uma estatística de interesse usada para caracterizar sistemas caóticos espaço-temporais é a *dimensão do espaço dos vetores de propagação (bread vector dimension)* [94, 95, 96]. Pode-se analisar a dimensão local de um

sistema caótico espaço-temporal bidimensional usando a dimensão do espaço dos vetores de propagação. No presente estudo, esta dimensão é aplicada para as redes bidimensional e tridimensional de mapas acoplados. No caso da rede tridimensional, é considerada uma fatia bidimensional, com o objetivo de calcular a dimensão local.

O cálculo

O cálculo da dimensão do espaço dos vetores de propagação é feita tal como mostrado a seguir. Considere um sistema distribuído espacialmente em duas dimensões, cujo estado num instante t_1 é definido sobre uma coleção de pontos (i, j) . Aqui são considerados os $M - 1$ vizinhos mais próximos para cada ponto (i, j) numa rede quadrada, com $M = 25$. O mapa logístico é um sistema dinâmico com apenas uma variável; assim, para especificar o estado num ponto da rede, incluindo seus vizinhos, é necessário um vetor de estado M -dimensional, chamado de vetor de propagação.

Agora que um ponto da rede foi escolhido (em conjunto com seus vizinhos) num instante de tempo específico, é gerada uma perturbação no vetor de estado M -dimensional. Esta perturbação é feita no instante t_0 e o estado é evoluído por um passo de tempo. Verifica-se qual foi a propagação da perturbação e este vetor é normalizado. A normalização é adotada porque o único interesse é na direção de propagação da perturbação. Esta perturbação (propagada e normalizada) é usada para continuar propagando a perturbação. O processo deve ser repetido até que se chegue ao instante de tempo t_1 . Depois de algum transiente, a perturbação converge, assim é obtido um vetor de propagação. Devem ser obtidos k vetores de propagação. Portanto devem ser criadas novas perturbações, e todo o processo é repetido para cada perturbação, para que sejam obtidos os k vetores de propagação. A partir destes k vetores locais normalizados (módulo igual a um), escreve-se a matriz $M \times k$ B . Obtém-se assim a matriz de covariância $k \times k$ do sistema, que é dada por $C = B^T B$.

Os autovalores da matriz de covariância são ordenados, tal que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$. Então são definidos os valores singulares de B assim: $\sigma_l = \sqrt{\lambda_l}$, $l = 1, 2, \dots, k$. Os autovalores λ_l representam a quantidade de variância no conjunto dos k vetores de propagação unitários. Então a dimensão do espaço dos vetores de propagação pode ser definida como [94]:

$$\psi_{i,j}(\sigma_1, \sigma_2, \dots, \sigma_k) = \frac{\left(\sum_{l=1}^k \sigma_l\right)^2}{\sum_{l=1}^k \sigma_l^2}. \quad (9.6)$$

Como cada um dos k vetores bred possui norma igual a um, ψ assume valores no intervalo $[0, k]$. No presente estudo, o número de vetores de bred foi fixado em $k = 5$.

Para um entendimento do significado da dimensão do espaço dos vetores de propagação, serão apresentados alguns exemplos. Se metade dos vetores estiverem aproximadamente alinhados numa direção, enquanto que a outra metade estiverem aproximadamente alinhados numa outra direção, então a dimensão do vetor bred será aproximadamente 2. Se a maioria dos vetores bred estiverem alinhados numa direção, e apenas alguns estiverem alinhados numa segunda direção, então a dimensão do vetor bred assumirá um valor entre 1 e 2. O valor máximo ocorre se os k vetores bred estiverem apontando em diferentes direções, o valor da dimensão do vetor bred será k .

Resultados para redes de mapas acoplados

A dimensão local para a rede de mapas acoplados foi calculada nos casos bidimensional e tridimensional, considerando diferentes conjuntos de condições iniciais. No caso da rede bidimensional, a dimensão local obtida ficou entre um mínimo de 1,03 e um máximo de 2,4, enquanto que para a rede tridimensional a dimensão local assumiu valores entre um mínimo de $\sim 1,04$ e um máximo de $\sim 2,1$. Como este tipo de dimensão está relacionada com preditabilidade [95, 96], conclui-se que a rede tridimensional de mapas acoplados é mais previsível que a rede bidimensional.

9.2.2 Dimensão de imersão

O método desenvolvido por Cao [97] foi implementado para obter a dimensão de imersão para as redes de mapas acoplados. Nas simulações, foi observado que este procedimento foi mais robusto e objetivo do que os resultados ambíguos baseados na integral de correlação.

A idéia é começar com alguma dimensão de imersão inicial d e um vetor r_j do conjunto de vetores de estados. Um simples algoritmo de previsão, similar àquele usado no teste de predição cruzada da não-estacionaridade (veja a seguir), consiste em encontrar o vizinho mais próximo $r_{\alpha(j)}$ de r_j e então evolui-lo um passo de tempo à frente, obtendo $r_{\alpha(j)+1}$. O erro nesta previsão é $\|r_{j+1} - r_{\alpha(j)+1}\|$ e o erro sobre todo o conjunto de vetores de estado, $E(d)$, é simplesmente a soma de cada erro individual. Então a dimensão de imersão é o valor de d que minimiza a função $E(d)$.

Esta é uma forma simples e inteligente de determinar a dimensão de imersão, que não requer grandes conjuntos de dados*.

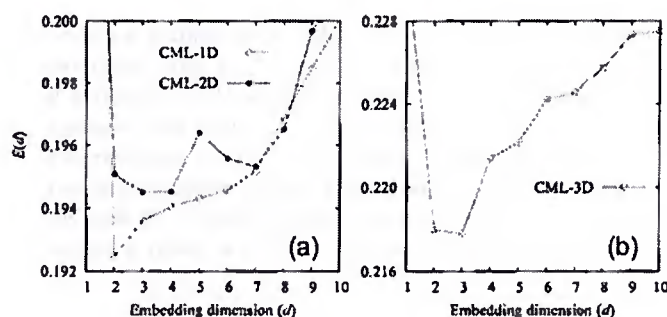


Figura 9.3: Estimativa da dimensão de imersão ótima para as séries temporais obtidas das redes de mapas acoplados com 1, 2 e 3 dimensões.

Na figura 9.3 é mostrada a dimensão de imersão das redes de mapas acoplados. Foi encontrado que a rede unidimensional possui a dimensão mais baixa, seguida de perto pelas redes bi e tridimensionais. Além disso, observa-se que todas estas redes possuem dimensão compatível com sistemas caóticos.

9.2.3 Estacionaridade e análise de recorrência

Um dos testes mais importantes para um sistema dinâmico é a respeito da sua estacionaridade. Foram implementados três diferentes algoritmos, pois não existe um único critério para estes testes em dinâmica não-linear.

O gráfico de recorrência foi introduzido por Eckmann [101]. A idéia é obter propriedades de sistemas dinâmicos multidimensionais por meio das características de um gráfico bidimensional. Esta idéia consiste em graficar os instantes quando a trajetória visita uma dada região do espaço de fase. Sistemas dinâmicos não-lineares determinísticos exibem a propriedade fundamental que os estados tendem a retornar à vizinhança das regiões já visitadas pela trajetória.

*Veja, por exemplo, Ref. [98] uma estimativa do tamanho de uma série temporal que é necessária para dar resultados significativos no cálculo da dimensão usando a integral de correlação. Também na Ref. [99] uma estimativa ainda mais estrita é dada.

Considere a seguinte função

$$R_{ij} = \Theta(\epsilon - \|r_i - r_j\|), \quad (9.7)$$

onde os vetores de estado r_i, r_j são considerados em instantes de tempo discretos, com $1 \leq i, j \leq N$, sendo N o número total de estados. Aqui Θ é a função de Heaviside e ϵ é um número positivo, cujo valor é discutido abaixo. Um gráfico é então composto associando um ponto no plano com coordenadas (i, j) se o argumento da função for positivo. O bem conhecido mapa do padeiro [102] é examinado para comparação, assim como o gráfico da rede de mapas acoplados para dimensão 1 a 3. Os valores de ϵ são escolhidos iguais a $\sqrt{d_E} \times 10\%$ da flutuação da série temporal, onde d_E é a

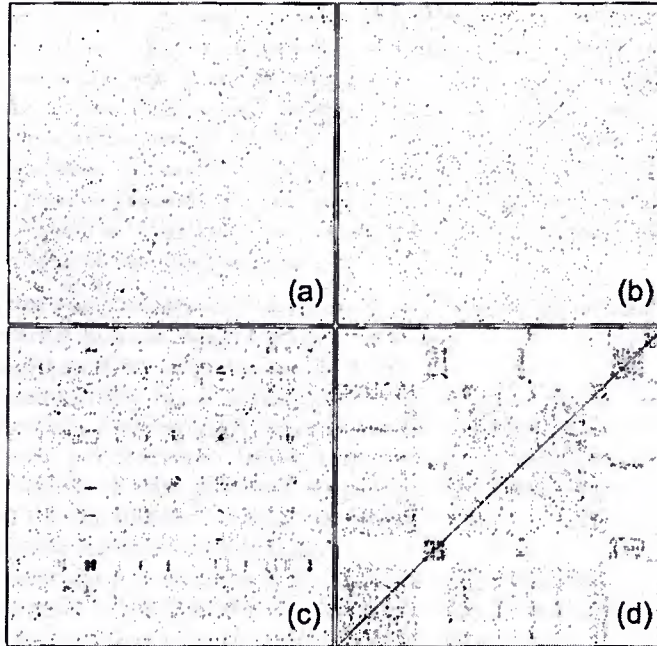


Figura 9.4: Recurrence plot for the (a) nonstationary baker's map [102], (b) 1D CML, (c) 2D CML and (d) 3D CML.

dimensão de imersão [100]. A dimensão de imersão para as redes é 3 e as análises gráficas não são alteradas quando este número é mudado. O critério

de não-estacionaridade é que a densidade de pontos nos cantos superior esquerdo e inferior direito do gráfico de recorrência deveriam diminuir [103]. Como pode ser visto nos gráficos, o resultado é inconclusivo na rede tridimensional. As demais redes (uni e bidimensional), de acordo com este critério, são estacionárias.

O método discutido por Gao [104] usa a propriedade de que, em sistemas dinâmicos não-lineares, o tempo de recorrência, adequadamente calculado, é aproximadamente constante para uma dinâmica estacionária. A idéia é calcular este tempo para uma trajetória, dada uma condição inicial dentro de uma bola de raio $r > 0$ (mais especificamente, tempo de recorrência do segundo tipo). Considera-se uma dimensão de imersão 2, mas os resultados são robustos com respeito a mudança neste parâmetro. Para eliminar a dependência nas condições iniciais, um certo procedimento é necessário. Impondo-se a normalização do tempo de retorno, então diversas condições iniciais são tomadas. Uma bola de raio r é centrada em cada condição inicial, e o tempo para o primeiro retorno para cada bola é calculado. O primeiro tempo de retorno é a média destes. Então o segundo, terceiro, etc, tempos de retorno são calculados. O critério para estacionaridade é que o gráfico do tempo necessário para os retornos deveria ser constante como uma função do número de recorrências. Conclui-se, de acordo com este método, que as redes de mapas acoplados são estacionárias.

Outro teste para estacionaridade [102] requer o particionamento da série temporal e a implementação de algum algoritmo de previsibilidade (ou qualquer outra estatística sensível a diferenças na dinâmica). O algoritmo é usado como um critério para avaliar como os dados de um subconjunto da partição podem ser usados para prever outros subconjuntos (daí o nome predição cruzada, que é usado por Schreiber [102]). Aqui a escolha da dimensão de imersão não afeta as conclusões principais, assim ela foi escolhida como 3, com o atraso igual a 1. Suponha que a série temporal x_n , $n = 1, 2, \dots, N$, seja dividida em segmentos contínuos de tamanho l e que uma estatística γ_{ij} seja construída para cada par de segmentos (i, j) . Para o presente propósito, foi escolhido um algoritmo de previsão simples sobre os vetores de estado [105] no qual γ_{ij} é representado como uma superfície de erro sobre a grade (i, j) , ou seja, o erro de usar o segmento i para prever a dinâmica do segmento j . Um gráfico desta superfície dá uma idéia de se diferentes dinâmicas estão ocorrendo em diferentes segmentos: quando o erro de previsão é pequeno, então os segmentos possuem dinâmicas similares. Para sistemas não-estacionários é de se esperar que a superfície de erro possua um valor baixo na diagonal, pois um segmento está sendo usado para prever a sua própria dinâmica (predição dentro da amostra). Para sistemas estacionários, a dinâmica de-

veria ser similar em todos os segmentos e uma superfície de erro irregular como resultado. Das várias superfícies de erro analisadas, não há evidência de não-estacionaridade nas redes de mapas acoplados.

Capítulo 10

Conclusões

10.1 Reamostragem

Verificou-se ao longo deste trabalho, que o algoritmo desenvolvido consegue produzir séries temporais que possuem um atrator similar ao atrator da série original. A inspeção da similaridade entre os espaços de fase foi realizada visualmente, pois não existe nenhum método quantitativo geral para esta finalidade.

Também foi observado que o algoritmo é capaz de realizar a reamostragem com um número de vizinhos bastante inferior ao algoritmo proposto por Golia e Sandri [6, 7]. Enquanto o algoritmo que eles utilizaram necessitava de 80 vizinhos para reamostrar a série de Lorenz, o algoritmo que aqui foi desenvolvido precisava de algo em torno de uma dezena de vizinhos. Na reamostragem que foi apresentada aqui, utilizou-se apenas 9 vizinhos.

Além disso, foi verificado que o algoritmo de reamostragem consegue reproduzir a distribuição para o expoente de Lyapunov. A comparação das duas distribuições (aquela baseada em várias realizações do sistema original e aquela baseada nas reamostragens) mostrou que a reamostragem reproduz os resultados obtidos por várias realizações.

Portanto o algoritmo de reamostragem pode ser utilizado para se obter a incerteza e intervalos de confiança associados ao expoente de Lyapunov.

10.1.1 Trabalho futuro

No futuro pretende-se realizar um estudo envolvendo séries empíricas, para verificar o funcionamento do algoritmo e para descobrir se ele é capaz de reproduzir a dinâmica de séries obtidas de sistemas reais.

10.2 Redes de mapas acoplados

Conforme está mostrado aqui, assim como no trabalho de Muruganandam et al. [106], foi feita uma análise comparativa das séries temporais caóticas geradas pelas redes de mapas acoplados com dimensão 1, 2 e 3 baseadas em certas ferramentas frequentemente empregadas, tal como a análise de flutuação, densidade espectral de potência, dimensão local e análise de quantificação da recorrência.

Foi encontrado que o gráfico da flutuação para a rede tridimensional mostra um comportamento anômalo, tal que nenhuma região de escala pode ser encontrada para n menor que aproximadamente 400. Entretanto, existe uma ampla região de escala tanto para o mapa unidimensional como para o mapa bidimensional. Além disso, ocorre um aumento notável na persistência se for aumentada a dimensão. Em outras palavras, a rede tridimensional de mapas acoplados é mais persistente do que as redes com dimensão 1 e 2. Outras propriedades importantes analisadas revelaram que as redes de mapas acoplados possuem dimensão baixa e, no caso tridimensional, mostra uma clara indicação de ruído $1/f$. Todas as simulações foram replicadas em diferentes pontos das redes, e as conclusões são robustas a este respeito.

Se for analisada a série temporal como um todo, então os testes na seção 9.2.3 mostram que o sistema é estacionário. Entretanto, no gráfico da série temporal (figura 8.2(c)) é evidente que deve haver alguma mudança da variância ao longo da evolução. Verificou-se assim que ocorre a não-estacionaridade para valores suficientemente baixos de n , que pode ser vista como uma quebra na escala do gráfico de flutuação.

10.2.1 Trabalho futuro

O próximo passo a ser realizado é um amplo estudo sobre a previsibilidade de uma série temporal gerada por uma rede de mapas acoplados usando algoritmos especificamente designados para sistemas caóticos de baixa dimensão. Além disso, deve ser feito um estudo utilizando outros regimes para

os parâmetros das redes de mapas acoplados. Principalmente onde ocorra intermitência espaço-temporal.



Índice Remissivo

- aprendizado
 - ansioso, 31
 - preguiçoso, 31
- arquitetura, 31, 36
- atraso temporal, 21
- atrator, 13
 - estranho, 13
- bijetora, 21
- campo vetorial, 12, 16
- difecomorfismo, 14
- dimensão
 - de capacidade, 13, 21
 - de inersão, 8
 - de mergulho, 21
- dimensão de contagem de caixa, 13
- dimensão do espaço dos vetores de propagação, 98
- dimensão do núcleo da matriz, 58
- distância de sombreamento, 71
- equivalência
 - diferenciável, 17
 - topológica, 17
- equivalência diferenciável, 16
- equivalência topológica, 16
- espaço de fase, 11
 - reconstruído, 18
- espaço de fases
 - reconstruído, 19
- espectro de potências, 24
 - contínuo, 24
- filtro
 - passa-baixa, 24
- fluxo, 13
- fractal, 13
- função
 - de autocorrelação, 8
 - mensuração, 18
- graus de liberdade, 11
- graus de liberdade ativos, 21
- hardware, 36
- homeomorfismo, 14, 17
- informação mútua, 8
- kernel da matriz, 58
- média ponderada, 38
- método de Takens, 20
- matriz
 - imagem, 58
 - posto, 58
- modelo
 - preguiçoso global, 32
- morfismo, 14
- núcleo da matriz, 58
- parâmetros, 31, 36
- ponto fixo
 - foco, 17
 - nó, 17
 - nó impróprio, 17
- problema de valor inicial, 13

- pseudo-órbita, 71
- pulso aleatório, 71
- regressão ponderada local, 38
- séries temporais
 - análise, 1
 - classificação, 1
 - previsão, 1
- sensibilidade às condições iniciais, 13
- singular
 - numericamente quase singular, 57
 - valor singular, 57
 - vetor singular, 57
- subespaço do ruído, 62
- subespaço do sinal, 62
- variedade, 63
- vetor tangente, 12

Referências Bibliográficas

- [1] T. Schreiber, *Interdisciplinary Application of Nonlinear Time Series Methods*, Physics Reports **308**, 1-64 (1999).
- [2] A. Wolf, J. B. Swift, H. L. Swinney, J. A. Vastano, *Determining Lyapunov exponents from a time series*, Physica D **16**, 285-317 (1992).
- [3] M. T. Rosenstein, J. J. Collins, C. J. De Luca, *A practical method for calculating largest Lyapunov exponents from small data sets*, Physica D **65**, 117-134 (1993).
- [4] P. Grassberger, I. Procaccia, *Characterization of Strange Attractors*, Phys. Rev. Lett. **50**, 346-349 (1983).
- [5] R. Gençay, *A statistical framework for testing chaotic dynamics via Lyapunov exponents*, Physica D **89**, 261-266 (1996).
- [6] S. Golia, M. Sandri, *Resampling Chaotic Time Series*, Physical Review Letters **78** 4197-4200 (1997).
- [7] S. Golia, M. Sandri, *A Resampling Algorithm for Chaotic Time Series*, Statistics and Computing **11**, 241-255 (2001).
- [8] E.J. Kostelich, *Bootstrap Estimates of Chaotic Dynamics*, Physical Review E **64** 16213 (2001).
- [9] T. Schreiber, H. Kantz, *Observing and predicting chaotic signals: Is 2% noise too much?*, em Y. Kravtsov, J. Kadtko (editores), *Predictability of Complex Dynamical Systems*, Springer, New York (1996).
- [10] D.S. Broomhead, G.P. King, *Extracting dynamics from experimental data*, Physica D **20**, 217-236 (1986).
- [11] V.I. Arnold, *Ordinary Differential Equations*, MIT Press, Cambridge (1978).

- [12] E. Ott, *Chaos in dynamical systems*, Cambridge Univ. Press, Cambridge (1993).
- [13] M.W. Hirsch, S. Smale, R.L. Devaney, *Differential Equations, Dynamical Systems and an Introduction to Chaos*, Academic Press, California (2004).
- [14] M.W. Hirsch, S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, California (1974).
- [15] J.P. Eckmann, D. Ruelle, *Ergodic theory of chaos and strange attractors*, Review of Modern Physics (57) No. 3, Part I, 617-656 (1985).
- [16] A. Asperti, G. Longo, *Category Types and Structures: An Introduction to Category Theory for the working computer scientist*, MIT Press, Cambridge (1991).
- [17] E. Lorenz, *Deterministic Nonperiodic Flow*, J. Atmos Sci. 20 130-141 (1963).
- [18] N. Packard, J. Crutchfield, D. Farmer, R. Shaw, *Geometry from a time series* Physical Review Letters 45, 712 (1980).
- [19] F. Takens, *Detecting strange attractors in turbulence*, in Lectures Notes in Mathematics, D.A. Rand, L.S. Young, eds, Springer, Berlin (1981).
- [20] T. Sauer, J. Yorke, M. Casdagli, *Embedology*, J. Stat. Phys. 65, 579 (1991).
- [21] A. M. Fraser and H. L. Swinney, *Independent coordinates for strange attractors from mutual information*, Phys. Rev. A 33, 1134 (1986).
- [22] M. B. Kennel, R. Brown, H. D. I. Abarbanel, *Determining embedding dimension for phase-space reconstruction using a geometrical construction*, Phys. Rev. A 45, 3403 (1992).
- [23] H. Kautz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge Univ. Press, Cambridge (1997).
- [24] H. Whitney, *Differentiable manifolds*, Ann. Math. 37, 645 (1936).
- [25] E. Ott, T. Sauer, J.A. Yorke *Coping with chaos: Analysis of chaotic data and the exploitation of chaotic systems*, Wiley, New York (1994).
- [26] S.W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, 2nd edition, California Technical Publishing, San Diego (1999).

- [27] T. Sauer, *Noise reduction for nonlinear systems*, Physica D 58, 193-201 (1992).
- [28] R.H. Landau, M.J Páez, *Computational Physics Problem Solving with Computers*, John Wiley & Sons, New York (1997).
- [29] A.S. Weigend, N.A. Gershenfeld (editores), *Time Series Prediction*. Addison-Wesley Publishing Company, 1994.
- [30] A.S. Weigend, N.A. Gershenfeld, The future of time series. Em A.S. Weigend, N.A. Gershenfeld (editores), *Time Series Prediction*. Addison-Wesley Publishing Company, 1994.
- [31] E.A. Wan, Time series prediction by using a connectionist network with internal delay lines. Em A.S. Weigend, N.A. Gershenfeld (editores), *Time Series Prediction*, Santa Fe Institute Studies in the Sciences of Complexity, páginas 195-217. Addison-Wesley, 1994.
- [32] Tim Sauer, Time series prediction by using delay coordinate embedding. Em A.S. Weigend, N.A. Gershenfeld (editores), *Time Series Prediction*, Santa Fe Institute Studies in the Sciences of Complexity, 175-193. Addison-Wesley, 1994.
- [33] J.A.K. Suykens and J. Vandewalle (editores), *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modelings: Theory and Applications*, Katholieke Universiteit, Leuven, Belgium, Jullho de 1998.
- [34] J. McNames, J. Suykens, J. Vandewalle, *Winning Entry of the K.U. Leuven time-series prediction competition*, International Journal of Bifurcation and Chaos, Vol.9, No.8, pp.1485-1500 (1999).
- [35] C. M. Bishop, *Neural Networks for Statistical Pattern Recognition*, Oxford University Press, Oxford (1994).
- [36] T. Schreiber, *Detecting and analysing nonstationarity in a time series with nonlinear cross-predictions*, Phys. Rev. Lett. 78, 843 (1997).
- [37] A. Provenzale, L. A. Smith, R. Vio, and G. Murante, *Distinguishing between low-dimensional dynamics and randomness in measured time series*, Physica D 58, 31 (1992).
- [38] G.H. Golub, C.F. Van Loan, *Matrix Computation*, The Johns Hopkins, Baltimore(1996).

- [39] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge (2000).
- [40] T. Schreiber, Efficient neighbor searching in nonlinear time series analysis, *Int. J. Bifurcation and Chaos* 5, 349 (1995).
- [41] C. G. Atkeson, A. W. Moore, S. Schaal, Locally Weighted Learning, *Artificial Intelligence Review* 11, 11-73 (1997).
- [42] Aha, Editorial, *Artificial Intelligence Review* 11, 7-10 (1997).
- [43] G. Bontempi, M. Birattari, H. Bersini, *Lazy Learning: A local method for supervised learning*, em L. C. Jain, J. Kacprzyk (editores) *New Learning Paradigms in Soft Computing*, Studies in Fuzziness and Soft Computing, vol. 84, pp. 97-137. Springer, Heidelberg, Alemanha (2001).
- [44] Martin Casdagli, Deirdre Des Jardins, Stephen Eubank, J. Doyne Farmer, John Gibson, James Theiler, Nonlinear modeling of chaotic time series: Theory and applications. Em Jong Hyun Kim and John Stringer (editores), *Applied Chaos*, páginas 335-380. John Wiley & Sons, Boston, (1992).
- [45] J. D. Farmer, John J. Sidorowich. Exploiting chaos to predict the future and reduce noise. Em Yee Chung Lee, editor, *Evolution, Learning and Cognition*, páginas 277-330. World Scientific, Singapore, (1988).
- [46] J. Doyne Farmer, John J. Sidorowich. Predicting chaotic time series. *Physical Review Letters* 59, 845-848, (1987).
- [47] Zhong Liu, Xiaolin Ren, and Zhiwen Zhu. Equivalence between different local prediction methods of chaotic time series. *Physics Letters A* 227, 37-40, (1997).
- [48] Martin C. Casdagli and Andreas S. Weigend. Exploring the continuum between deterministic and stochastic modeling. Em Andreas S. Weigend, Neil A. Gershenfeld (editores), *Time Series Prediction*, Santa Fe Institute Studies in the Sciences of Complexity, 347-366, Addison-Wesley (1994).
- [49] D. Kugiuntzis Lillekjendlic, N. Christophersen. Chaotic time series. Part II. System identification and prediction. *Modeling, Identification and Control*, 15(4):225-243, (1994).

- [50] Martin C. Casdagli and Andreas S. Weigend. Exploring the continuum between deterministic and stochastic modeling. In Andreas S. Weigend, Neil A. Gershenfeld (editores), *Time Series Prediction*, Santa Fe Institute Studies in the Sciences of Complexity, pages 347-366. Addison-Wesley, 1994.
- [51] C. Ziehmann, L.A. Smith, J. Kurths, *The bootstrap and Lyapunov exponent in deterministic chaos*, Physica D **126**, 49-59 (1999).
- [52] T. Sauer, C. Grebogi, J.A. Yorke, *How long do numerical solutions remain valid?*, Phys. Rev. Lett **79**, 59-62 (1997).
- [53] J.D. Farmer, J.J. Sidorowich, *Optimal shadowing and noise reduction*, Physica D **47**, 373-392 (1991).
- [54] S. Giannerini, R. Rosa, *Generating replications of chaotic time series*, Nonlinear Dynamics, Psychology, and Life Sciences **5**, 77-87 (2001).
- [55] S. Giannerini, R. Rosa, *New resampling method to assess the accuracy of the maximal Lyapunov exponent estimation*, Physica D **155**, 101-111 (2001).
- [56] S. Giannerini, R. Rosa, *Assessing chaos in time series: statistical aspects and perspectives*, Studies in Nonlinear Dynamics & Econometrics **8**, 1215 (2004).
- [57] U. Lall, A. Sharma, *A nearest neighbor bootstrap for resampling hydrologic time series*, Water Resources Research **32**, 679-693 (1996).
- [58] Devaney, R. L. *Introduction to Chaotic Dynamical Systems*, Boulder, CO: Westview Press (1989).
- [59] Banks, J. et al. *On Devaney's definition of chaos*, Amer. Math. Monthly. **99**, 332 (1992).
- [60] H. Kantz, *A robust method to estimate the maximal Lyapunov exponent of a time series*, Phys. Lett. A **185**, 77 (1994).
- [61] M. Menezes, G. Francisco, *Resampling and Lyapunov exponent in deterministic time series*, IFT Preprint (2009).
- [62] P. Grassberger, T. Schreiber, *Phase-transitions in coupled map lattices*, Physica D **50** (2) 177-188 (1991).

- [63] M. C. Cross, P. C. Hohenberg, Pattern-formation outside of equilibrium, *Rev. Mod. Phys.* **65** (3) 851-1112 (1993).
- [64] P. Marcq, H. Chate, P. Manneville, Universal critical behavior in two-dimensional coupled map lattices, *Phys. Rev. Lett.* **77** (19) 4003-4006 (1996).
- [65] C. C. Brewster, J. C. Allen, Spatiotemporal model for studying insect dynamics in large-scale cropping systems, *Environ. Entomol.* **26** (3) 473-482 (1997).
- [66] J. P. Lachaux, L. Pezard, L. Garnero, C. Pette, B. Renault, F. J. Varela, J. Martinerie, Spatial extension of brain activity fools the single-channel reconstruction of EEG dynamics, *Hum. Brain Mapp.* **5** (1) 26-47 (1997).
- [67] T. Yanagita, K. Kaneko, Rayleigh-Benard convection patterns, chaos, spatiotemporal chaos and turbulence, *Physica D* **82** (3) 288-313 (1995).
- [68] P. W. Colovas, C. D. Andereck, Turbulent bursting and spatiotemporal intermittency in the counterrotating taylor-couette system, *Phys. Rev. E* **55** (3) 2736-2741 (1997).
- [69] Y. Cuche, R. Livi, A. Politi, Phase transitions in 2D linearly stable coupled map lattices, *Physica D* **103** (1-4) 369-380 (1997).
- [70] M. J. Keeling, I. Mezic, R. J. Hendry, J. McGlade, D. A. Rand, Characteristic length scales of spatial models in ecology via fluctuation analysis, *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **352** (1361) 1589-1601 (1997).
- [71] A. Hilgers, C. Beck, Hierarchical coupled map lattices as cascade models for hydrodynamical turbulence, *Europhys. Lett.* **45** (5) 552-557 (1999).
- [72] P. G. Lind, J. A. M. Corte-Real, J. A. C. Gallas, Using advection to control the velocity of patterns in rings of maps, *Physica D* **168** 93-105 (2002).
- [73] K. Kaneko, Spatiotemporal chaos in one-dimensional and two-dimensional coupled map lattices, *Physica D* **37** (1-3) 60-82 (1989).
- [74] T. Bohr, O. B. Christensen, Size dependence, coherence, and scaling in turbulent coupled-map lattices, *Phys. Rev. Lett.* **63** (20) 2161-2164 (1989).

- [75] P. Palaniyandi, P. Muruganandam, M. Lakshmanan, Desynchronized wave patterns in synchronized chaotic regions of coupled map lattices, *Phys. Rev. E* **72** 037205 (2005).
- [76] L.H.A. Monteiro, *Sistemas Dinâmicos*, Editora Livraria da Física, São Paulo (2002).
- [77] B. B. Mandelbrot, J. Berger, J.-P. Kahane, J. Peyriere, *Multifractals and $1/f$ Noise: Wild Self-Affinity in Physics*, Springer, New York, 1999.
- [78] B. Mandelbrot, F. Damerau, M. Frame, K. McCamy, *Gaussian Self-Affinity and Fractals*, Springer, New York, 2001.
- [79] M. S. El Naschie, The VAK of vacuum fluctuation, spontaneous self-organization and complexity theory interpretation of high energy particle physics and the mass spectrum, *Chaos, Solitons & Fractals* **18** 401 (2003).
- [80] M. S. El Naschie, Elementary prerequisites for E-infinity, *Chaos, Solitons & Fractals* **30** 579 (2006).
- [81] H. Chate, P. Manneville, Emergence of effective low-dimensional dynamics in the macroscopic behavior of coupled map lattices, *Europhys. Lett.* **17** (4) 291–296 (1992).
- [82] H. Chate, P. Manneville, Collective behaviors in spatially extended systems with local interactions and synchronous updating, *Prog. Theor. Phys.* **87** (1) 1–60 (1992).
- [83] R. F. Voss, J. Clarke, Flicker ($1/f$) noise: equilibrium temperature and resistance fluctuations., *Phys. Rev. D* **13** 556 (1976).
- [84] B. Pilgram, D. T. Kaplan, A comparison of estimators for $1/f$ noise, *Physica D* **114** 108 – 122 (1998).
- [85] P. A. Rikvold, R. K. P. Zia, Punctuation equilibria and $1/f$ noise in a biological model with individual-based dynamics, *Phys. Rev. E* **68** 031913 (2003).
- [86] C.-K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. S. M. Simons, H. E. Stanley, Long-range correlations in nucleotide sequences, *Nature (London)* **356** 168 (1992).

- [87] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, A. L. Goldberger, Mosaic organization of DNA nucleotides, *Phys. Rev. E* **49** 1685 (1994).
- [88] C.-K. Peng, J.M. Hausdorff, A.L. Goldberger, Fractal mechanisms in neural control: Human heartbeat and gait dynamics in health and disease. Em: J. Walleczek (editor), *Nonlinear Dynamics, Self-Organization, and Biomedicine*, Cambridge, Cambridge University Press, (1999).
- [89] K. Hu, P. C. Ivanov, Z. Chen, P. Carpena, H. E. Stanley, Effect of trends on detrended fluctuation analysis, *Phys. Rev. E* **64** 011114 (2001).
- [90] M. J. Feigenbaum, Universal behavior in nonlinear systems, *Physica D: Nonlinear Phenomena* **7** (1-3) 16-39 (1983).
- [91] I. Procaccia, H. Schuster, Functional renormalization-group theory of universal $1/f$ noise in dynamical systems, *Phys. Rev. A* **28** (2) 1210 (1983).
- [92] J. D. Keeler, J. D. Farmer, Robust space-time intermittency and $1/f$ noise, *Physica D* **23** (1-3) 413-435 (1986).
- [93] U. Frisch, *Turbulence*, Cambridge University Press, Cambridge, 1995.
- [94] D. J. Patil, B. R. Hunt, E. Kalnay, J. A. Yorke, E. Ott, Local low dimensionality of atmospheric dynamics, *Phys. Rev. Lett.* **86** (26) 5878 (2001).
- [95] G. Francisco, P. Muruganandam, Local dimension and finite time prediction in spatiotemporal chaotic systems, *Phys. Rev. E* **67** (6) 066204 (2003).
- [96] P. Muruganandam, G. Francisco, Local dimension and finite time prediction in coupled map lattices, *Pramana-J. Phys.* **64** 381-387 (2005).
- [97] L. Cao, A. S. Soofi, Nonlinear deterministic forecasting of daily dollar exchange rate, *Int. J. Forecasting* **15** 421-430 (1999).
- [98] J.-P. Eckmann, D. Ruelle, Fundamental limitations for estimating dimensions and lyapunov exponents in dynamical systems, *Physica D: Nonlinear Phenomena* **56** (2-3) 185-187 (1992).

- [99] A. Stefanovska, S. Strle, P. Kroselj, On the overestimation of the correlation dimension, *Phys. Lett. A* **235** (4) 24–30 (1997).
- [100] C. Letellier, Estimating the shannon entropy: Recurrence plots versus symbolic dynamics, *Phys. Rev. Lett.* **96** (25) 254102–4 (2006).
- [101] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems, *Europhys. Lett.* **4** 973–977 (1987).
- [102] T. Schreiber, Detecting and analyzing nonstationarity in a time series using nonlinear cross predictions, *Phys. Rev. Lett.* **78** (5) 843 (1997).
- [103] M. Marwan, Ph.D. thesis, University of Potsdam (2003).
- [104] J. B. Gao, Recurrence time statistics for chaotic systems and their applications, *Phys. Rev. Lett.* **83** (16) 3178–3181 (1999).
- [105] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, 2nd Edition, Cambridge University Press, Cambridge (2003).
- [106] P. Muruganandam, G. Francisco, M. Menezes, F.F. Ferreira, *Low dimensional behavior in three-dimensional coupled map lattices* *Chaos Solitons and Fractals* - (2009).

