

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Programa de Pós-Graduação em Matemática

DISSERTAÇÃO DE MESTRADO

**Cálculo Numérico da Forma Normal de
Floquet e Aplicações em Controle de Sistemas
Dinâmicos**

Marcus Vinicius de Oliveira

Belo Horizonte
2013

Cálculo Numérico da Forma Normal de Floquet e Aplicações em Controle de Sistemas Dinâmicos

Marcus Vinicius de Oliveira

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Mestre em Matemática

Orientadora: Dr^a. Sylvie Marie Oliffson Kamphorst Leal da Silva

BANCA EXAMINADORA

Dr^a. Sylvie Marie Oliffson Kamphorst Leal da Silva - DMAT-UFMG

Dr. Armando Gil Magalhães Neves - DMAT-UFMG

Dr^a. Sônia Pinto de Carvalho - DMAT-UFMG

Dr. Ricardo Hiroshi Caldeira Takahashi - DMAT-UFMG

Belo Horizonte, 08 de Novembro de 2013

Aos meus pais

Agradecimentos

A Deus por ter me permitido chegar até aqui e me guiado durante todo o caminho.

Aos meus pais, Mário e Dirce, e minha irmã, Kênia, pelo apoio e carinho durante toda a minha vida.

À professora Sylvie Oliffson Kamphorst, pela orientação e paciência durante todo o programa, em especial neste trabalho.

Às pessoas que tornaram o Programa de Iniciação Científica e Mestrado - PICME uma realidade, em especial aos seus coordenadores e ao Programa de Pós-Graduação em Matemática da UFMG.

Aos colegas dos cursos de graduação em Engenharia Elétrica e mestrado em Matemática da UFMG, pelo apoio durante as jornadas diárias na Universidade.

Aos funcionários, monitores e coordenadores do Centro de Cálculo Eletrônico da UFMG.

A todos os professores com os quais eu tive a honra de aprender.

Aos membros da banca pela gentileza de participar da avaliação deste trabalho.

A Petrobras pela liberação para a defesa sem restrições.

A todos que, por esquecimento, não tenham sido citados.

"O único lugar em que sucesso vem antes de trabalho é no dicionário"

Albert Einstein

Resumo

Este trabalho apresenta um método numérico rigoroso para calcular a Forma Normal de Floquet $X(t) = Q(t)e^{tR}$ de uma solução Fundamental de Equações Diferenciais Ordinárias Lineares τ -periódicas $x' = A(t)x$, com $A(t + \tau) = A(t) \forall t \in \mathbb{R}$. Este problema pode ser resolvido através da solução de uma equação do tipo $f(x) = 0$ em que $f : \Omega^s \rightarrow \Omega^s$, com Ω^s um espaço de Banach adequadamente definido. Dessa forma, o método apresentado neste trabalho busca obter uma solução aproximada \bar{x} para $f(x) = 0$ e $r > 0$ tal que seja garantida a existência de $x^* \in B_{(\bar{x}, 0^\infty)}(r) \subset \Omega^s$ com $f(x^*) = 0$. A técnica baseia-se na definição de um operador $T : \Omega^s \rightarrow \Omega^s$ cujos pontos fixos são soluções da equação $f(x) = 0$ e tal que $T : B_{(\bar{x}, 0^\infty)}(r) \rightarrow B_{(\bar{x}, 0^\infty)}(r)$ seja uma contração, possibilitando a aplicação do Teorema do Ponto Fixo de Banach. Assim, garante-se a existência de ponto fixo de T e, conseqüentemente, a existência de uma solução para $f(x) = 0$, em $B_{(\bar{x}, 0^\infty)}(r)$. Para calcular r é utilizada a técnica dos Polinômios *Radii*, apresentada em [11], [27] e [5]. Para a realização dos cálculos é aplicada a Aritmética de Intervalos, de forma a garantir que erros de arredondamento não comprometam o rigor do método. Um algoritmo escrito em MATLAB[®] 2008 foi utilizado para implementação prática.

São apresentados exemplos de aplicação da Forma Normal de Floquet no Controle de dois sistemas dinâmicos, o Pêndulo Forçado e o Oscilador Duffing. O objetivo é fazer com que órbitas desses sistemas se aproximem assintoticamente de uma trajetória periódica desejada $y : \mathbb{R} \rightarrow \mathbb{R}$, ou seja, $\lim_{t \rightarrow \infty} [x(t) - y(t)] = 0$. Esse problema é constantemente substituído em Engenharia Elétrica por uma versão linearizada, o que conduz à utilização da Forma Normal de Floquet. O algoritmo desenvolvido como implementação do método demonstrado no Capítulo 2 é utilizado para obter a forma normal de Floquet das matrizes fundamentais das equações obtidas no Capítulo 4. Embora trate-se de um problema de controle baseado numa aproximação linear, a aplicação de um algoritmo rigoroso serve como motivação de aplicação.

O Capítulo 1 apresenta uma introdução à Teoria da Floquet e resultados utilizados no restante do trabalho. O Capítulo 2 contém a descrição do método numérico de cálculo da Forma Normal de Floquet e sua demonstração. O Capítulo 3 apresenta detalhes da implementação do método em um algoritmo computacional. O Capítulo 4 contém a aplicação da Forma Normal de Floquet ao controle do Pêndulo Forçado e do Oscilador Duffing, como citado anteriormente. Nos apêndices são apresentados conceitos e demonstrações auxiliares utilizadas ao longo do trabalho, inclusive uma generalização da Teoria de Floquet para uma classe de sistemas não lineares

Palavras-chave: Teoria de Floquet, Sistemas Lineares Periódicos, Método Numérico Rigoroso, Polinômios *Radii*, Controle.

Abstract

This work presents a rigorous numerical method to compute the Floquet Normal Form $X(t) = Q(t)e^{tR}$ for a given Fundamental Solution of a τ -periodic Linear Differential Equation. This problem is replaced by solving an equation $f(x) = 0$ such that $f : \Omega^s \rightarrow \Omega^s$, is defined in a suitable Banach Space Ω^s . The Method aims to find an approximate solution \bar{x} to the equation $f(x) = 0$ and $r > 0$ such that there exists a point $x^* \in B_{(\bar{x}, 0^\infty)}(r) \subset \Omega^s$, $f(x^*) = 0$. The technique is based on the definition of an operator $T : \Omega^s \rightarrow \Omega^s$ whose fixed points are solutions of the equation $f(x) = 0$. Thus, the numerical technique allows to calculate r so that $T : B_{(\bar{x}, 0^\infty)}(r) \rightarrow B_{(\bar{x}, 0^\infty)}(r)$ is a contraction, providing conditions to apply the Banach Fixed Point Theorem to ensure the existence of a fixed point of the operator T , and hence the existence of a solution to the equation $f(x) = 0$ in $B_{(\bar{x}, 0^\infty)}(r)$. *Radii* Polinomials, presented in [11], [27] and [5], are used to compute r . In order to prevent a loss of accuracy due to rounding errors, we use interval arithmetic for performing the calculations. A MATLAB[®] 2008 code is used to implement the numerical method described.

It's also addressed the application of the Floquet Normal Form in the control of two classical dynamic systems, Forced Pendulum and Duffing Oscillator. The goal is that the orbits of these systems become asymptotically close to a desired trajectory $y : \mathbb{R} \rightarrow \mathbb{R}$, $\lim_{t \rightarrow \infty} [x(t) - y(t)] = 0$. This problem is often replaced in Electrical Engineering by its linearized version, which leads to the use of Floquet Normal Form.

Chapter 1 provides an introduction to the Floquet theory and other results used in the rest of the work. Chapter 2 contains a description of the numerical method for calculating the Floquet normal form and its proof. Chapter 3 provides details of the method as a computational algorithm. Chapter 4 contains the application of Floquet normal form to the control of the Forced Pendulum and the Duffing Oscillator, as previously mentioned. Appendices present concepts and auxiliary statements used throughout the work.

Key Words: Floquet Theory, Periodic Linear Systems, Rigorous Numerical Method, *Radii* Polinomials, Control.

Sumário

1	Introdução	1
1.1	Resultados Fundamentais sobre EDO's não-autônomas	1
1.2	Fundamentos da Teoria de Floquet	4
2	Cálculo Numérico da Forma Normal de Floquet	10
2.1	Estratégia de Solução do Problema	10
2.2	Métodos <i>Quasi-Newton</i> para Solução de Equações em Espaços de Banach	13
2.3	Construção do Espaço de Banach Utilizado	14
2.4	Definição do Operador T e Condições para Existência de Ponto Fixo	17
2.5	Polinômios <i>Radii</i>	24
2.6	Cálculo de Y e Z	25
3	Detalhes de Implementação	31
3.1	O Algoritmo para Cálculo da Forma Normal de Floquet	31
3.2	Cálculo do Ponto Inicial	33
3.3	Cálculo dos Polinômios <i>Radii</i>	35
3.4	Aritmética de Intervalos	36
3.5	Resultados experimentais	37
4	Aplicações em Controle de Sistemas Periódicos	40
4.1	Controle Linear do Oscilador de Duffing	40
4.2	Controle Linear do Pêndulo Forçado Periodicamente	44
5	Conclusões	48
	<u>APÊNDICES</u>	48
A	Logaritmo de Matrizes	49
B	Teoria de Floquet Generalizada	54
C	Séries de Fourier	57
D	Método de Runge-Kutta para solução aproximada de EDOs	62
E	Introdução à Aritmética Intervalar	64
E.1	Aritmética Intervalar Real	64
E.2	Aritmética Intervalar Complexa	66
E.3	Representação de Vetores e Matrizes	68

Lista de Figuras

2.1	Representação de $J : X \rightarrow X$ como uma matriz de dimensão infinita.	22
3.1	Resumo dos principais passos da estratégia de solução adotada	33
3.2	Aproximação de $Q(t)$ usando interpolação linear, $\tilde{Q}(t)$	34
3.3	Resultados de convergência e não convergência para $A(t)$ da Equação 3.9	38
3.4	Gráfico de $ \Psi(t) _\infty$ definido pela Equação 3.10	38
4.1	Solução da Equação de Duffing com parâmetros $\xi = \frac{1}{4}$, $\beta = \alpha = \gamma = \omega = 1$ com controle para $(x_1(0), x_2(0)) = (2, 3)$ e sem controle ($u = 0$).	43
4.2	Ação de controle para o Oscilador de Duffing para controle implementado conhecendo-se exatamente os parâmetros	44
4.3	Gráficos de $x(t)$ e $y(t)$ para o Oscilador Duffing com $x(0) = -1,5$ e $x'(0) = 2$	44
4.4	Diagrama esquemático de um pêndulo forçado verticalmente	45
4.5	Órbita desejada e órbita obtida com o controle implementado para pólos em -5 e -6 para parâmetros $\beta = 0, 1$, $p = 2$ e $\omega = 1$	46
4.6	Órbita desejada e órbita obtida com o controle implementado para valores característicos em -10 e -12	46
4.7	Ação de controle para o pêndulo com o controle implementado para pólos em -5 e -6	47
D.1	Solução pelo método de Runge-Kutta. Diferença entre $C(i)$ e C_0	63
E.1	Representação de intervalos complexos na forma retangular (a) e circular (b)	66
E.2	Região correspondente à multiplicação de dois intervalos complexos	67

Notação

1. $x(\bullet; t_0, x_0)$: Aplicação $x : \mathbb{R} \rightarrow \mathbb{R}^n$ tal que $x(t_0) = x_0$.
2. $f^{(k)}(t) = \frac{d^k f}{dt^k}(t)$
3. $A_k = A_{k,1} + iA_{k,2} = \left\langle A(t), e^{ik\frac{2\pi}{T}t} \right\rangle_{L^2}$: Coeficiente de $e^{ik\frac{2\pi}{T}t}$ na série de Fourier Complexa de $A(t)$.
 $A_{k,1} = \text{Re}(A_k)$, $A_{k,2} = \text{Im}(A_k)$
4. $(AQ)_k = (AQ)_{k,1} + i(AQ)_{k,2} = \sum_{\alpha+\beta=k} A_\alpha Q_\beta = \sum_{\alpha+\beta=k} (A_{\alpha,1} + iA_{\alpha,2})(Q_{\beta,1} + iQ_{\beta,2})$ com $\alpha, \beta \in \mathbb{Z}$, $\forall k \in \mathbb{Z}$
5. $M(n, \mathbb{K})$: Matriz quadrada $n \times n$ com entradas no corpo \mathbb{K} . $\mathbb{K} = \mathbb{R}$ ou \mathbb{C}
6. $M(n, m, \mathbb{K})$: Matriz $n \times m$ com entradas no corpo \mathbb{K} . $\mathbb{K} = \mathbb{R}$ ou \mathbb{C}
7. $X = [M(n, \mathbb{K}) \times M(n, \mathbb{K})]^{\mathbb{N}}$: Espaço das seqüências de duplas de matrizes com entradas no corpo \mathbb{K} indexadas por números naturais. $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .
8. $X^m = [M(n, \mathbb{K}) \times M(n, \mathbb{K})]^m$: m-úplas ordenadas de pares de matrizes reais $n \times n$ com entradas em um corpo \mathbb{K} . $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} .
9. $(\bar{x}, 0^\infty) = ((x_{0,1}, x_{0,2}), (x_{1,1}, x_{1,2}), (x_{2,1}, x_{2,2}), \dots, (x_{m-1,1}, x_{m-1,2}), (0, 0), (0, 0), \dots)$, em que $\bar{x} = ((x_{0,1}, x_{0,2}), (x_{1,1}, x_{1,2}), (x_{2,1}, x_{2,2}), \dots, (x_{m-1,1}, x_{m-1,2})) \in [M(n, \mathbb{R}) \times M(n, \mathbb{R})]^m$.
10. $sg(l) = \begin{cases} 1; & l \geq 0 \\ -1; & l < 0 \end{cases}$
11. $C^0(\mathbb{X}, \mathbb{Y}) = \{f : \mathbb{X} \rightarrow \mathbb{Y}; f \text{ é Contínua}\}$. \mathbb{X} e \mathbb{Y} são espaços topológicos.
12. $C([-L, L])$: Espaço das funções reais contínuas definidas no conjunto $[-L, L]$
13. $CP([-L, L])$: Espaço das funções reais contínuas por partes definidas no conjunto $[-L, L]$
14. $|A| = |a_{ij}|$: Matriz de valores absolutos das entradas da matriz A
15. $|R|_\infty = \max(|r_{ij}|)$: Maior módulo das entradas de R .
16. $\|A\|_\infty$: Norma infinito padrão de uma matriz A , definida como $\|A\|_\infty = \max_i \sum_j |a_{ij}|$.
17. $A \leq_{cw} B$: Se $A = (a_{ij})$ e $B = (b_{ij})$, $A \leq_{cw} B \Leftrightarrow a_{ij} \leq b_{ij}, \forall i, j = 1, 2, \dots, n$.
18. $A \leq_{cw} b$: Se $A = (a_{ij})$ e $b \in \mathbb{R}$, $A \leq_{cw} b \Leftrightarrow a_{ij} \leq b \forall i, j = 1, 2, \dots, n$
19. $|y|_2 = \sqrt{\sum_{k=1}^n |y_k|^2}$, em que $y = (y_1, y_2, \dots, y_n) \in \mathbb{C}^n$.
20. $\|A\|_{OP} = \sup_{|x|=1} |Ax|_2$, em que $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$.
21. \mathbb{I}_n : Matriz $n \times n$ que tem todas as entradas iguais a 1.
22. $\omega_k = \begin{cases} |k|; & k \neq 0 \\ 1; & k = 0 \end{cases}; k \in \mathbb{Z}$
23. $\|x\|_s = \sup \{|x_k| (\omega_k)^s\} = \sup \left\{ |R|_\infty, |Q_0|_\infty, \sup_{k \geq 1} \{|Q_{k,1}|_\infty (\omega_k)^s, |Q_{k,2}|_\infty (\omega_k)^s\} \right\}$

24. A^+ : Pseudoinversa de Moore-Penrose da matriz $A \in M(n, m, \mathbb{C})$, definida através da Proposição 4.1.

25. $\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$

26. $o(h^k)$: Utilizado para denotar um termo que decai com velocidade k , ou seja, $o : I \rightarrow \mathbb{R}^n$ tal que $\lim_{h \rightarrow 0} \frac{o(h^k)}{h^k} = 0$.

Capítulo 1

Introdução

Neste capítulo serão enunciados e demonstrados alguns resultados importantes sobre a Teoria de Equações Diferenciais Ordinárias Lineares e, em especial, aquelas cujos coeficientes dependem periodicamente do tempo, Teoria de Floquet. O tratamento desse tipo de equação foi introduzido por Gaston Floquet em 1883, [10]. Os principais resultados deste Capítulo são os Teoremas 1.2 e 1.4. A abordagem apresentada em [6] será utilizada como base para este trabalho.

1.1 Resultados Fundamentais sobre EDO's não-autônomas

Seja $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ uma aplicação contínua, i.e, as aplicações $a_{ij} : \mathbb{R} \rightarrow \mathbb{R}$, $i, j = 1, 2, \dots, n$ correspondentes às entradas da matriz $A(t)$ são contínuas. Consideremos a EDO linear não-autônoma $x' = A(t)x$.

Teorema 1.1 (Existência e Unicidade de Equações Diferenciais Ordinárias Lineares). *Seja $A : I \rightarrow M(n, \mathbb{R})$ uma aplicação contínua em um intervalo $I = [a, b]$. Então, dados $(t_0, x_0) \in I \times \mathbb{R}^n$, o Problema de Valor Inicial Equação (1.1) tem uma única solução $x : I \rightarrow \mathbb{R}^n$. Em particular, tomando $I = \mathbb{R}$, a Equação (1.1) tem solução única definida em todo o \mathbb{R} .*

$$\begin{cases} x' = A(t)x \\ x(t_0) = x_0 \end{cases} \quad (1.1)$$

Proposição 1.1. *O conjunto S de soluções da equação $x' = A(t)x$ é um subespaço vetorial de dimensão n do espaço das funções contínuas $C^0(\mathbb{R}, \mathbb{R}^n)$.*

Demonstração. Segue imediatamente do Teorema 1.1 que toda solução de $x' = A(t)x$ está definida em todo \mathbb{R} . Assim, como todos os elementos de S são funções com mesmo domínio e contradomínio, o fato de que S é um espaço vetorial segue imediatamente da linearidade da equação $x' = A(t)x$. Para provar que $\dim(S) = n$, consideremos $t_0 \in \mathbb{R}$ e $\Psi_{t_0} : \mathbb{R}^n \rightarrow S$ definido por

$$\Psi_{t_0}(v) = x(\bullet; t_0, v)$$

ou seja $\Psi_{t_0}(v)$ é a solução de $x' = A(t)x$ com $x(t_0) = v$. Pelo Teorema de Existência e Unicidade de Soluções, Ψ_{t_0} está bem definido. Dados $v, w \in \mathbb{R}^n$ e $c \in \mathbb{R}$, seja

$$\gamma(\bullet) = x(\bullet; t_0, v) + cx(\bullet; t_0, w)$$

então $\gamma'(t) = A(t)\gamma(t)$ e $\gamma(t_0) = v + cw$. Pelo Teorema de Existência e unicidade $\Psi_{t_0}(v + cw) = \gamma(\bullet)$. Como $x(\bullet; t_0, v) = \Psi_{t_0}(v)$ e $x(\bullet; t_0, w) = \Psi_{t_0}(w)$, temos que $\Psi_{t_0}(v + cw) = \Psi_{t_0}(v) + c\Psi_{t_0}(w)$, provando que Ψ_{t_0} é um homomorfismo.

Dado $\phi \in S$ então $\phi = \Psi_{t_0}(\phi(t_0))$ e logo Ψ_{t_0} é sobrejetivo.

Se $\alpha(t) = \Phi_{t_0}(v) = \Psi_{t_0}(w) = \beta(t)$ então $v = \alpha(t_0) = \beta(t_0) = w$. Assim, Ψ_{t_0} é injetivo e portanto um isomorfismo entre \mathbb{R}^n e S , provando que $\dim(S) = n$. \square

Definição 1.1 (Matriz Fundamental). *Seja $\beta = \{x_1, x_2, \dots, x_n\}$ uma base de do espaço S de soluções da equação $x' = A(t)x$. Uma aplicação $X : \mathbb{R} \rightarrow M(n, \mathbb{R})$ em que $X(t) = \begin{bmatrix} x_1(t) & x_2(t) & \dots & x_n(t) \end{bmatrix}$ é a matriz cujas colunas são os vetores do conjunto β é dita Matriz Fundamental de Soluções da equação $x' = A(t)x$. Se $X(0) = I$ então X é dita Matriz Fundamental Principal.*

Lema 1.1. *Sejam $x_1, x_2, \dots, x_n : \mathbb{R} \rightarrow \mathbb{R}^n$ soluções da equação $x' = A(t)x$ e $X : \mathbb{R} \rightarrow M(n, \mathbb{R})$ tal que $X(t)$ é a matriz cujas colunas são $x_1(t), \dots, x_n(t)$ e $t_0 \in \mathbb{R}$ arbitrário. São equivalentes:*

(i) $\det(X(t_0)) \neq 0$.

(ii) O conjunto $\{x_1(t_0), x_2(t_0), \dots, x_n(t_0)\}$ é base de \mathbb{R}^n .

(iii) X é uma matriz fundamental de $x' = A(t)x$.

(iv) $\det(X(t)) \neq 0$ para todo $t \in \mathbb{R}$.

Além disso, se X é uma matriz fundamental da equação $x' = A(t)x$ então a solução do PVI (1.1) é $x(\bullet; t_0, x_0) = X(\bullet)[X(t_0)]^{-1}x_0$.

Demonstração. A equivalência será demonstrada da forma (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i).

(i) \Rightarrow (ii) : Como existe $t_0 \in \mathbb{R}$ tal que $\det(X(t_0)) \neq 0$, temos que $x_1(t_0), x_2(t_0), \dots, x_n(t_0)$ são linearmente independentes formando uma base de \mathbb{R}^n

(ii) \Rightarrow (iii) : Suponhamos por absurdo que x_1, x_2, \dots, x_n sejam linearmente dependentes, ou seja, existem $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$ tais que $\alpha_1 x_1(t) + \alpha_2 x_2(t) + \dots + \alpha_n x_n(t) = 0 \forall t \in \mathbb{R}$ com algum $\alpha_i \neq 0$, o que implica que $x_1(t_0), x_2(t_0), \dots, x_n(t_0)$ são linearmente dependentes, um absurdo. Como pela hipótese geral do Lema x_1, x_2, \dots, x_n são soluções da equação $x' = A(t)x$, segue que X é uma matriz fundamental da equação $x' = A(t)x$.

(iii) \Rightarrow (iv) : Como X é uma matriz fundamental, suas colunas formam uma base para S . Para provar essa implicação consideremos o inverso do isomorfismo definido na demonstração da Proposição 1.1. Seja $t \in \mathbb{R}$, então $\Psi_t^{-1}(x_i) = x_i(t)$. Como $\{x_1, x_2, \dots, x_n\}$ forma uma base para o espaço de soluções da EDO $x' = A(t)x$, $\{\Psi_t^{-1}(x_1), \Psi_t^{-1}(x_2), \dots, \Psi_t^{-1}(x_n)\}$ forma uma base de \mathbb{R}^n , provando que $\det(X(t)) \neq 0 \forall t \in \mathbb{R}$.

(iv) \Rightarrow (i) : Como $\det(X(t)) \neq 0$ para todo $t \in \mathbb{R}$, $\det(X(t_0)) \neq 0$.

Para provar a segunda parte do Lema observemos inicialmente que $x(t, t_0, x_0)$ está bem definido para todo $t \in \mathbb{R}$. Além disso, $x'(t, t_0, x_0) = X'(t)[X(t_0)]^{-1}x_0 = A(t)X(t)[X(t_0)]^{-1}x_0 = A(t)x(t, t_0, x_0)$ e $x(t_0, t_0, x_0) = x_0$, como queríamos. \square

Lema 1.2. Se $X(t)$ é uma matriz fundamental para a equação $x' = A(t)x$ então:

$$\det(X(t)) = \det(X(t_0)) \exp \left[\int_{t_0}^t \text{Tr}(A(s)) ds \right]$$

Demonstração. Como a aplicação $t \mapsto X(t)$ é diferenciável, temos que

$$X(t+h) = X(t) + X'(t)h + o(h)$$

Como $X'(t) = A(t)X(t)$, pode-se escrever que

$$X(t+h) = X(t) + A(t)X(t)h + o(h)$$

em que $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$. Assim,

$$X(t+h) = (I + hA(t))X(t) + o(h)$$

utilizando a definição de determinante através de permutações $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$,

$$\det B = \sum_{\sigma} \left(\text{sgn}(\sigma) \prod_{i=1}^n b_{i, \sigma(i)} \right)$$

obtemos

$$\det[X(t+h)] = \det[(I + hA(t))X(t)] + o(h)$$

Usando a propriedade de que o determinante de um produto é igual ao produto dos determinantes e a definição de determinante através de permutações, temos que

$$\det[X(t+h)] = (1 + h \cdot \text{Tr}A(t)) \det[X(t)] + o(h)$$

e logo

$$\frac{d}{dt} \det[X(t)] = \lim_{h \rightarrow 0} \frac{\det[X(t+h)] - \det[X(t)]}{h} = \text{Tr}A(t) \det[X(t)]$$

Ou seja $\det[X(t)] = \det[X(t_0)] e^{\int_{t_0}^t \text{Tr}A(s) ds}$. □

Observação 1.1. Se $G, H : \mathbb{R} \rightarrow M(n, \mathbb{R})$ são aplicações diferenciáveis, a aplicação $K : \mathbb{R} \rightarrow M(n, \mathbb{R})$ definida por $K(t) = G(t)H(t)$ também o é.

De fato, pela definição de derivada temos que

$$K'(t) = \lim_{h \rightarrow 0} \frac{1}{h} (K(t+h) - K(t)) = \lim_{h \rightarrow 0} \frac{1}{h} (G(t+h)H(t+h) - G(t)H(t))$$

somando e subtraindo $G(t+h)H(t)$ temos que

$$\begin{aligned} K'(t) &= \lim_{h \rightarrow 0} \frac{1}{h} [G(t+h)H(t+h) - G(t+h)H(t)] + \lim_{h \rightarrow 0} \frac{1}{h} [G(t+h)H(t) - G(t)H(t)] \\ &= G'(t)H(t) + G(t)H'(t) \end{aligned}$$

Lema 1.3. Se $X : \mathbb{R} \rightarrow M(n, \mathbb{R})$ é uma matriz fundamental para a equação $x' = A(t)x$ então $Y : \mathbb{R} \rightarrow M(n, \mathbb{R})$ definida por $Y(t) = X(t)B$ também o é, para toda matriz $B \in M(n, \mathbb{R})$ não-singular.

Demonstração. Como $Y(t) = X(t)B$, tem-se que $Y'(t) = X'(t)B = A(t)X(t)B = A(t)Y(t)$. Assim, Y também é solução da equação matricial $Y' = A(t)Y$. Uma vez que $X(t)$ é matriz fundamental, os vetores coluna que a compõem são linearmente independentes para cada $t \in \mathbb{R}$. Como B é não-singular, $\det(B) \neq 0$ e portanto $\det(Y(t)) = \det(X(t)B) = \det(X(t))\det(B) \neq 0$. Assim, pelo Lema 1.1 conclui-se que $Y(t)$ também é uma matriz fundamental para $x' = A(t)x$. □

Lema 1.4. *Sejam $X, Y : I \rightarrow M(n, \mathbb{R})$ duas matrizes fundamentais da equação $x' = A(t)x$. Então, existe $C \in M(n, \mathbb{R})$ tal que $Y(t) = X(t)C \forall t \in \mathbb{R}$.*

Demonstração. Como X e Y são matrizes fundamentais, $X(t)$ e $Y(t)$ são não-singulares para todo $t \in I$, pelo Lema 1.1. Assim, definamos $C(t) = X^{-1}(t)Y(t)$, ou seja $Y(t) = X(t)C(t)$. Assim,

$$Y'(t) = X(t)C'(t) + X'(t)C(t)$$

ou seja, $A(t)Y(t) = X(t)C'(t) + A(t)X(t)C(t) = X(t)C'(t) + A(t)Y(t)$, de onde vem que $X(t)C'(t) = 0$. Multiplicando ambos os lados da igualdade por $X^{-1}(t)$ segue que $C'(t) = 0$, ou seja C não depende de t , o que completa a demonstração. \square

1.2 Fundamentos da Teoria de Floquet

Uma vez apresentados resultados genéricos sobre EDO's não-autônomas, esta seção aborda o caso em que $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ é uma aplicação contínua e τ -periódica, definindo a EDO (1.2). A partir deste ponto $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ denota uma aplicação contínua e τ -periódica.

$$x' = A(t)x \quad A(t) = A(t + \tau); \quad \forall t \in \mathbb{R} \quad (1.2)$$

Proposição 1.2. *Seja $X : \mathbb{R} \rightarrow M(n, \mathbb{R})$ uma matriz fundamental para a Equação (1.2). Então $Z : \mathbb{R} \rightarrow M(n, \mathbb{R})$ definida por $Z(t) = X(t + \tau) \forall t \in \mathbb{R}$ também é uma matriz fundamental para a Equação (1.2).*

Demonstração. Como $A(t + \tau) = A(t) \forall t \in \mathbb{R}$ temos que $Z'(t) = X'(t + \tau) = A(t + \tau)X(t + \tau) = A(t)Z(t)$. Além disso, o fato de que $\det(Z(t)) \neq 0 \forall t \in \mathbb{R}$ segue imediatamente do Lema 1.1. Assim, $Z(t)$ é matriz fundamental da Equação (1.2). \square

Proposição 1.3. *Se $C \in M(n, \mathbb{C})$ é uma matriz não singular, existe uma matriz $B \in M(n, \mathbb{C})$ tal que $e^B = C$. Além disso se $C \in M(n, \mathbb{R})$ é não-singular então existe uma matriz não singular $B \in M(n, \mathbb{R})$ tal que $e^B = C^2$.*

A demonstração da Proposição anterior está apresentada no Apêndice A.

Teorema 1.2 (Forma Normal de Floquet). *Se $X(t)$ é uma matriz fundamental para a equação $x' = A(t)x$, então existe uma matriz complexa $B \in M(n, \mathbb{C})$ e uma aplicação τ -periódica $P : \mathbb{R} \rightarrow M(n, \mathbb{C})$ tal que*

$$X(t) = P(t)e^{tB} \quad \forall t \in \mathbb{R} \quad (1.3)$$

A escrita de X no formato da equação anterior é chamada Forma Normal de Floquet Complexa de X . Existe ainda uma matriz real $R \in M(n, \mathbb{R})$ e uma aplicação 2τ -periódica $Q : \mathbb{R} \rightarrow M(n, \mathbb{R})$ tal que

$$X(t) = Q(t)e^{tR} \quad \forall t \in \mathbb{R} \quad (1.4)$$

A escrita de X no formato da equação anterior é chamada Forma Normal de Floquet Real de X . Além disso, $P(t)$ na Equação (1.3) e $Q(t)$ na Equação (1.4) são não-singulares para todo $t \in \mathbb{R}$.

Demonstração. Pela Proposição 1.2, se $X(t)$ é uma matriz fundamental de $x' = A(t)x$ então $X(t + \tau)$ também o é. Assim, pelo Lema 1.4, existe uma matriz não-singular $C \in M(n, \mathbb{R})$ tal que $X(t + \tau) = X(t)C$, e portanto $X(t + 2\tau) = X(t)C^2$. Dessa forma pela Proposição 1.3 existem matrizes constantes $B \in M(n, \mathbb{C})$ e $R \in M(n, \mathbb{R})$

tais que $C = e^{\tau B}$ e $C^2 = e^{2\tau R}$. Definindo as aplicações P e Q por $P(t) = X(t)e^{-tB}$ e $Q(t) = X(t)e^{-tR}$ e observando que tB comuta com τB temos que

$$P(t + \tau) = X(t + \tau)e^{-tB - \tau B} = X(t + \tau)C^{-1}e^{-tB} = X(t)e^{-tB} = P(t) \quad \forall t \in \mathbb{R}$$

$$Q(t + 2\tau) = X(t + 2\tau)e^{-tR - 2\tau R} = X(t + 2\tau)[C^2]^{-1}e^{-tR} = X(t)e^{-tR} = Q(t) \quad \forall t \in \mathbb{R}$$

Assim, P é uma aplicação τ -periódica e Q é uma aplicação 2τ -periódica. Além disso, pelas definições de $P(t)$ e $Q(t)$ segue que

$$X(t) = Q(t)e^{tR} = P(t)e^{tB} \quad \forall t \in \mathbb{R}$$

Além disso, como $X(t) = Q(t)e^{tR} = P(t)e^{tB}$ temos que $\det(X(t)) = \det(Q(t))\det(e^{tR}) = \det(P(t))\det(e^{tB})$. Já que $X(t)$ é não singular por ser uma matriz fundamental, $\det(Q(t)) \neq 0$ e $\det(P(t)) \neq 0 \quad \forall t \in \mathbb{R}$, provando que ambas são não-singulares. É importante observar o Exemplo apresentado na Observação A.1, que afirma que o logaritmo de uma matriz não-singular pode não ser único. Mais que isso, o logaritmo de uma matriz que é o quadrado de uma matriz não-singular também pode não ser único. \square

Em termos práticos, o Teorema da Forma Normal de Floquet estabelece que qualquer matriz fundamental da EDO não-autônoma $x' = A(t)x$ pode ser obtida através da matriz fundamental principal de uma EDO autônoma via uma mudança de coordenadas periódica. De fato, definindo $Y(t) = [Q(t)]^{-1}X(t)$ e notando que a matriz fundamental principal de $y' = Ry$ é $Y(t) = e^{tR}$ temos que $X(t) = Q(t)e^{tR}$. O mesmo procedimento pode ser aplicado para a Forma Normal de Floquet Complexa. No restante deste trabalho, será utilizada a denominação *Forma Normal de Floquet* para se referir à Forma Normal de Floquet Real da matriz fundamental principal da equação $x' = A(t)x$. É possível generalizar o Teorema de Floquet para uma classe de sistemas não-lineares através da chamada Teoria de Floquet Generalizada, assunto brevemente abordado no Apêndice B.

O Exemplo a seguir apresenta um caso em que $A(t)$ foi escolhida de forma que fosse conhecida explicitamente a Forma Normal de Floquet (Real e Complexa). Apesar de ser um exemplo simples, ele será utilizado para comparar resultados numéricos e analíticos no Capítulo 3.

Exemplo 1.1. *Seja a equação $x' = A(t)x$ com*

$$A(t) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{2 - \cos(t) + \sin(t)}{2 - \cos(t)} \end{bmatrix} \quad (1.5)$$

Sua Forma Normal de Floquet é dada por

$$X(t) = \begin{bmatrix} 1 & 0 \\ 0 & 2 - \cos(t) \end{bmatrix} e^{tI_2} = e^t \begin{bmatrix} 1 & 0 \\ 0 & 2 - \cos(t) \end{bmatrix} \quad (1.6)$$

É fácil ver que $X'(t) = A(t)X(t)$ e $X(0) = I$. Além disso (1.6) corresponde tanto à Forma Normal de Floquet Real quanto complexa.

Sejam $v \in \mathbb{R}^n$ e $X(t)$ uma matriz fundamental da EDO $x' = A(t)x$. Então $\gamma(t) = X(t)[X(0)]^{-1}v$ é a solução do Problema de Valor Inicial $\begin{cases} x' = A(t)x \\ x(0) = v \end{cases}$. Assim $\gamma(\tau) = X(\tau)[X(0)]^{-1}v$ corresponde ao fluxo no tempo τ do vetor v . É importante notar que $\gamma(\tau)$ não depende da matriz fundamental X escolhida, permitindo definir o Operador de Monodromia Φ .

Definição 1.2 (Operador de Monodromia). *Seja X uma matriz fundamental da Equação (1.2), τ -periódica. O Operador de Monodromia $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ é definido por*

$$\Phi(v) = X(\tau)[X(0)]^{-1}v \quad (1.7)$$

Teorema 1.3 (Teorema de Floquet). *Existe $\mu \in \mathbb{C}$ tal que a EDO periódica $x' = A(t)x$ admite uma solução não-trivial $y(t)$ tal que $y(t + \tau) = \mu y(t)$.*

Demonstração. Seja $X(t)$ uma matriz fundamental para a equação $x' = A(t)x$. Então $X(t + \tau)$ também é matriz fundamental e existe uma matriz C não-singular tal que $X(t + \tau) = X(t)C$. Como C é não-singular 0 não é autovalor de C . Seja μ um autovalor de C e v o autovetor associado a μ . Definamos $y(t) = X(t)v$, solução da Equação (1.2). De fato,

$$y'(t) = X'(t)v = A(t)X(t)v = A(t)y(t)$$

Além disso

$$y(t + \tau) = X(t + \tau)v = X(t)Cv = X(t)(\mu v) = \mu X(t)v = \mu y(t)$$

A solução y não é trivial, já que caso $y(t) = 0 \forall t \in \mathbb{R}$, então existiria $t_0 \in \mathbb{R}$ tal que $0 = y(t_0) = X(t_0)v$ e como X é matriz fundamental segue que $0 = v$, o que é um absurdo pois v é um autovetor (em \mathbb{C}^n), logo $v \neq 0$. \square

Proposição 1.4 (Multiplicadores de Floquet). *Os valores de $\mu \in \mathbb{C}$ do Teorema 1.3 são denominados Multiplicadores de Floquet e independem da matriz fundamental escolhida. Além disso, se μ é um Multiplicador de Floquet então μ é um autovalor do Operador de Monodromia.*

Demonstração. Seja $Y(t)$ outra matriz fundamental para a Equação (1.2). Dessa forma pelo Lema 1.4 existe uma matriz $P \in M(n)$ tal que $Y(t) = X(t)P$. Assim, $Y(t + \tau) = X(t + \tau)P = X(t)CP$, o que implica $Y(t + \tau) = X(t + \tau)P = X(t)CP = Y(t)P^{-1}CP$. Como as matrizes C e $P^{-1}CP$ são semelhantes, seus autovalores são iguais e portanto os valores de μ no Teorema 1.3 independem da escolha de uma matriz fundamental. Para provar que μ é autovalor do operador de monodromia basta observar que por construção μ é autovalor de C tal que $X(t + \tau) = X(t)C$ para uma matriz fundamental $X(t)$ escolhida arbitrariamente. Dessa forma, fazendo $t = 0$ temos $C = X(\tau)[X(0)]^{-1}$ e portanto μ é autovalor de Φ . \square

As noções de Estabilidade Assintótica e Estabilidade no sentido de Lyapunov utilizadas neste trabalho são aquelas apresentadas em [28].

Definição 1.3 (Estabilidade Assintótica e Estabilidade de Lyapunov). *Uma solução $y(t)$ de uma equação diferencial $x' = f(t, x)$ é dita Estável no sentido de Lyapunov se para todo $\epsilon > 0$ existe $\delta > 0$ tal que para qualquer solução $z(t)$ tal que $\|y(t_0) - z(t_0)\| < \delta$, $\|y(t) - z(t)\| < \epsilon$ para $t > t_0$, $t_0 \in \mathbb{R}$. A solução $y(t)$ é dita assintoticamente estável se $\lim_{t \rightarrow \infty} \|y(t) - z(t)\| = 0$*

Teorema 1.4 (Estabilidade da Origem). *Seja $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ uma aplicação contínua e τ -periódica e $X(t) = Q(t)e^{tR}$ Forma Normal de Floquet Real para essa EDO. Então, a mudança de coordenadas $x = Q(t)y$ transforma as soluções da EDO não-autônoma $x' = A(t)x$ nas soluções da EDO autônoma $y' = Ry$. Além disso:*

- (i) *Se todos os autovalores de R têm parte real negativa, então a origem é uma solução assintoticamente estável da equação $x' = A(t)x$.*
- (ii) *Se todos os autovalores de R têm parte real não positiva, e se as multiplicidades geométrica e algébrica de cada autovalor imaginário puro são iguais então a origem é estável no sentido de Lyapunov da equação $x' = A(t)x$.*
- (iii) *Se existe autovalor com parte real positiva, então a origem é uma solução instável da equação $x' = A(t)x$.*

Demonstração. Seja $x(t)$ uma solução de $x' = A(t)x$. Então $x(t) = Q(t)e^{tR}Q^{-1}(0)x(0)$. Usando a transformação $y = Q^{-1}(t)x$ temos que $y(t) = e^{tR}Q^{-1}(0)x(0)$ que é uma solução da EDO $y' = Ry$. Reciprocamente, se $y(t)$ é solução da EDO $y' = Ry$ então $y(t) = e^{tR}y(0)$. Usando a transformação $x = Q(t)y$, temos que $x(t) = Q(t)e^{tR}y(0)$, que é uma solução da EDO $x' = A(t)x$.

Para provar o ítem (i) seja $x_0 \in \mathbb{R}^m$. A solução de $x' = A(t)x$ com condição inicial $x(0) = x_0$ é $x(t) = Q(t)e^{tR}Q^{-1}(0)x_0$, e dessa forma

$$\|x(t)\| = \left\| Q(t) e^{tR} [Q(0)]^{-1} x_0 \right\| \leq \max_{v \neq 0} \left\| Q(t) \frac{v}{\|v\|} \right\| \left\| e^{tR} [Q(0)]^{-1} x_0 \right\|$$

Como todos os autovalores de R têm parte real negativa então $\lim_{t \rightarrow \infty} \|e^{tR}x_0\| = 0$. Além disso, como $Q(t)$ é contínua e periódica, logo limitada, temos que $\lim_{t \rightarrow \infty} \|x(t)\| = 0$, provando o ítem (i).

Para provar o ítem (ii) é necessário observar a Forma Canônica de Jordan de R , uma vez que se $R = KJK^{-1}$ então $e^{tR} = Ke^{tJ}K^{-1}$. A matriz J pode ser escrita como $J = \begin{bmatrix} B_n & 0 \\ 0 & B_0 \end{bmatrix}$, em que B_n é uma matriz diagonal por blocos formada pelos blocos de Jordan de R associados aos autovalores com parte real negativa e B_0 a matriz diagonal por blocos formada pelos blocos de Jordan associados aos autovalores com parte real nula. Como as multiplicidades algébrica e geométrica dos autovalores de R com parte real nula são iguais, os blocos de Jordan referentes a esses autovalores são diagonais. Assim, B_0 é uma matriz diagonal formada cujas entradas tem parte real nula. Dessa forma, $e^{tJ} = \begin{bmatrix} e^{tB_n} & 0 \\ 0 & e^{tB_0} \end{bmatrix}$. Como $\lim_{t \rightarrow \infty} \|e^{tB_n}\| = 0$ e $\|e^{tB_0}\|$ é limitado. Desse modo existe $M > 0$ tal que $M = \sup_{v \neq 0} \left\| Q(t) Ke^{tJ}K^{-1} [Q(0)]^{-1} \frac{v}{\|v\|} \right\|$. Sejam $x(t) = Q(t)Ke^{tJ}K^{-1}[Q(0)]^{-1}x_0$ e $y(t) = Q(t)Ke^{tJ}K^{-1}[Q(0)]^{-1}y_0$ as soluções da EDO $x' = A(t)x$ com condições iniciais $x(0) = x_0$ e $y(0) = y_0$. Se $\|x_0 - y_0\| < \delta$ então, definindo $\epsilon = M\delta$,

$$\|x(t) - y(t)\| = \left\| Q(t) Ke^{tJ}K^{-1} [Q(0)]^{-1} (x_0 - y_0) \right\| < M\delta = \epsilon$$

de onde segue o ítem (ii).

O ítem (iii) segue imediatamente das argumentações já utilizadas, pois considerando a matriz diagonal $\lambda_j I$ em que λ_j é um valor característico de R com parte real positiva. Assim, $e^{t\lambda_j I} = e^{\lambda_j t} I$, uma matriz diagonal cujos termos não-nulos crescem arbitrariamente. Fixada a matriz que faz a transformação para a Forma Canônica de Jordan, K , e argumentando que $Q(t)$ é não-singular $\forall t \in \mathbb{R}$, temos que $\|Q(t)\| > \epsilon > 0$, temos que $X(t) = \|Q(t)e^{tR}\| \rightarrow \infty$, o que completa a demonstração. \square

Para aplicar o Teorema 1.4 é necessário calcular a matriz R o que é, em geral, uma tarefa difícil. Uma pergunta natural é: é possível associar os autovalores de $A(t)$ para cada $t \in \mathbb{R}$ com a estabilidade da origem como solução de $x' = A(t)x$?

Exemplo 1.2. Seja a EDO π -periódica $x' = A(t)x$ em que

$$A(t) = \begin{bmatrix} -1 + \frac{3}{2} \cos^2(t) & 1 - \frac{3}{2} \sin(t) \cos(t) \\ -1 - \frac{3}{2} \sin(t) \cos(t) & -1 + \frac{3}{2} \sin^2(t) \end{bmatrix}$$

Os autovalores de $A(t)$ são $\frac{1}{4}(-1 \pm \sqrt{7}i)$, ambos com parte real negativa. No entanto,

$$x(t) = e^{\frac{t}{2}} \begin{bmatrix} -\cos(t) \\ \sin(t) \end{bmatrix}$$

é solução de $x' = A(t)x$ e $\|x(t)\|$ é ilimitado quando t cresce. Este exemplo mostra, portanto, que a relação entre os autovalores de $A(t)$ e a estabilidade de origem não é imediata como no caso de equações lineares autônomas. No entanto, a Proposição 1.5 fornece uma propriedade dos multiplicadores de Floquet útil para analisar a estabilidade da origem como solução de $x' = A(t)x$.

Proposição 1.5. *Se uma EDO τ -periódica $x' = A(t)x$ tem os multiplicadores de Floquet $\lambda_1, \lambda_2, \dots, \lambda_n$ então:*

$$\lambda_1 \lambda_2 \dots \lambda_n = \int_0^\tau \text{Tr}[A(s)] ds \quad (1.8)$$

Em particular, se $\int_0^\tau \text{Tr}(A(s)) ds > 1$ então a origem é uma solução instável da equação $x' = A(t)x$.

Demonstração. Pela forma normal de Floquet dada pelo Teorema 1.2, $X(t) = P(t)e^{tB}$. Assim, $X(0) = P(0)$ e $X(\tau) = P(\tau)e^{\tau B}$. Mas como $P(0) = P(\tau)$ temos que $X(\tau) = P(0)e^{\tau B}$. Aplicando o Lema 1.2

$$\det(P(0)e^{\tau B}) = \det(P(0)) \exp\left(\int_0^\tau \text{Tr}(A(s)) ds\right)$$

$$\lambda_1 \lambda_2 \dots \lambda_n = \det(e^{\tau B}) = \exp\left(\int_0^\tau \text{Tr}(A(s)) ds\right)$$

Se $\int_0^\tau \text{Tr}(A(s)) ds > 1$ temos $\lambda_1 \lambda_2 \dots \lambda_n > 1$. Logo, existe λ_i multiplicador de Floquet com $|\lambda_i| > 1$. Assim, pelo Teorema 1.4, a origem é uma solução instável de $x' = A(t)x$. \square

Lema 1.5 (Lema de Gronwall). *Sejam $I = [t_0, t_1]$, $\alpha > 0$ e $u, v : I \rightarrow \mathbb{R}$ duas funções positivas contínuas tais que*

$$u(t) \leq \alpha + \int_{t_0}^t u(s)v(s) ds$$

então, para cada $t \in [t_0, t_1]$ vale que

$$u(t) \leq \alpha e^{\int_{t_0}^t v(s) ds}$$

Em particular, se $\alpha = 0$ então $u(t) = 0 \forall t \in [t_0, t_1]$.

Demonstração. Primeiramente vamos considerar $\alpha > 0$. Definindo $\theta(t) = \alpha + \int_{t_0}^t u(s)v(s) ds$ temos que pelo Teorema Fundamental do Cálculo $\frac{d\theta}{dt} = u(t)v(t)$. Como por hipótese $u(t) \leq \theta(t)$, $\frac{d\theta}{dt} \leq \theta(t)v(t)$. Já que estamos analisando o caso em que $\alpha > 0$ e $u(t), v(t) \geq 0$ então $\theta(t) > 0 \forall t \in [t_0, t_1]$. Assim,

$$\frac{1}{\theta(t)} \frac{d\theta}{dt} \leq v(t) \Rightarrow \frac{d}{dt} (\log(\theta(t))) \leq v(t) \Rightarrow \int_{t_0}^t \frac{d}{ds} (\log(\theta(s))) ds \leq \int_{t_0}^t v(s) ds$$

e aplicando o Teorema Fundamental do Cálculo,

$$\log(\theta(t)) - \log(\theta(t_0)) \leq \int_{t_0}^t v(s) ds \Rightarrow \log\left(\frac{\theta(t)}{\theta(t_0)}\right) \leq \int_{t_0}^t v(s) ds$$

como a função exponencial é crescente,

$$\frac{\theta(t)}{\theta(t_0)} \leq e^{\int_{t_0}^t v(s) ds} \Rightarrow \theta(t) \leq \theta(t_0) e^{\int_{t_0}^t v(s) ds}$$

Mas $\alpha = \theta(t_0)$ e $u(t) \leq \theta(t)$. Dessa forma, $u(t) \leq \alpha e^{\int_{t_0}^t v(s) ds}$, provando o resultado para $\alpha > 0$. Para provar o resultado para $\alpha = 0$ basta utilizar o resultado para $\alpha_n = \frac{1}{n}$, já que

$$u(t) \leq \alpha_n e^{\int_{t_0}^t v(s) ds} = \frac{1}{n} e^{\int_{t_0}^t v(s) ds} \Rightarrow \lim_{n \rightarrow \infty} u(t) \leq \lim_{n \rightarrow \infty} \frac{1}{n} e^{\int_{t_0}^t v(s) ds}$$

como $e^{\int_{t_0}^t v(s) ds}$ é finito, segue que $u(t) = 0 \forall t \in [t_0, t_1]$. \square

Proposição 1.6. *O fluxo da equação $x' = A(t)x$ depende continuamente de A em que $\|A\|_{\text{sup}} = \sup_{0 \leq t \leq \tau} \|A(t)\|_{OP}$ e $\|A(t)\|_{OP} = \sup_{\|x\|=1} |A(t)x|$, em que $|\bullet|$ é a norma usual de \mathbb{R}^n*

Demonstração. Seja $\epsilon > 0$, $\phi(\bullet, 0, x_0)$ o fluxo da equação $x' = A(t)x$ e $\bar{\phi}(\bullet, 0, x_0)$ o fluxo da equação $x' = \bar{A}(t)x$. Então, temos que

$$\phi(t, 0, x_0) = x_0 + \int_0^t A(s)\phi(s, 0, x_0) ds \quad \bar{\phi}(t, 0, x_0) = x_0 + \int_0^t \bar{A}(s)\bar{\phi}(s, 0, x_0) ds$$

dessa forma

$$\phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) = \int_0^t [A(s)\phi(s, 0, x_0) - \bar{A}(s)\bar{\phi}(s, 0, x_0)] ds$$

somando e subtraindo $A(s)\bar{\phi}(s, 0, x_0)$, temos que

$$\phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) = \int_0^t [A(s)\phi(s, 0, x_0) - A(s)\bar{\phi}(s, 0, x_0) + A(s)\bar{\phi}(s, 0, x_0) - \bar{A}(s)\bar{\phi}(s, 0, x_0)] ds$$

separando a integral

$$\phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) = \int_0^t [A(s)\phi(s, 0, x_0) - A(s)\bar{\phi}(s, 0, x_0)] ds + \int_0^t [A(s)\bar{\phi}(s, 0, x_0) - \bar{A}(s)\bar{\phi}(s, 0, x_0)] ds$$

Como $\left\| \int_0^t [A(s)\bar{\phi}(s, 0, x_0) - \bar{A}(s)\bar{\phi}(s, 0, x_0)] ds \right\| \leq t \max_{0 \leq s \leq t} \{ \| (A(s) - \bar{A}(s)) \|_{OP} \| \bar{\phi}(s, 0, x_0) \| \}$. Definindo o intervalo $I = [0, 2\tau]$ temos que $\forall t \in I$ vale

$$\| \phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) \| \leq \int_0^t \| A(s) \|_{OP} \| \phi(s, 0, x_0) - \bar{\phi}(s, 0, x_0) \| ds + 2\tau \max_{0 \leq s \leq 2\tau} \{ \| (A(s) - \bar{A}(s)) \|_{OP} \| \bar{\phi}(s, 0, x_0) \| \}$$

Aplicando o Lema de Gronwall com $\alpha = 2\tau \max_{0 \leq s \leq 2\tau} \{ \| (A(s) - \bar{A}(s)) \|_{OP} \| \bar{\phi}(s, 0, x_0) \| \}$, $u(t) = \| \phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) \|$ e $v(t) = \| A(t) \|_{OP}$.

$$\| \phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) \| \leq 2\tau \max_{0 \leq s \leq 2\tau} \{ \| (A(s) - \bar{A}(s)) \|_{OP} \| \bar{\phi}(s, 0, x_0) \| \} e^{\int_0^t \| A(s) \|_{OP} ds}$$

Definindo $K = 2\tau \max_{0 \leq s \leq 2\tau} \{ \| \bar{\phi}(s, 0, x_0) \| \} e^{\int_0^{2\tau} \| A(s) \|_{OP} ds}$ temos que

$$\| \phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) \| \leq K \max_{0 \leq s \leq 2\tau} \{ \| (A(s) - \bar{A}(s)) \|_{OP} \} = K \| A - \bar{A} \|_{\text{sup}}$$

Definindo $\delta = \frac{\epsilon}{K}$ temos que $\| A - \bar{A} \|_{\text{sup}} < \delta \Rightarrow \| \phi(t, 0, x_0) - \bar{\phi}(t, 0, x_0) \| < \epsilon$, provando que o fluxo da equação $x' = A(t)x$ é contínuo com relação a A .

□

Teorema 1.5 (Dependência contínua dos multiplicadores de Floquet). *Os multiplicadores de Floquet da equação $x' = A(t)x$ dependem continuamente da aplicação A .*

Demonstração. Seja $X(t)$ a matriz fundamental principal de $x' = A(t)x$. Pelo enunciado da Proposição 1.4, os Multiplicadores de Floquet são os autovalores da matriz C tal que $X(t + \tau) = X(t)C$. Assim, fazendo $t = 0$ temos que $X(\tau) = X(0)C = C$. Se $\{e_1, e_2, \dots, e_n\}$ é a base canônica de \mathbb{R}^n então

$$X(t) = \begin{bmatrix} \phi(t, 0, e_1) & \phi(t, 0, e_2) & \dots & \phi(t, 0, e_n) \end{bmatrix}$$

Dessa forma, pela continuidade do fluxo demonstrada na Proposição anterior e pelo fato de que os autovalores dependem continuamente da matriz, segue o resultado. □

Capítulo 2

Cálculo Numérico da Forma Normal de Floquet

Seja $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ uma aplicação contínua e τ -periódica. O cálculo da Forma Normal de Floquet de uma solução fundamental $X(t) = Q(t)e^{tR}$ não é, em geral, uma tarefa fácil. Este Capítulo apresenta um método numérico rigoroso para calcular $Q(t)$ e R da Forma Normal de Floquet real da equação $x' = A(t)x$, ou seja, a Forma Normal de Floquet da matriz fundamental principal de $x' = A(t)x$. Os conceitos de Análise Funcional utilizados correspondem à abordagem de [4].

Proposição 2.1. *Seja $X(t)$ a Matriz Fundamental Principal da EDO τ -periódica $x' = A(t)x$. Então, existem $Q : \mathbb{R} \rightarrow M(n, \mathbb{R})$ 2τ -periódica e $R \in M(n, \mathbb{R})$ tais que.*

$$\begin{cases} Q'(t) = A(t)Q(t) - Q(t)R \\ Q(0) = I \end{cases} \quad (2.1)$$

Além disso, $X(t) = Q(t)e^{tR}$ é uma Forma Normal de Floquet para a EDO $x' = A(t)x$

Demonstração. O Teorema de Floquet garante a existência de $Q : \mathbb{R} \rightarrow M(n, \mathbb{R})$ 2τ -periódica e $R \in M(n, \mathbb{R})$ tais que $X(t) = Q(t)e^{tR}$. Como $X(t)$ é Matriz Fundamental Principal temos que $X(0) = I$ logo temos $Q(0)e^{0R} = Q(0) = I$. Além disso, como $X' = A(t)X$, e R comuta com $e^{tR} \forall t \in \mathbb{R}$, $X'(t) = Q'(t)e^{tR} + Q(t)Re^{tR} = A(t)Q(t)e^{tR}$. Desse modo, como e^{tR} é sempre invertível pois $\det(e^{tR}) = e^{\det(tR)} \neq 0$, temos que $Q'(t) = A(t)Q(t) - Q(t)R$. Dessa forma, Q é 2τ -periódica e a Equação 2.1 é satisfeita, completando a demonstração da proposição. \square

2.1 Estratégia de Solução do Problema

Como o objetivo é obter R e $Q(t)$ para todo $t \in \mathbb{R}$ que satisfazem a Equação 2.1, será utilizada a decomposição de $Q(t)$ em Séries de Fourier. Não é a primeira vez que uma metodologia semelhante é aplicada para o cálculo da Forma Normal de Floquet. Em [22], é apresentada uma técnica que aplica Polinômios de Chebychev ao invés de Séries de Fourier para a representação de $Q(t)$. No presente trabalho será seguida a abordagem de [27] e [5] cuja metodologia consiste em transformar o problema de encontrar $Q(t)$ e R em um problema de solução de

uma equação do tipo $f(x) = 0$, que será resolvida utilizando um argumento de contração via Teorema do Ponto Fixo de Banach.

Como $A(t)$ é uma aplicação contínua τ -periódica e $Q(t)$ é uma aplicação contínua 2τ -periódica, é conveniente tomar as séries de Fourier de período 2τ de ambas, de forma que $A(t)$ e $Q(t)$ são completamente definidas pelos coeficientes de sua série de Fourier, dadas por

$$A(t) = \sum_{k=-\infty}^{\infty} A_k e^{ik\frac{2\pi}{2\tau}t} \quad Q(t) = \sum_{k=-\infty}^{\infty} Q_k e^{ik\frac{2\pi}{2\tau}t}$$

em que $Q_k = Q_{k,1} + iQ_{k,2}$ e $A_k = A_{k,1} + iA_{k,2}$ com $Q_{k,1}, Q_{k,2}, A_{k,1}, A_{k,2} \in M(n, \mathbb{R})$ para $k \neq 0$ e $A_0, Q_0 \in M(n, \mathbb{R})$. Além disso, a aplicação $t \mapsto A(t)Q(t)$ também é 2τ -periódica, sendo sua série de Fourier dada pela Equação a seguir, em que $(AQ)_k = (AQ)_{k,1} + i(AQ)_{k,2}$ com $(AQ)_{k,1}, (AQ)_{k,2} \in M(n, \mathbb{R})$ são os coeficientes da série de Fourier dessa aplicação.

$$A(t)Q(t) = \sum_{k=-\infty}^{\infty} (AQ)_k e^{ik\frac{2\pi}{2\tau}t}$$

Embora seja possível garantir a convergência pontual da Série de Fourier com a hipótese de continuidade, vamos supor que $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ é de classe C^1 . Isso não traz problemas para os casos que se deseja tratar neste trabalho, conforme será verificado mais a frente. O Apêndice C apresenta alguns resultados sobre séries de Fourier, entre os quais a obtenção de $(AQ)_k$ através dos termos A_l e Q_r . Derivando termo a termo a série de Fourier de $Q(t)$, $Q'(t) = \sum_{k=-\infty}^{\infty} ik\frac{2\pi}{2\tau}Q_k e^{ik\frac{2\pi}{2\tau}t}$. Substituindo na Equação 2.1 temos que

$$\sum_{k=-\infty}^{\infty} ik\frac{2\pi}{2\tau}Q_k e^{ik\frac{2\pi}{2\tau}t} = Q'(t) = A(t)Q(t) - Q(t)R = \sum_{k=-\infty}^{\infty} (AQ)_k e^{ik\frac{2\pi}{2\tau}t} - \sum_{k=-\infty}^{\infty} Q_k e^{ik\frac{2\pi}{2\tau}t} R$$

Como o conjunto $\left\{ e^{ik\frac{2\pi}{2\tau}t} \right\}_{k \in \mathbb{Z}}$ é ortonormal, para cada $k \in \mathbb{Z}$ vale que

$$ik\frac{2\pi}{2\tau}Q_k = (AQ)_k - Q_k R \Rightarrow ik\frac{2\pi}{2\tau}Q_k + Q_k R - (AQ)_k = 0$$

Portanto, para $k = 0$ a equação acima fica

$$Q_0 R - (AQ)_0 = 0$$

e para $k \neq 0$ temos uma equação para a parte real e outra para a parte imaginária,

$$\begin{cases} -k\frac{\pi}{\tau}Q_{k,2} + Q_{k,1}R - (AQ)_{k,1} = 0 \\ k\frac{\pi}{\tau}Q_{k,1} + Q_{k,2}R - (AQ)_{k,2} = 0 \end{cases} \quad k \neq 0$$

Além disso, para que seja satisfeita a Equação 2.1 devemos ter $Q(0) = I$, ou seja,

$$Q(0) = \sum_{k=-\infty}^{\infty} Q_k = I \Rightarrow Q_0 + 2 \sum_{k=1}^{\infty} Q_{k,1} - I = 0$$

Embora as equações tenham sido apresentadas para $k \in \mathbb{Z}$, pelas propriedades da Série de Fourier no Apêndice C temos $Q_{k,1} = Q_{-k,1}$ e $Q_{k,2} = -Q_{-k,2} \forall k \neq 0$. Dessa forma, o problema de obter a Forma Normal de Floquet consiste em calcular R , Q_0 , $Q_{k,1}$ e $Q_{k,2}$ para $k \geq 1$.

A partir deste ponto vamos utilizar a seguinte notação para facilitar a apresentação dos resultados

$$X = [M(n, \mathbb{R}) \times M(n, \mathbb{R})]^{\mathbb{N}}$$

Proposição 2.2. *Sejam $x, y \in X$ dados por $x = ((x_{0,1}, x_{0,2}), (x_{1,1}, x_{1,2}), \dots, (x_{k,1}, x_{k,2}), \dots)$, $y = ((y_{0,1}, y_{0,2}), (y_{1,1}, y_{1,2}), \dots, (y_{k,1}, y_{k,2}), \dots)$ e $\lambda \in \mathbb{R}$. As operações de soma e multiplicação por escalares Reais definidas por*

$$\begin{aligned} x + y &= ((x_{0,1} + y_{0,1}, x_{0,2} + y_{0,2}), (x_{1,1} + y_{1,1}, x_{1,2} + y_{1,2}), \dots, (x_{k,1} + y_{k,1}, x_{k,2} + y_{k,2}), \dots) \\ \lambda x &= ((\lambda x_{0,1}, \lambda x_{0,2}), (\lambda x_{1,1}, \lambda x_{1,2}), \dots, (\lambda x_{k,1}, \lambda x_{k,2}), \dots) \end{aligned}$$

tornam X um espaço vetorial de dimensão infinita.

Demonstração. Inicialmente, cabe destacar que as operações estão bem definidas. Além disso, o fato de que as operações consistem apenas em soma e multiplicação por escalar termo a termo, as propriedades necessárias a um espaço vetorial estão mantidas. O fato de que a dimensão é infinita pode ser demonstrado através de um isomorfismo natural entre X e \mathbb{R}^∞ definido enumerando as entradas de cada componente de um vetor $x \in X$ como no exemplo abaixo

$$x = \left(\left(\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}, \begin{bmatrix} a_5 & a_6 \\ a_7 & a_8 \end{bmatrix} \right), \left(\begin{bmatrix} a_9 & a_{10} \\ a_{11} & a_{12} \end{bmatrix}, \begin{bmatrix} a_{13} & a_{14} \\ a_{15} & a_{16} \end{bmatrix} \right), \dots \right) \leftrightarrow (a_1, a_2, a_3, a_4, \dots)$$

□

Proposição 2.3. *Seja $f : X \rightarrow X$ definida abaixo. Existe $x \in X$ tal que $f(x) = 0$. Além disso, se $x = ((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots, (Q_{k,1}, Q_{k,2})) \in X$ com $f(x) = 0$ temos que $Q(t)e^{tR}$ é Forma Normal de Floquet da Equação $x' = A(t)x$, em que $Q(t) = Q_0 + \sum_{k \neq 0} (Q_{k,1} + iQ_{k,2}) e^{ik\frac{2\pi}{\tau}t}$.*

$$f = ((f_*, f_0), (f_{1,1}, f_{1,2}), (f_{2,1}, f_{2,2}), \dots) \quad (2.2)$$

$$\begin{cases} f_*(x) = Q_0 + 2 \sum_{k=1}^{\infty} Q_{k,1} - I & \begin{cases} f_{k,1}(x) = -k\frac{\pi}{\tau}Q_{k,2} + Q_{k,1}R - (AQ)_{k,1} \\ f_{k,2}(x) = k\frac{\pi}{\tau}Q_{k,1} + Q_{k,2}R - (AQ)_{k,2} \end{cases} \\ f_0(x) = Q_0R - (AQ)_0 & k \geq 1 \end{cases}$$

Demonstração. Seja $Q(t)e^{tR}$ Forma Normal de Floquet de $x' = A(t)x$ e $Q(t) = \sum_{k=-\infty}^{\infty} (Q_{k,1} + iQ_{k,2}) e^{ik\frac{2\pi}{\tau}t}$ a série de Fourier de Q . Então, definindo $x = ((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots)$ temos que $f(x) = 0$. Por construção, a definição de Q faz com que $Q(t)e^{tR}$ seja Forma Normal de Floquet da equação $x' = A(t)x$. □

A Proposição 2.3 garante que a obtenção de uma Forma Normal de Floquet da equação $x' = A(t)x$ pode ser feita mediante a solução de uma equação do tipo $f(x) = 0$ com $f : X \rightarrow X$, X um espaço vetorial de dimensão infinita. A Proposição 2.3 garante ainda a existência de solução dessa equação em X . A metodologia que será empregada para obtê-la consiste em definir um operador linear $T : X \rightarrow X$ cujos pontos fixos são as soluções de $f(x) = 0$. Um método bastante eficaz de obtenção de pontos fixos de operadores lineares é o Teorema do Ponto Fixo de Banach. No entanto, para sua aplicação, é necessário que o operador seja definido em um espaço de Banach. Dessa forma, será definido um subespaço vetorial $\Omega^s \subset X$ que, munido de uma norma $\|\bullet\|_s$, torna-se um espaço de Banach. Assim, utilizando o Teorema do Ponto Fixo de Banach será construído um método numérico rigoroso para o Cálculo das soluções de $f(x) = 0$ e, conseqüentemente, uma Forma Normal de Floquet para $x' = A(t)x$.

2.2 Métodos *Quasi-Newton* para Solução de Equações em Espaços de Banach

Lema 2.1. *Sejam E um espaço de Banach, $J : E \rightarrow E$ uma aplicação linear invertível e $f : E \rightarrow E$. Defina o operador $T : E \rightarrow E$ por $T(x) = x - Jf(x)$. Então $x \in E$ é tal que $f(x) = 0$ se e somente se $T(x) = x$.*

Demonstração. (\Rightarrow): Suponha que $f(x) = 0$. Então $T(x) = x - J0 = x$.

(\Leftarrow): Se $T(x) = x - Jf(x) = x$ então $Jf(x) = 0$. Como J é invertível é injetiva logo $f(x) = 0$. \square

Teorema 2.1 (Teorema do Ponto Fixo de Banach). *Seja (\mathbb{X}, d) um espaço métrico completo e $F \subset \mathbb{X}$ um conjunto fechado não vazio. Se $T : F \rightarrow F$ é uma contração, i.e. existe $\lambda \in \mathbb{R}$, $0 \leq \lambda < 1$ tal que $d(T(x), T(y)) \leq \lambda d(x, y) \forall x, y \in \mathbb{X}$. Então, existe um único $x^* \in F$ tal que $T(x^*) = x^*$. Além disso, o ponto fixo x^* pode ser obtido como limite da sequência definida iterativamente $x_{k+1} = T(x_k)$, $k = 0, 1, 2, \dots$ em que $x_0 \in \mathbb{X}$ é arbitrário.*

Demonstração. Unicidade: Sejam $x, y \in F$ pontos fixos de T com $x \neq y$. Assim, $d(x, y) = d(T(x), T(y)) \leq \lambda d(x, y) < d(x, y)$, o que é um absurdo.

Existência: Seja $x_0 \in F$ e considere $x_{k+1} = T(x_k)$, $\forall k \geq 1$. Assim, usando que T é uma contração tem-se que

$$d(x_{k+m}, x_k) \leq \sum_{i=0}^{m-1} d(x_{m+k-i}, x_{m+k-i-1}) \leq \sum_{i=0}^{m-1} \lambda^{m+k-i-1} d(x_1, x_0)$$

Assim, temos que

$$d(x_{k+m}, x_k) \leq d(x_1, x_0) \left[\lambda^{m+k-1} \left(\frac{\lambda^{-m} - 1}{\lambda^{-1} - 1} \right) \right] = d(x_1, x_0) \lambda^k \left(\frac{1 - \lambda^m}{1 - \lambda} \right)$$

Mas como $0 < 1 - \lambda^m < 1$,

$$d(x_{k+m}, x_k) < d(x_1, x_0) \left(\frac{\lambda^k}{1 - \lambda} \right) \xrightarrow{k \rightarrow \infty} 0$$

Dessa forma (x_k) é uma sequência de Cauchy. Como \mathbb{X} é um espaço completo, $x_k \rightarrow x^*$ para algum $x^* \in \mathbb{X}$. Mas como $(x_k)_{k \geq 0}$ é uma sequência no conjunto fechado F , $x^* \in F$. Como $x_{k+1} = T(x_k)$, passando o limite, uma vez que T é contínua em virtude de ser uma contração, temos que

$$x^* = \lim_{k \rightarrow \infty} (x_{k+1}) = T \left(\lim_{k \rightarrow \infty} (x_k) \right) = T(x^*)$$

provando a existência de ponto fixo para T em F . \square

Teorema 2.2. *Sejam E um espaço de Banach, $f : E \rightarrow E$ e $x_0 \in E$. Se existem $J : E \rightarrow E$ aplicação linear invertível e $F \subset E$ um conjunto fechado com $x_0 \in F$ tal que o operador $T(x) = x - Jf(x)$ é tal que $T(F) \subset F$ e $T|_F : F \rightarrow F$ é uma contração então a sequência $(x_k)_{k \geq 0}$ definida recursivamente*

$$x_{k+1} = T(x_k) \quad k \geq 0$$

converge para algum $\tilde{x} \in E$ tal que $f(\tilde{x}) = 0$, que independe de x_0 .

Demonstração. Pelo Lema 2.1, as soluções da equação $f(x) = 0$ em E são os pontos fixos de T . Se for possível definir o operador J e o conjunto F conforme o enunciado, fica claro, pelo Teorema do Ponto Fixo de Banach que a sequência $(x_k)_k$ converge para algum $\tilde{x} \in E$ que é ponto fixo de T , logo $f(\tilde{x}) = 0$, completando o argumento. \square

O Lema 2.1 garante que os pontos fixos do operador T definido por $T(x) = x - Jf(x)$ com $J : E \rightarrow E$ são as soluções de $f(x) = 0$ em E . Já o Teorema 2.2 fornece um esquema iterativo para a obtenção de um ponto fixo para T , desde que $T|_F$ seja uma contração para algum $F \subset E$ fechado. Como f do problema é diferenciável, T também o é, por sua definição. Dessa forma, temos que para $x, y \in E$

$$T(x) - T(y) = DT(y)(x - y) + \rho(x, y) \|x - y\|$$

com $\lim_{\|x-y\| \rightarrow 0} \rho(x, y) = 0$. Se $\|\bullet\|$ é uma norma em E e $\|\bullet\|_{OP}$ a norma operatorial clássica então,

$$\|T(x) - T(y)\| \leq \|DT(y)\|_{OP} \|x - y\| + \|\rho(x, y)\| \|x - y\|$$

Dessa forma,

$$\|T(x) - T(y)\| \leq (\|DT(y)\|_{OP} + \|\rho(x, y)\|) \|x - y\|$$

Assim, se encontrarmos um conjunto fechado $F \subset E$ tal que $\|DT(y)\|_{OP} + \|\rho(x, y)\| \leq \alpha < 1$ o Teorema de Banach pode ser aplicado, garantindo a existência de um ponto fixo em F . Assim, como $\rho(x, y)$ é pequeno desde que $\|x - y\|$ seja pequeno, é necessário um controle sobre $\|DT(y)\|_{OP}$. Mas,

$$DT(y) = I - JDf(y)$$

Assim, se tomarmos J uma aplicação linear invertível tal que $JDf(y) \approx I$, teremos $DT(y) \approx 0$ e consequentemente $\|DT(y)\|_{OP} \approx 0$. Dessa forma, para que seja possível aplicar o Teorema do Ponto Fixo de Banach, a escolha de J do Lema 2.1 não é livre, mas sim uma aproximação para $Df^{-1}(y)$. No caso do método de Newton, o esquema iterativo é idêntico ao do Teorema 2.2, porém o operador T varia a cada iteração de modo que $x_{k+1} = T^k(x_k)$ em que $T^k(x_k) = x_k - Df^{-1}(x_k)f(x_k)$. Neste trabalho o operador T é fixo, razão pela qual essa categoria de métodos recebe o nome de *Métodos Quasi-Newton*.

2.3 Construção do Espaço de Banach Utilizado

Definindo a função ω_k , por (2.3), e fixado $s > 0$ podemos definir por (2.4) uma norma $\|\bullet\|_s : X \rightarrow \mathbb{R}^+$ para o espaço vetorial X . A escolha de s não é livre e está relacionada ao decaimento dos coeficientes da série de Fourier da aplicação A como pode ser observado no Lema 2.2. Neste trabalho, $|R|_\infty = \max(|r_{ij}|)$ denota o maior dentre os módulos das entradas da matriz R .

$$\omega_k = \begin{cases} |k|; & k \neq 0 \\ 1; & k = 0 \end{cases} \quad (2.3)$$

$$\|x\|_s = \sup \{|x_k| (\omega_k)^s\} = \sup \left\{ |R|_\infty, |Q_0|_\infty, \sup_{k \geq 1} \{|Q_{k,1}|_\infty (\omega_k)^s, |Q_{k,2}|_\infty (\omega_k)^s\} \right\} \quad (2.4)$$

A motivação para definir a norma dada pela Equação anterior é atribuir norma finita apenas à séries de Fourier com decaimento mais rápido que k^s para s especificado. A Proposição 2.4 prova o fato de que Ω^s definido por (2.5) é um subespaço normado completo de X e portanto é um Espaço de Banach.

$$\Omega^s = \{x \in X, \|x\|_s < \infty\} \quad (2.5)$$

Proposição 2.4. *O subconjunto Ω^s definido por (2.5) munido da norma definida por (2.4) é um Espaço de Banach.*

Demonstração. A demonstração será dividida em três partes:

$\|\bullet\|_s$ é uma norma: Primeiramente é necessário provar que $\|\bullet\|_s$ é uma norma. De fato, $\|x\|_s = 0 \Leftrightarrow x = 0$, uma vez que se $\|x\|_s = 0$ então não pode haver nenhuma entrada de coordenada de x diferente de zero. $\|x\|_s \geq 0$ por sua definição. Para provar que vale a desigualdade triangular, sejam $x = ((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots)$ e $y = ((r, q_0), (q_{1,1}, q_{1,2}), \dots)$

$$\begin{aligned} \|x + y\|_s &= \max \{ |R + r|_\infty, |Q_0 + q_0|_\infty, \sup \{ |Q_{k,1} + q_{k,1}|_\infty \omega_k^s, |Q_{k,2} + q_{k,2}|_\infty \omega_k^s \} \} \\ &\leq \max \{ |R|_\infty, |Q_0|_\infty, \sup \{ |Q_{k,1}|_\infty \omega_k^s, |Q_{k,2}|_\infty \omega_k^s \} \} + \\ &\quad + \max \{ |r|_\infty, |q_0|_\infty, \sup \{ |q_{k,1}|_\infty \omega_k^s, |q_{k,2}|_\infty \omega_k^s \} \} = \|x\|_s + \|y\|_s \end{aligned}$$

Se $\lambda \in \mathbb{R}$ então

$$\|\lambda x\|_s = \max \{ |\lambda R|_\infty, |\lambda Q_0|_\infty, \sup \{ |\lambda Q_{k,1}|_\infty \omega_k^s, |\lambda Q_{k,2}|_\infty \omega_k^s \} \} = |\lambda| \|x\|_s$$

Ω^s é um subespaço vetorial de X : Se $x, y \in \Omega^s$ então $\|x + y\|_s \leq \|x\|_s + \|y\|_s < \infty$, logo $x + y \in \Omega^s$. Além disso, se $\lambda \in \mathbb{R}$, $\|\lambda x\|_s = |\lambda| \|x\|_s < \infty$, provando que Ω^s é um subespaço vetorial normado de X .

Ω^s é completo: Seja $(x_m)_{m \geq 1}$ uma seqüência de Cauchy em Ω^s . Defina $(r^{ij})_m, (q_0^{ij})_m, (q_{k,1}^{ij})_m$ e $(q_{k,2}^{ij})_m$ as seqüências formadas pelas entradas das matrizes $R, Q_0, Q_{k,1}$ e $Q_{k,2}$, respectivamente, de cada x_m . Afirmamos que cada entrada converge. De fato, pela definição da norma $\|\bullet\|_s$ como um supremo de valores absolutos o fato de que $(x_m)_m$ é uma seqüência de Cauchy implica imediatamente o fato de que $(r^{ij})_m, (q_0^{ij})_m, (q_{k,1}^{ij})_m, (q_{k,2}^{ij})_m$ forma seqüências de Cauchy em \mathbb{R} . Desse modo, como \mathbb{R} é completo cada uma das infinitas seqüências de entradas converge para um número real, fazendo com que $x_m \rightarrow x \in X$. Resta provar que $x \in \Omega^s$. De fato, como $x_m \rightarrow x$ existe $N_0 \in \mathbb{N}$ tal que para todo $m \geq N_0$ temos $x_m \in B_x(1) \subset X$. Tome $d \geq N_0$ e portanto $x_d \in B_x(1)$. Assim $x \in B_{x_d}(1)$ e como $x_d \in \Omega^s$ por ser um dos termos da seqüência (x_m) então $\|x\|_s \leq \|x_d\|_s + 1$. Logo $x \in \Omega^s$, como queríamos. \square

Proposição 2.5. Para cada $s > 0$ vale que $\Omega^s \supset \Omega^{s+1}$.

Demonstração. Seja $x \in \Omega^{s+1}$, então $\sup_{k \geq 0} |x_k|_\infty \omega_k^{s+1} < \infty$. Como pela definição de $\omega_k, \omega_k^{s+1} \geq \omega_k^s \forall s > 0$ segue que $\|x\|_{s+1} \geq \|x\|_s$. Assim, $\Omega^s \supset \Omega^{s+1}$. \square

O Lema 2.2 estabelece a condição necessária sobre a série de Fourier de A para que a solução da Equação $f(x) = 0$ esteja em Ω^s . Observe que se $A(t)$ for analítica $Q(t)$ também será e dessa forma os coeficientes da série de Fourier de $Q(t)$ decairão mais rapidamente que qualquer potência, fazendo com que $((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots) \in \Omega^s$ para todo $s \geq 1$. No entanto, não é necessário que $A(t)$ seja analítica para concluir que a solução de $f(x) = 0$ seja um elemento de Ω^s . Este Lema conduz à escolha de s , que até então não estava definido. Para $A(t)$ analítica $s > 0$ pode ser escolhido arbitrário.

Lema 2.2. Seja A_k o k -ésimo termo da série de Fourier de $A(t)$. Se existe $C > 0$ e $s > 0$ tais que $|A_k|_\infty < C \omega_k^{-s}$ então para todo $x \in X$ solução de $f(x) = 0$ tem-se que $x \in \Omega^s$.

Demonstração. Considere o termo $Q_k \in M(n, \mathbb{C})$ da expansão em série de Fourier de $Q(t)$.

$$\frac{1}{2\tau} \int_0^{2\tau} Q(t) e^{-ik \frac{\pi}{\tau} t} dt = \frac{1}{2\tau} \left[-\frac{\tau}{ik\pi} Q(t) e^{-ik \frac{\pi}{\tau} t} \right]_0^{2\tau} + \frac{1}{2k\pi} \int_0^{2\tau} Q'(t) e^{-ik \frac{\pi}{\tau} t} dt$$

Substituindo a condição de que $Q(t)e^{tR}$ é solução de $Q' = A(t)Q - QR$ temos que

$$\frac{1}{2\tau} \int_0^{2\tau} Q(t) e^{-ik \frac{\pi}{\tau} t} dt = \frac{1}{2k\pi} \int_0^{2\tau} A(t) Q(t) e^{-ik \frac{\pi}{\tau} t} dt - \frac{1}{2k\pi} \int_0^{2\tau} Q(t) R e^{-ik \frac{\pi}{\tau} t} dt$$

Assim, tomando a norma $|\bullet|_\infty$ de ambos os lados fica:

$$\left| \int_0^{2\tau} Q(t) e^{-ik\frac{\pi}{\tau}t} dt \left[\frac{1}{2\tau}I + \frac{1}{2k\pi}R \right] \right|_\infty = \left| \frac{1}{2k\pi} \int_0^{2\tau} A(t) Q(t) e^{-ik\frac{\pi}{\tau}t} dt \right|_\infty$$

Denotando $A(t) = (a_{ij}(t))$ e $Q(t) = (q_{ij}(t))$ temos que, pela definição de $|\bullet|_\infty$,

$$\left| \int_0^{2\tau} A(t) Q(t) e^{-ik\frac{\pi}{\tau}t} dt \right|_\infty = \max_{i,j \leq n} \left\{ \left| \int_0^{2\tau} [A(t) Q(t)]_{ij} e^{-ik\frac{\pi}{\tau}t} dt \right| \right\} = \max_{i,j \leq n} \left\{ \left| \int_0^{2\tau} \sum_{r=1}^n a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right\}$$

usando a linearidade da integração,

$$\max_{i,j \leq n} \left\{ \left| \int_0^{2\tau} \sum_{r=1}^n a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right\} = \max_{i,j \leq n} \left\{ \left| \sum_{r=1}^n \left[\int_0^{2\tau} a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right] \right| \right\}$$

e pela desigualdade triangular,

$$\max_{i,j \leq n} \left\{ \left| \sum_{r=1}^n \left[\int_0^{2\tau} a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right] \right| \right\} \leq \max_{i,j \leq n} \left\{ \sum_{r=1}^n \left| \int_0^{2\tau} a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right\}$$

Dessa forma, é claro que

$$\max_{i,j \leq n} \left\{ \left| \sum_{r=1}^n \left[\int_0^{2\tau} a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right] \right| \right\} \leq \max_{i,j \leq n} \left\{ \sum_{r=1}^n \left[\left| \int_0^{2\tau} a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right] \right\}$$

e logo

$$\max_{i,j \leq n} \left\{ \sum_{r=1}^n \left[\left| \int_0^{2\tau} a_{ir}(t) q_{rj}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right] \right\} \leq \max_{i \leq n} \left\{ \sum_{r=1}^n \left[\left(\max_{t \in [0, 2\tau]} |q_{rj}(t)| \right) \left| \int_0^{2\tau} a_{ir}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right] \right\}$$

definindo $G = \max_{t \in [0, 2\tau], r \leq n} |q_{rj}(t)|$ fica

$$\max_{i \leq n} \left\{ \sum_{r=1}^n \left[\left(\max_{t \in [0, 2\tau]} |q_{rj}(t)| \right) \left| \int_0^{2\tau} a_{ir}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right] \right\} \leq G \max_{i \leq n} \left\{ \sum_{r=1}^n \left[\left| \int_0^{2\tau} a_{ir}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right] \right\}$$

assim, existe $D \in \mathbb{R}$ com

$$G \max_{i \leq n} \left\{ \sum_{r=1}^n \left[\left| \int_0^{2\tau} a_{ir}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right] \right\} \leq nG \max_{i,r \leq n} \left\{ \left| \int_0^{2\tau} a_{ir}(t) e^{-ik\frac{\pi}{\tau}t} dt \right| \right\} \leq nG |A_k|_\infty \leq nG C k^{-s} = D k^{-s}$$

Provando que existe $D > 0$ tal que

$$\left| \int_0^{2\tau} Q(t) e^{-ik\frac{\pi}{\tau}t} dt \left[\frac{1}{2\tau}I + \frac{1}{2k\pi}R \right] \right|_\infty \leq D k^{-s}$$

O passo fundamental da demonstração é observar que os termos fora da diagonal principal de $\left[\frac{1}{2\tau}I + \frac{1}{2k\pi}R \right]$ se aproximam de zero quando k cresce. Passando o limite com $k \rightarrow \infty$ segue que

$$\lim_{k \rightarrow \infty} \left| \omega_k^s \int_0^{2\tau} Q(t) e^{-ik\frac{\pi}{\tau}t} dt \left[\frac{1}{2\tau}I + \frac{1}{2k\pi}R \right] \right|_\infty = \lim_{k \rightarrow \infty} \frac{1}{2\tau} |Q_k \omega_k^s|_\infty \leq D$$

Uma vez que o limite de $|Q_k \omega_k^s|_\infty$ é finito, existe $K_0 > 0$ tal que para todo $k \geq K_0$, $|Q_k \omega_k^s|_\infty \leq 2\tau D + 1$. Defina

$$M = \max \left\{ |R|_\infty, |Q_0|_\infty, \max_{1 \leq k \leq K_0} \{ |Q_{k,1}|_\infty \omega_k^s, |Q_{k,2}|_\infty \omega_k^s \}, 2\tau D + 1 \right\}$$

Desse modo, $\|(R, Q_0, Q_{1,1}, Q_{1,2}, \dots)\| \leq M < \infty$, provando que $x \in \Omega^s$, como queríamos demonstrar. \square

Seja $\mathcal{A} = (A_k)_{k \geq 0}$ a sequência formada pelos coeficientes da série de Fourier de $A(t)$. Então defina uma extensão de s -norma, dada por

$$\|\mathcal{A}\|_s = \sup_{k \geq 0} |A_k|_\infty \omega_k^s$$

Lema 2.3. *Se $\|\mathcal{A}\|_p < \infty$ para $p \geq 2$. Então $f(\Omega^s) \subset \Omega^{s-1} \forall 2 \leq s \leq p$.*

Demonstração. Seja $2 \leq s \leq p$ e suponha que $x \in \Omega^s$. Assim, como $\|\mathcal{A}\|_s < \infty$ tem-se que existe $C_1 > 0$ tal que para todo $k \geq 1$, $|A_k| < C_1 \omega_k^{-s}$. Desse modo, pelo Lema 2.1 de [11], $|(AQ)_k|_\infty \leq \frac{D}{\omega_k^s}$ para algum $C_2 > 0$. Assim usando a definição de $f_k = (f_{k,1}, f_{k,2})$ tem-se que

$$|f_k(x)|_\infty \leq C_3 k |Q_k|_\infty + C_4 |Q_k|_\infty + C_2 \omega_k^{-s}$$

Usando e o fato de que $|A_k| < C \omega_k^{-s}$ conclui-se que $|Q_k|_\infty \omega_k^s \leq G$ para algum $G > 0$, de modo que $|Q_k|_\infty < G \omega_k^{-s}$. Assim, $C_3 k |Q_k|_\infty < C_3 G \omega_k^{-s+1}$, $C_4 |Q_k|_\infty < C_4 G \omega_k^{-s+1}$ e $C_2 \omega_k^{-s} < C_2 \omega_k^{-s+1}$. Tome $C > \max(C_3 G, C_4 G, C_2)$, o que conduz a $f_k(x) < C \omega_k^{-s+1}$, completando a demonstração. \square

2.4 Definição do Operador T e Condições para Existência de Ponto Fixo

O operador $T : X \rightarrow X$ apresentado na Seção 2.2 é definido como $T(x) = x - Jf(x)$ em que $J : X \rightarrow X$ é um operador linear invertível. No entanto, para que seja possível aplicar o Teorema do Ponto Fixo de Banach é necessário que $T(\Omega^s) \subset \Omega^s$ e que ainda que T seja uma contração quando restrito a um conjunto fechado $F \subset \Omega^s$. Esta seção trata da definição completa desse operador.

2.4.1 Definição do Operador T

Seja $m \in \mathbb{N} \setminus \{0\}$ fixo, o conjunto $X^m = [M(n, \mathbb{R}) \times M(n, \mathbb{R})]^m$ e a aplicação $\Pi_m : X \rightarrow X^m$ definida por

$$\Pi_m((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots, (Q_{k,1}, Q_{k,2}), \dots) = ((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots, (Q_{m-1,1}, Q_{m-1,2})) \quad (2.6)$$

Além disso, defina $0^\infty = ((0, 0), (0, 0), \dots)$. Dado $x^m \in X_m$ pode-se obter um ponto $x \in X$ adjuntando zeros, ou seja, $x = (x^m, 0^\infty) \in X$. Uma vez definido X^m , define-se a aplicação $f^{(m)} : X^m \rightarrow X^m$ por (2.7).

$$f^{(m)}(x^m) = \Pi_m(f((x^m, 0^\infty))) \quad (2.7)$$

Dessa forma, um problema aproximado em dimensão finita consiste em obter $\tilde{x} \in X^m$ tal que $f^{(m)}(\tilde{x}) = 0$. No entanto, como o objetivo é utilizar um método numérico para a solução do problema, será possível apenas aproximar a solução do problema em dimensão finita, ou seja, obter $\bar{x} \in X^m$ tal que $f^{(m)}(\bar{x}) \approx 0$.

Calculando a derivada da aplicação $f^{(m)}$ em \bar{x} e aplicando a um ponto $v \in X^m$ temos que

$$Df^{(m)}(\bar{x})v = \left(\left(Df_*^{(m)}(\bar{x}) \cdot v, Df_0^{(m)}(\bar{x}) \cdot v \right), \left(Df_{1,1}^{(m)}(\bar{x}) \cdot v, Df_{1,2}^{(m)}(\bar{x}) \cdot v \right), \dots, \left(Df_{m-1,1}^{(m)}(\bar{x}) \cdot v, Df_{m-1,2}^{(m)}(\bar{x}) \cdot v \right) \right)$$

Para utilizar uma notação matricial vamos assumir a correspondência biunívoca abaixo

$$((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots, (Q_{m-1,1}, Q_{m-1,2})) \leftrightarrow \begin{bmatrix} R & Q_0 & Q_{1,1} & Q_{1,2} & \dots & Q_{m-1,1} & Q_{m-1,2} \end{bmatrix}^T$$

A Equação 2.9 representa a derivada de $f^{(m)}$ calculada no ponto \bar{x} . A notação dada por (2.8) é utilizada para a determinação da derivada

$$\begin{aligned} \frac{\partial \Delta}{\partial (R, Q_0)}(\bar{x}) &= \begin{bmatrix} \frac{\partial f_*}{\partial R}(\bar{x}) & \frac{\partial f_*}{\partial Q_0}(\bar{x}) \\ \frac{\partial f_0}{\partial R}(\bar{x}) & \frac{\partial f_0}{\partial Q_0}(\bar{x}) \end{bmatrix} & \frac{\partial \Delta}{\partial Q_j}(\bar{x}) &= \begin{bmatrix} \frac{\partial f_*}{\partial Q_{j,1}}(\bar{x}) & \frac{\partial f_*}{\partial Q_{j,2}}(\bar{x}) \\ \frac{\partial f_0}{\partial Q_{j,1}}(\bar{x}) & \frac{\partial f_0}{\partial Q_{j,2}}(\bar{x}) \end{bmatrix} \\ \frac{\partial f_k}{\partial (R, Q_0)}(\bar{x}) &= \begin{bmatrix} \frac{\partial f_{k,1}}{\partial R}(\bar{x}) & \frac{\partial f_{k,1}}{\partial Q_0}(\bar{x}) \\ \frac{\partial f_{k,2}}{\partial R}(\bar{x}) & \frac{\partial f_{k,2}}{\partial Q_0}(\bar{x}) \end{bmatrix} & \frac{\partial f_k}{\partial Q_j}(\bar{x}) &= \begin{bmatrix} \frac{\partial f_{k,1}}{\partial Q_{j,1}}(\bar{x}) & \frac{\partial f_{k,1}}{\partial Q_{j,2}}(\bar{x}) \\ \frac{\partial f_{k,2}}{\partial Q_{j,1}}(\bar{x}) & \frac{\partial f_{k,2}}{\partial Q_{j,2}}(\bar{x}) \end{bmatrix} \end{aligned} \quad (2.8)$$

$$Df^{(m)}(\bar{x}) = \begin{bmatrix} \frac{\partial \Delta}{\partial (R, Q_0)}(\bar{x}) & \frac{\partial \Delta}{\partial Q_1}(\bar{x}) & \dots & \frac{\partial \Delta}{\partial Q_{m-1}}(\bar{x}) \\ \frac{\partial f_1}{\partial (R, Q_0)}(\bar{x}) & \frac{\partial f_1}{\partial Q_1}(\bar{x}) & \dots & \frac{\partial f_1}{\partial Q_{m-1}}(\bar{x}) \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial f_{m-1}}{\partial (R, Q_0)}(\bar{x}) & \frac{\partial f_{m-1}}{\partial Q_1}(\bar{x}) & \dots & \frac{\partial f_{m-1}}{\partial Q_{m-1}}(\bar{x}) \end{bmatrix} \quad (2.9)$$

Aplicando a derivada de $f^{(m)}(\bar{x})$ ao vetor $V = (r, q_0, q_{1,1}, q_{1,2}, \dots, q_{m-1,1}, q_{m-1,2})$ temos o resultado apresentado na Equação 2.10.

$$Df^{(m)}(\bar{x})V = \begin{bmatrix} \frac{\partial f_*}{\partial R}r + \frac{\partial f_*}{\partial Q_0}q_0 + \dots + \frac{\partial f_*}{\partial Q_{m-1,1}}q_{m-1,1} + \frac{\partial f_*}{\partial Q_{m-1,2}}q_{m-1,2} \\ \frac{\partial f_0}{\partial R}r + \frac{\partial f_0}{\partial Q_0}q_0 + \dots + \frac{\partial f_0}{\partial Q_{m-1,1}}q_{m-1,1} + \frac{\partial f_0}{\partial Q_{m-1,2}}q_{m-1,2} \\ \dots \\ \frac{\partial f_{m-1,1}}{\partial R}r + \frac{\partial f_{m-1,1}}{\partial Q_0}q_0 + \dots + \frac{\partial f_{m-1,1}}{\partial Q_{m-1,1}}q_{m-1,1} + \frac{\partial f_{m-1,1}}{\partial Q_{m-1,2}}q_{m-1,2} \\ \frac{\partial f_{m-1,2}}{\partial R}r + \frac{\partial f_{m-1,2}}{\partial Q_0}q_0 + \dots + \frac{\partial f_{m-1,2}}{\partial Q_{m-1,1}}q_{m-1,1} + \frac{\partial f_{m-1,2}}{\partial Q_{m-1,2}}q_{m-1,2} \end{bmatrix} \quad (2.10)$$

Uma vez calculada numericamente a matriz jacobiana $Df^{(m)}(\bar{x})$, pode-se calcular uma aproximação numérica invertível para $[Df^{(m)}]^{-1}$, i.e., $J_m \in M(2n^2m, \mathbb{R})$ tal que $J_m \cdot Df^{(m)} \approx I$. Estão definidas as derivadas parciais de cada componente de $f^{(m)}$ em relação às variáveis correspondentes ao espaço de dimensão finita obtido após a projecção. No entanto, as derivadas de todas as componentes de $f = (f_*, f_0, f_1, \dots, f_n, \dots)$ estão bem definidas em relação a todas as componentes de $x = ((R, Q_0), (Q_{1,1}, Q_{1,2}), \dots)$. Assim, para $k \geq m$ denote

$$\Lambda_k = \frac{\partial f_k}{\partial Q_k}(\bar{x}) = \begin{bmatrix} \frac{\partial f_{k,1}}{\partial Q_{k,1}}(\bar{x}) & \frac{\partial f_{k,1}}{\partial Q_{k,2}}(\bar{x}) \\ \frac{\partial f_{k,2}}{\partial Q_{k,1}}(\bar{x}) & \frac{\partial f_{k,2}}{\partial Q_{k,2}}(\bar{x}) \end{bmatrix}.$$

A Definição 2.1 e o Lema 2.4 são resultados instrumentais, sendo utilizados na demonstração do Lema 2.5.

Definição 2.1 (Diagonal dominância). *Seja $A = (a_{ij})$ uma matriz $n \times n$ com entradas complexas. A é dita diagonal dominante (ou diagonal dominante por linhas) se*

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}|$$

Se valer ($>$) ao invés de \geq então é dita estritamente diagonal dominante (por linhas).

O Teorema 2.3 demonstrado por S. Gerschgorin em [12] estabelece as regiões do plano complexo às quais podem pertencer os autovalores de matrizes complexas.

Teorema 2.3 (de Gerschgorin). *Seja A uma matriz complexa $n \times n$ com entradas a_{ij} , $i = 1, 2, \dots, n$. Defina $R_i = \sum_{j \neq i} |a_{ij}|$. Então, todo valor característico λ de A é tal que $|\lambda - a_{ii}| \leq R_i$ para algum $i = 1, 2, \dots, n$. Os conjuntos $D_i = \{z \in \mathbb{C}; |z - a_{ii}| \leq R_i\}$ é denominado Disco de Gerschgorin.*

Demonstração. Sejam $\lambda \in \mathbb{C}$ um valor característico de A e $x \in \mathbb{C}$ tal que $Ax = \lambda x$ com $x = (x_1, \dots, x_n) \neq 0$. Denote $|x_i| = \max\{|x_j|, j = 1, 2, \dots, n\}$. Assim, $\sum_{j=1}^n a_{ij}x_j = \lambda x_i, \forall i = 1, 2, \dots, n$. Ou seja, $\sum_{j \neq i} a_{ij}x_j = \lambda x_i - a_{ii}x_i$.

Mas $x \neq 0 \implies x_i \neq 0$ e pela definição de x_i , $\frac{|x_j|}{|x_i|} \leq 1$. Assim, aplicando a desigualdade triangular temos que

$$|\lambda - a_{ii}| = \left| \frac{\sum_{j \neq i} a_{ij} x_j}{x_i} \right| = \left| \sum_{j \neq i} a_{ij} \frac{x_j}{x_i} \right| \leq \sum_{j \neq i} \left| a_{ij} \frac{x_j}{x_i} \right| = \sum_{j \neq i} |a_{ij}| \left| \frac{x_j}{x_i} \right| \leq \sum_{j \neq i} |a_{ij}| = R_i$$

completando a demonstração. \square

Lema 2.4 (Invertibilidade via Diagonal-dominância). *Sejam $A = (a_{ij})$ uma matriz $n \times n$ estritamente diagonal dominante e $B = (b_{ij})$ uma matriz $n \times m$. Então A é invertível e vale que*

$$|A^{-1}B|_{\infty} \leq \max_i \frac{\sum_{j=1}^m |b_{ij}|}{|a_{ii}| - \sum_{j \neq i} |a_{ij}|}$$

Em particular $|A^{-1}|_{\infty} \leq \max_i \frac{1}{|a_{ii}| - \sum_{j \neq i} |a_{ij}|}$

Demonstração. Antes de demonstrar a desigualdade, é necessário argumentar sobre a invertibilidade de A . Uma matriz diagonal dominante sempre é não-singular, resultado conhecido como Teorema de Levy-Desplanques. Para provar que A é não-singular basta provar que 0 não é autovalor de A . De fato, como A é estritamente diagonal dominante, $|a_{ii}| > R_i$ e portanto 0 não pertence a nenhum Disco de Gershgorin, de modo que, pelo Teorema anterior, 0 não é autovalor de A e, portanto, A é não-singular.

Consideremos $\|\bullet\|_{\infty}$ a norma operatorial definida por $\|K\|_{\infty} = \sup_{|x|_{\infty}=1} |Kx|_{\infty}$. A aplicação $\beta : S^{m-1} \rightarrow \mathbb{R}$ definida por $\beta(x) = |A^{-1}Bx|_{\infty}$ é contínua e está definida no compacto S^{m-1} . Logo existe $x = (x_1, x_2, \dots, x_m)^T \in S^{m-1}$ tal que

$$\|A^{-1}B\|_{\infty} = |A^{-1}Bx|_{\infty} = |y|_{\infty} = |y_{i_0}|$$

em que $y = A^{-1}Bx = (y_1, y_2, \dots, y_n)^T$. Como $y = A^{-1}Bx$, tem-se que $Ay = Bx$, o que implica que

$$\sum_{j=1}^m b_{i_0 j} x_j = (Bx)_{i_0} = (Ay)_{i_0} = \sum_{j=1}^n a_{i_0 j} y_j = a_{i_0 i_0} y_{i_0} + \sum_{j \neq i_0} a_{i_0 j} y_j$$

Aplicando a desigualdade triangular temos que

$$\max_{1 \leq j \leq m} (|x_j|) \sum_{j=1}^m |b_{i_0 j}| \geq \left| \sum_{j=1}^m b_{i_0 j} x_j \right| = \left| a_{i_0 i_0} y_{i_0} + \sum_{j \neq i_0} a_{i_0 j} y_j \right| \geq |a_{i_0 i_0} y_{i_0}| - \left| \sum_{j \neq i_0} a_{i_0 j} y_j \right|$$

Mas como $\left| \sum_{j \neq i_0} a_{i_0 j} y_j \right| \leq \left| \sum_{j \neq i_0} a_{i_0 j} \right| |y_{i_0}|$ vale que

$$\max_{1 \leq j \leq m} (|x_j|) \sum_{j=1}^m |b_{i_0 j}| \geq \left(|a_{i_0 i_0}| - \left| \sum_{j \neq i_0} a_{i_0 j} \right| \right) |y_{i_0}|$$

Como A é estritamente diagonal dominante, $|a_{i_0 i_0}| - \left| \sum_{j \neq i_0} a_{i_0 j} \right| > 0$ temos que

$$|y_{i_0}| \leq \frac{\sum_{j=1}^m |b_{i_0 j}|}{|a_{i_0 i_0}| - \left| \sum_{j \neq i_0} a_{i_0 j} \right|} \max_{1 \leq j \leq m} (|x_j|)$$

Como $|y_{i_0}| = |A^{-1}B|_\infty$ e $\max_{1 \leq j \leq m} (|x_j|) = |x|_\infty = 1$ segue que

$$|A^{-1}B|_\infty \leq \frac{\sum_{j=1}^m |b_{i_0 j}|}{|a_{i_0 i_0}| - \left| \sum_{j \neq i_0} a_{i_0 j} \right|} \leq \max_i \frac{\sum_{j=1}^m |b_{ij}|}{|a_{i_0 i_0}| - \left| \sum_{j \neq i} a_{ij} \right|}$$

Se tomarmos $B = \mathbb{I}$ então $\sum_{j=1}^m |b_{ij}| = 1 \forall i$, o que implica que

$$|A^{-1}|_\infty \leq \max_i \frac{1}{|a_{i_0 i_0}| - \left| \sum_{j \neq i} a_{ij} \right|}$$

completando a demonstração. \square

Lema 2.5. *Seja $\mathcal{A} = ((A_{0,1}, A_{0,2}), (A_{1,1}, A_{1,2}), \dots)$. Se $\|\mathcal{A}\|_s < \infty$ existem $K > 0$ e $C_\Lambda > 0$ tais que para cada $k \geq K$ o operador Λ_k é invertível e $\|(\Lambda_k)^{-1}\|_\infty < \frac{C_\Lambda}{k}$.*

Demonstração. Pela definição

$$\Lambda_k = \frac{\partial f_k}{\partial Q_k} = \begin{bmatrix} \frac{\partial f_{k,1}}{\partial Q_{k,1}} & \frac{\partial f_{k,1}}{\partial Q_{k,2}} \\ \frac{\partial f_{k,2}}{\partial Q_{k,1}} & \frac{\partial f_{k,2}}{\partial Q_{k,2}} \end{bmatrix}$$

Para calcular as derivadas é necessário calcular $(AQ)_{k,1}$ e $(AQ)_{k,2}$. Assim,

$$\begin{cases} (AQ)_k = (A_{2k,1} + iA_{2k,2})(Q_{k,1} - iQ_{k,2}) + A_0(Q_{k,1} + iQ_{k,2}) + W \\ (AQ)_{k,1} = A_{2k,1}Q_{k,1} + A_0Q_{k,1} + A_{2k,2}Q_{k,2} + W_1 \\ (AQ)_{k,2} = -A_{2k,1}Q_{k,2} + A_{2k,2}Q_{k,1} + A_0Q_{k,2} + W_2 \end{cases} \quad (2.11)$$

de modo que W_1 e W_2 não dependem de $Q_{k,1}$ e $Q_{k,2}$.

$$\Lambda_k = \begin{bmatrix} \lambda_{1,1} & -k \frac{2\pi}{2\tau} \mathbb{I}_{n^2} + \lambda_{1,2} \\ k \frac{2\pi}{2\tau} \mathbb{I}_{n^2} + \lambda_{2,1} & \lambda_{2,2} \end{bmatrix}, \quad (2.12)$$

em que as entradas $\lambda_{1,1}$ e $\lambda_{2,2}$ são combinações lineares de \bar{R} , $\mathcal{A}_0, \mathcal{A}_{2k}$ de modo que $|\lambda_{1,1}|_\infty, |\lambda_{2,2}|_\infty < |\bar{R}|_\infty + |\mathcal{A}_0|_\infty + |\mathcal{A}_{2k}|_\infty$. Além disso, $\lambda_{2,1} = \lambda_{1,2}$ dependem apenas de A_{2k} . Como a invertibilidade de Λ_k é equivalente a invertibilidade de $\hat{\Lambda}_k$ dada por:

$$\hat{\Lambda}_k = \begin{bmatrix} k \frac{2\pi}{2\tau} \mathbb{I}_{n^2} + \lambda_{2,1} & \lambda_{2,2} \\ \lambda_{1,1} & -k \frac{2\pi}{2\tau} \mathbb{I}_{n^2} + \lambda_{1,2} \end{bmatrix}.$$

Como $|\lambda_{1,1}|_\infty, |\lambda_{2,2}|_\infty < |\bar{R}|_\infty + |\mathcal{A}_0|_\infty + |\mathcal{A}_{2k}|_\infty$ e $|\lambda_{1,2}|_\infty < |\mathcal{A}_{2k}|_\infty$, a hipótese de que $\|\mathcal{A}\|_s < \infty$ implica que $|\lambda_{i,j}|_\infty$ são uniformemente limitados em k . Dessa forma existe K tal que $\hat{\Lambda}_k$ é diagonal dominante por linhas para $k \geq K$, já que para k suficientemente grande $k \frac{\pi}{\tau}$ é maior que a soma dos valores absolutos de todos os elementos de $\lambda_{i,j}$, $i, j = 1, 2$. Assim, $\det(\hat{\Lambda}_k) \neq 0$, o que implica que $\hat{\Lambda}_k$ é invertível para todo $k \geq K$.

Denote por $a_{i,i}$ os elementos da diagonal principal de $\hat{\Lambda}_k$. Dessa forma, se $\hat{\Lambda}_k$ é diagonal dominante, isto é, se $|a_{i,i}| > \sum_{j \neq i} |\hat{\Lambda}_k(i, j)|$ para todo $i = 1, \dots, 2n^2$, dessa forma usando o resultado o 2.4 temos que

$$\|\hat{\Lambda}_k^{-1}\|_\infty \leq \max_i \left\{ \frac{1}{|a_{i,i}| - \sum_{j \neq i} |\hat{\Lambda}_k(i, j)|} \right\}.$$

Dessa forma, para $k \geq K$

$$\|\Lambda_k^{-1}\|_\infty = \|\hat{\Lambda}_k^{-1}\|_\infty \leq \frac{C_\Lambda}{k}$$

para uma constante C_Λ que depende inicialmente de τ , \bar{R} , $|A_0|_\infty$ e $|A_{2k}|_\infty$. No entanto, deve-se atentar para o fato de que $|A_{2k}|_\infty \leq \frac{\|A\|_{s^*}^s}{\omega_k^s}$. Para o cálculo de C_Λ , afirmamos que existem $a > 0$ e $b \geq 0$ tais que para $k \geq K$

$$\|\hat{\Lambda}_k^{-1}\|_\infty \leq \max_i \left\{ \frac{1}{|a_{ii}| - \sum_{j \neq i} |\hat{\Lambda}_k(i, j)|} \right\} \leq \frac{1}{ak - b}$$

De fato, para $k \geq K$ temos que

$$|a_{ii}(k)| \geq k \frac{\pi}{\tau} - |\lambda_{1,2}(K)|_\infty$$

e

$$\sum_{j \neq i} |\hat{\Lambda}_k(i, j)| \leq (2n-1) (|\bar{R}|_\infty + |A_0|_\infty + |A_{2k}|_\infty) \leq (2n-1) \left(|\bar{R}|_\infty + |A_0|_\infty + \frac{\|A\|_{s^*}^s}{(2K)^{s^*}} \right)$$

$$\|\Lambda_k^{-1}\|_\infty \leq \frac{1}{k \frac{\pi}{\tau} - \left[|\lambda_{1,2}(K)|_\infty + (2n-1) \left(|\bar{R}|_\infty + |A_0|_\infty + \frac{\|A\|_{s^*}^s}{(2K)^{s^*}} \right) \right]}$$

em que $\lambda_{1,2}(K)$ é o termo $\Lambda_{1,2}$ quando $k = K$. Assim, ficam definidos a e b . Queremos determinar C_Λ de modo que $\frac{1}{ak-b} \leq \frac{C_\Lambda}{k}$ para $k \geq K$. Mas, veja que $\frac{1}{ak-b} \leq \frac{C_\Lambda}{k} \Rightarrow C_\Lambda \geq \frac{k}{ak-b}$. Assim, como

$$C_\Lambda \geq \frac{1}{a} \frac{ak - b + b}{ak - b} = \frac{1}{a} + \frac{b}{a} \left(\frac{1}{ak - b} \right)$$

e $\frac{1}{ak-b}$ decresce quando k cresce, basta tomar

$$C_\Lambda = \frac{1}{a} + \frac{b}{a} \left(\frac{1}{aK - b} \right)$$

em que os termos a e b não dependem de k □

Uma vez que \mathcal{A} é dado, pelo Lema 2.5 existe K tal que para todo $k \geq K$ Λ_k é invertível. Dessa forma, tomando $m > K$ e definindo o operador J dado pela Equação 2.13.

$$(Jx)_k = \begin{cases} (J_m x^m)_k, & \text{se } k = 0, \dots, m-1 \\ \Lambda_k^{-1} x_k, & \text{se } k \geq m \end{cases} \quad (2.13)$$

A Figura 2.1 representa graficamente a construção do operador J como uma matriz infinita composta por quatro blocos, dois nulos, um bloco de dimensão finita formado por $J_m \approx [Df^{(m)(\bar{x})}]^{-1}$ e um bloco diagonal. Naturalmente, trata-se de uma representação sem validade matemática, uma vez que uma matriz com uma quantidade infinita de linhas e colunas não é um objeto comumente definido. No entanto, essa interpretação ajuda na compreensão da construção realizada. Uma vez definido o operador J , define-se o operador $T : X \rightarrow X$ pela Equação 2.14.

$$T(x) = x - Jf(x) \quad (2.14)$$

Teorema 2.4. *Seja $\mathcal{A} = ((A_{0,1}, A_{0,2}, (A_{1,1}, A_{1,2}), \dots)) \in X$ é tal que $\|\mathcal{A}\|_p < \infty$ para $p \geq 2$. Então para cada $2 \leq s \leq p$ tem-se que $T(\Omega^s) \subset \Omega^s$.*

Demonstração. Como as condições do Lema 2.3 são satisfeitas, se $x \in \Omega^s$, $f(x) \in \Omega^{s-1}$. Pela definição de $T(x) = x - Jf(x)$. Assim, é suficiente mostrar que para cada $y \in \Omega^{s-1}$ temos $Jy \in \Omega^s$. De fato, se $y \in \Omega^{s-1}$, tem-se que $\|y\|_{s-1} < \infty$. Uma vez que o operador T já está definido, já está escolhido o valor de m e portanto

$$J = \begin{bmatrix} J_m = [Df^{(m)}]^{-1} & 0 \\ 0 & \Lambda_m^{-1} \Lambda_{m+1}^{-1} \Lambda_{m+2}^{-1} \dots \end{bmatrix}$$

Figura 2.1: Representação de $J : X \rightarrow X$ como uma matriz de dimensão infinita.

seja $k \geq m$. Assim $|(Jy)_k|_\infty = |\Lambda_k^{-1}y_k|_\infty \leq \|\Lambda_k^{-1}\|_\infty|y_k|_\infty$. Como existe $K > 0$ tal que $\forall k \geq K$ temos $\|\Lambda_k^{-1}\|_\infty \leq \frac{C_\Lambda}{k}$ para algum $C_\Lambda > 0$ e existe uma quantidade finita de números entre m e K , pode-se escolher C convenientemente de modo que $\|\Lambda_k^{-1}\|_\infty < \frac{C}{k}$ para cada $k \geq m$. Mas como $|(Jy)_k|_\infty \leq \frac{C}{k} \frac{\|y\|_{s-1}}{\omega_k^{s-1}}$ para $k \geq m$, segue que $|(Jy)_k|_\infty \leq \frac{C\|y\|_{s-1}}{\omega_k^{s-1}}$ para $k \geq m$. Como $y \in \Omega^{s-1}$ e existe apenas um número finito de índices menores ou iguais que m segue que existe $M > 0$ tal que $|(Jy)_k\omega_k^s|_\infty < M$, o que significa que $\|Jy\|_s < \infty$. Dessa forma, provamos que $T(\Omega^s) \subset \Omega^s$, como queríamos. \square

2.4.2 Existência de ponto fixo para o operador $T(x)$

Esta seção é dedicada à demonstração do Teorema 2.5, que estabelece condições suficientes para a Existência de uma vizinhança em Ω^s na qual o operador T seja uma contração, de modo que seja possível garantir através do Teorema 2.1 a existência de um ponto fixo. Este resultado segue a metodologia apresentada em [11], exposta no Lema 3.3. O mesmo procedimento é repetido em [5]. Sejam $Y = (Y_{0,1}, Y_{0,2}, Y_{1,1}, Y_{1,2}, \dots, Y_{k,1}, Y_{k,2}, \dots)$ e $Z(r) = (Z_{0,1}, Z_{0,2}, Z_{1,1}, Z_{1,2}, \dots, Z_{k,1}, Z_{k,2}, \dots)(r)$, i.e., $Y, Z \in X$ tais que valem as condições das Equações 2.15 e 2.16, em que o símbolo \leq_{cw} tem o seguinte significado: $A, B \in M(n, \mathbb{R})$ são tais que $A \geq_{cw} B$ se e somente se $a_{ij} \geq b_{ij} \forall i, j = 1, 2, \dots, n$. Para facilitar a notação Y e Z serão também denotados como $Y = (Y_0, Y_1, Y_2, \dots, Y_k, \dots)$ e $Z(r) = (Z_0, Z_1, Z_2, \dots, Z_k, \dots)(r)$ em que cada Y_i e Z_i representa uma dupla, $(Y_{i,1}, Y_{i,2})$ e $(Z_{i,1}, Z_{i,2})$ respectivamente. Neste ponto convém lembrar que \bar{x} consiste da solução obtida para o problema aproximado em dimensão finita completada com 0^∞ , ou seja $\bar{x} = (\bar{x}_0, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_m, 0, 0, \dots)$ em que $\bar{x}_i \in M(n\mathbb{R}) \times M(n, \mathbb{R})$.

$$|(T(\bar{x}) - \bar{x})_k| \leq_{cw} Y_k \quad (2.15)$$

$$\sup_{b_1, b_2 \in B_0(r)} |(DT(\bar{x} + b_1) b_2)_k| \leq_{cw} Z_k(r) \quad (2.16)$$

As condições de 2.15 e 2.16 impõe formas de controle sobre o quanto um ponto dista de sua imagem pelo operador T em termos de suas coordenadas e o resultado da aplicação da derivada em um ponto da bola de centro \bar{x} e raio r aplicada a um vetor de norma menor que r . O teorema que segue estabelece uma condição sobre Y e Z para que o operador T seja uma contração em $B_{\bar{x}}(r)$.

Teorema 2.5 (Existência de ponto fixo). *Fixado $s \geq 2$ e sejam Y e Z atendendo às condições dadas por (2.15) e (2.16) para todo $k \geq 0$. Se existe $r > 0$ tal que $\|Y + Z\|_s < r$ então $T(B_{\bar{x}}(r)) \subset B_{\bar{x}}(r)$ e existe um único $x^* \in B_{\bar{x}}(r)$ tal que $T(x^*) = x^*$.*

Demonstração. A primeira parte desta demonstração consiste na adaptação do Teorema do Valor Médio para o problema em questão. Em [11] é apresentado um argumento sobre uma versão do Teorema do Valor Médio para funções com valores em matrizes, diferente da desigualdade do valor médio demonstrada em [17]. No entanto, a igualdade não é necessária para os propósitos deste trabalho, sendo suficiente uma adaptação do Teorema do Valor Médio para funções reais de variável real, conforme o argumento que segue..

Sejam $x, y \in X$. Considere a função $\gamma_k : [0, 1] \rightarrow M(n, \mathbb{R})$ definida por $\gamma_k(t) = T_k(ty + (1-t)x)$. Além disso, defina $\gamma_k^{ij} : [0, 1] \rightarrow \mathbb{R}$ como o elemento da linha i e coluna j de $\gamma_k(t)$. Assim, pelo Teorema do Valor Médio temos que $\forall i, j = 1, 2, \dots, n$ existe $t_{ij} \in (0, 1)$ tal que

$$\gamma_k^{ij}(1) - \gamma_k^{ij}(0) = \frac{d}{dt} \gamma_k^{ij}(t_{ij})(1-0)$$

Como $\gamma_k'(t) = DT_k(ty + (1-t)x)(y-x)$, segue que existe para cada $i, j = 1, 2, \dots, n$ existe $t_{ij} \in (0, 1)$ tal que

$$\gamma_k^{ij}(1) - \gamma_k^{ij}(0) = [DT_k(t_{ij}y + (1-t_{ij})x)(y-x)]_{ij}$$

ou seja

$$[T_k(x) - T_k(y)]_{ij} = [DT_k(t_{ij}y + (1-t_{ij})x)(y-x)]_{ij}$$

que implica em

$$\left| [T_k(x) - T_k(y)]_{ij} \right| = \left| \left[DT_k(t_{ij}y + (1-t_{ij})x) \left(\frac{r(y-x)}{\|y-x\|_s} \right) \right]_{ij} \right| \frac{\|y-x\|_s}{r}$$

Mas, por hipótese que para cada entrada i, j vale

$$\left| \left[DT_k(t_{ij}y + (1-t_{ij})x) \left(\frac{r(y-x)}{\|y-x\|_s} \right) \right]_{ij} \right| \leq (Z_k(r))_{ij}$$

que pode ser reescrito como

$$|T_k(x) - T_k(y)| \leq \frac{Z_k}{r} \|y-x\|_s$$

Agora, aplicando a desigualdade triangular em cada entrada tem-se que

$$|(T(x) - \bar{x})_k| = |T_k(x) - \bar{x}_k| = |T_k(x) - T_k(\bar{x}) + T_k(\bar{x}) - \bar{x}_k| \leq_{cw} |T_k(x) - T_k(\bar{x})| + |T_k(\bar{x}) - \bar{x}_k|$$

Mas como $|T_k(x) - T_k(\bar{x})| \leq_{cw} \frac{Z_k(r)}{r} \|x - \bar{x}\|_s$ segue que $|T_k(x) - T_k(\bar{x})| \leq_{cw} Z_k(r)$, o que implica que

$$|T_k(x) - \bar{x}_k| \leq_{cw} Y_k + Z_k(r)$$

Dessa forma, segue que $|T_k(x) - \bar{x}_k|_\infty \leq |Y_k + Z_k(r)|_\infty$. Seja $x \in B_{\bar{x}}(r)$. Assim, temos que:

$$\|T(x) - \bar{x}\|_s = \sup_{k \geq 0} \{|T_k(x) - \bar{x}_k|_\infty \omega_k^s\} \leq \sup_{k \geq 0} \{|Y_k + Z_k(r)|_\infty \omega_k^s\} = \|Y + Z(r)\|_s$$

Por hipótese, $\|Y + Z(r)\|_s < r$, de modo que $\forall x \in B_{\bar{x}}(r)$ tem-se que $\|T(x) - \bar{x}\|_s < r$, ou seja $\forall x \in B_{\bar{x}} T(x) \in B_{\bar{x}}(r)$, ou seja, $T(B_{\bar{x}}(r)) \subset B_{\bar{x}}(r)$.

Para provar que T é de fato uma contração em $B_{\bar{x}}(r)$ observe que $Y_k \geq_{cw} 0, \forall k \geq 0$, ou seja, todas as entradas de Y_k são não-negativas para todo k . Dessa modo, $|Z_k(r)|_\infty \leq |Y_k + Z_k(r)|_\infty$

$$\|Z(r)\|_s = \sup_{k \geq 0} \{|Z_k(r)|_\infty \omega_k^s\} \leq \sup_{k \geq 0} \{|Y_k + Z_k(r)|_\infty \omega_k^s\} = \|Y + Z(r)\|_s < r$$

Definindo $0 < \alpha = \frac{\|Z(r)\|_s}{r} < \frac{r}{r} = 1$ temos que:

$$\|T_k(x) - T_k(y)\|_s \leq \frac{\|Z(r)\|_s}{r} \|x - y\|_s < \alpha \|x - y\|_s$$

Dessa forma, $T(B_{\bar{x}}(r)) \subset B_{\bar{x}}(r)$ e T é uma contração. Usando o Teorema do Ponto Fixo de Banach existe um único $x^* \in B_{\bar{x}}(r)$ tal que $T(x^*) = x^*$, que consequentemente satisfaz $f(x^*) = 0$. \square

2.5 Polinômios *Radii*

O Teorema 2.5 garante a existência de um ponto fixo do operador T em uma bola centrada numa aproximação para uma solução de $f(x) = 0$. No entanto, a condição suficiente para a garantia de existência desse ponto fixo envolve uma quantidade infinita de testes, já que para $\|Y + Z(r)\|_s < r$ significa $|Y_k + Z_k(r)| < \frac{r}{\omega_k^s}$, $\forall k \geq 0$, o que não pode ser verificado computacionalmente. Uma alternativa para tornar essa análise factível, uma vez que este trabalho se insere num contexto de Análise Numérica Rigorosa, é a definição dos chamados Polinômios *Radii*, seguindo a metodologia de [11]. Nesta seção o símbolo \mathbb{I}_n representa uma matriz quadrada de $n \times n$ com todas as entradas iguais a 1.

Suponha que para um dado M , exista Y_M e $Z_M(r)$ tais que $\forall k \geq M$ tem-se

$$\|(T(\bar{x}) - \bar{x})_k\|_\infty \leq \frac{M^s}{k^s} Y_M, \text{ e } \sup_{b_1, b_2 \in B(r)} |(DT(\bar{x} + b_1) b_2)_k|_\infty \leq \frac{M^s}{k^s} Z_M(r) \quad (2.17)$$

então definimos os Polinômios *Radii* pela Definição (2.2). Como na definição de Y e Z , os Polinômios *Radii* serão definidos como $p_k(r)$, sendo que correspondem a $(p_{k,1}, p_{k,2})(r)$, ou seja, cada $p_k : \mathbb{R} \rightarrow M(n, \mathbb{R})^2$.

Definição 2.2 (Polinômios *Radii*). *Dada as sequências $Y = (Y_0, Y_1, Y_2, \dots, Y_M, \dots)$ e $Z(r) = (Z_0, Z_1, Z_2, \dots, Z_M, \dots)(r)$ Polinômios *Radii* são definidos como sendo um conjunto de $M + 1$ polinômios com variável independente r definidos por:*

$$\begin{cases} p_k(r) = Y_k + Z_k(r) - \frac{r}{\omega_k^s} (\mathbb{I}_n, \mathbb{I}_n), & k = 0, 1, \dots, M - 1 \\ p_M(r) = Y_M + Z_M - \frac{r}{\omega_M^s} \end{cases} \quad (2.18)$$

O Teorema 2.6 fornece uma alternativa que substitui os infinitos testes necessários por M testes e a construção de Y_M e Z_M convenientes de forma que os Y_k e Z_k para $k > M - 1$ ficam definidos de forma adequada a satisfazer automaticamente as condições necessárias estabelecidas pelo Teorema 2.5, garantindo a existência do ponto fixo e conseqüentemente a solução para a equação do problema em $B_{\bar{x}}(r)$ em que r será tão pequeno quanto possível para que todos os M Polinômios *Radii* sejam tais que $p_k(r) <_{cw} 0$, o que justifica a definição 2.18.

Teorema 2.6 (Testes por Polinômios *Radii*). *Considere M e sejam Y e Z tais que Y_k e Z_k satisfazem (2.16) para $k = 0, 1, 2, \dots, M - 1$. Defina os demais Y_k e Z_k para $k \geq M$ como:*

$$Y_k = \frac{M^s}{k^s} Y_M [\mathbb{I}_n, \mathbb{I}_n] \text{ e } Z_k(r) = \frac{M^s}{k^s} Z_M [\mathbb{I}_n, \mathbb{I}_n] \quad (2.19)$$

com Y_M e Z_M satisfazendo as condições dadas por (2.17). Se existir $r > 0$ tal que $p_k(r) <_{cw} 0$ para todo $k = 0, 1, 2, \dots, M$, então existe um único $x^* \in B_{\bar{x}}(r)$ tal que $T(x^*) = x^*$ e conseqüentemente $f(x^*) = 0$

Demonstração. As relações $p_k(r) <_{cw} 0$ implicam que $|Y_k + Z_k(r)|_\infty < \frac{r}{\omega_k^s}$, pois $Y_k \geq_{cw} 0$ e $Z_k \geq_{cw} 0$. Como por hipótese $p_k(r) <_{cw} 0 \forall k = 0, 1, \dots, M - 1$ então de fato temos $|Y_k + Z_k(r)|_\infty < \frac{r}{\omega_k^s}$ para $0 \leq k \leq M - 1$. Para $k \geq M$ a definição de Y_M e Z_M faz com que para $k \geq M$, Y_k e Z_k acima definidos sejam tais que

$$|(T(\bar{x}) - \bar{x})_k| \leq_{cw} Y_k \text{ e } \sup_{b_1, b_2 \in B(r)} |(DT(\bar{x} + b_1) b_2)_k| \leq_{cw} Z_k(r)$$

Além disso,

$$|Y_k + Z_k(r)|_\infty = \frac{M^s}{k^s} (Y_M + Z_M) < \frac{M^s}{k^s} \frac{r}{M^s} = \frac{r}{k^s}, \forall k \geq M$$

Dessa forma temos que $\|Y + Z\|_s < r$, como queríamos. \square

2.6 Cálculo de Y e Z

Para a aplicação do Teorema 2.6 é necessário determinar Y_k e Z_k para $k = 0, 1, 2, \dots, M-1$, além de Y_M e Z_M . Para tanto será necessário relembrar o Lema 2.5. Existem alguns parâmetros a serem definidos. O primeiro deles é o parâmetro K do lema 2.5, que garante que para todo $k \geq K$ o operador Λ_k é invertível e existe C_Λ tal que $\|\Lambda_k^{-1}\|_\infty < \frac{C_\Lambda}{k}$. Devemos então supor que a dimensão do espaço de projeção, m seja suficientemente grande para que Λ_k seja invertível e $\|\Lambda_k^{-1}\|_\infty < \frac{C_\Lambda}{k}$ para $k \geq m$, ou seja, deve-se tomar $m > K$. Uma vez que o Lema 2.5 demonstra como obter K , pode-se definir uma dimensão de projeção com facilidade. O outro parâmetros que também deve ser mais precisamente especificado é M . Escolhendo $M \geq m$, T_k fica bem definido e não depende da aproximação para a inversa de $Df^{(m)}$ considerada, como pode ser observado em (2.14) e (2.13). A aproximação da inversa de $Df^{(m)}$ será importante, no entanto, para construir Y_k para $k = 0, 1, \dots, m-1$. Esta Seção será dividida em duas subseções dedicadas a construir Y e Z .

2.6.1 Construção de Y

Para definir Y é necessário definir $Y_k \forall k \geq 0$, ou seja, definir $Y_k \forall 0 \leq k \leq M-1$ e Y_M . Considere a definição do operador T , $T(x) = x - Jf(x)$, o que implica $T(x) - x = -Jf(x)$. Como queremos que dada uma solução aproximada \bar{x} tenhamos $|(T(\bar{x}) - \bar{x})_k| \leq_{cw} Y_k$, para $k = 0, 1, 2, \dots, m-1$ podemos definir $Y_k = |(J_m f^{(m)}(\bar{x}))_k|$. Recordando que por definição temos $(Jx)_k = \Lambda_k^{-1} Q_k$ para $k \geq m$, define-se Y_k conforme a Equação 2.20. Trata-se de uma definição pertinente, já que $|(T(\bar{x}) - \bar{x})_k| \leq_{cw} Y_k$ como queremos. É importante ressaltar que $Y_k = (Y_{k,1}, Y_{k,2})$ está bem definido por 2.20, já que J_m já foi calculado anteriormente.

$$Y_k = \begin{cases} |(J_m f^{(m)}(\bar{x}))_k|, & k = 0, 1, \dots, m-1 \\ |\Lambda_k^{-1} f_k(\bar{x})|, & k = m, m+1, \dots, M-1 \end{cases} \quad (2.20)$$

A definição de Y_M deve ser tal que $Y_M \frac{M^s}{k^s} > |\Lambda_k^{-1} f_k(\bar{x})|_\infty, \forall k > M$. Pela definição de norma operatorial $|\Lambda_k^{-1} f_k(\bar{x})|_\infty \leq \|\Lambda_k^{-1}\|_\infty |f_k(\bar{x})|_\infty$. Como m é a dimensão do espaço de projeção segue que $\bar{Q}_{k,1} = \bar{Q}_{k,2} = 0, \forall k \geq m$. Uma vez que \bar{x} tem apenas as primeira m componentes, a saber de 0 a $m-1$. Assim, tem-se que f_k é dado por (2.21).

$$f_k(R, \bar{Q}_0, \bar{Q}_{1,1}, \bar{Q}_{1,2}, \dots) = \begin{bmatrix} -k \frac{\pi}{T} \bar{Q}_{k,2} + \bar{Q}_{k,1} R - (A\bar{Q})_{k,1} \\ k \frac{\pi}{T} \bar{Q}_{k,1} + \bar{Q}_{k,2} R - (A\bar{Q})_{k,2} \end{bmatrix} = \begin{bmatrix} -(A\bar{Q})_{k,1} \\ -(A\bar{Q})_{k,2} \end{bmatrix} \quad (2.21)$$

Desenvolvendo $(A\bar{Q})_{k,1}$ e $(A\bar{Q})_{k,2}$ temos que

$$\begin{cases} (A\bar{Q})_{k,1} = A_{k,1} \bar{Q}_0 + \sum_{\beta=1}^{m-1} ((A_{k-\beta,1} + A_{k+\beta,1}) \bar{Q}_{\beta,1} + (A_{k+\beta,2} - A_{k-\beta,2}) \bar{Q}_{\beta,2}) \\ (A\bar{Q})_{k,2} = A_{k,2} \bar{Q}_0 + \sum_{\beta=1}^{m-1} ((A_{k+\beta,2} + A_{k-\beta,2}) \bar{Q}_{\beta,1} + (A_{k-\beta,1} - A_{k+\beta,1}) \bar{Q}_{\beta,2}) \end{cases}$$

Dessa forma,

$$|(A\bar{Q})_{k,1}| \leq_{cw} |A_{k,1} \bar{Q}_0| + \sum_{\beta=1}^{m-1} (|A_{k-\beta,1}| + |A_{k+\beta,1}|) |\bar{Q}_{\beta,1}| + (|A_{k+\beta,2}| + |A_{k-\beta,2}|) |\bar{Q}_{\beta,2}|$$

Usando o fato de que $|A_{k,1} \bar{Q}_0| \leq \frac{\|A\|_{s^*}}{\omega_k^{s^*}} \mathbb{I} \mathbb{I}_n |\bar{Q}_0|$ temos que

$$|(A\bar{Q})_{k,1}| \leq_{cw} \frac{\|A\|_{s^*}}{\omega_k^{s^*}} \mathbb{I} \mathbb{I}_n |\bar{Q}_0| + \sum_{\beta=1}^{m-1} \|A\|_{s^*} \mathbb{I} \mathbb{I}_n \left(\frac{1}{\omega_{k-\beta}^{s^*}} + \frac{1}{\omega_{k+\beta}^{s^*}} \right) |\bar{Q}_{\beta,1}| + \|A\|_{s^*} \mathbb{I} \mathbb{I}_n \left(\frac{1}{\omega_{k+\beta}^{s^*}} + \frac{1}{\omega_{k-\beta}^{s^*}} \right) |\bar{Q}_{\beta,2}|$$

Mas como $\left(\frac{1}{\omega_{k-\beta}^{s^*}} + \frac{1}{\omega_{k+\beta}^{s^*}}\right) \leq \left(\frac{1}{\omega_{k+\beta}^{s^*}} - \frac{1}{\omega_{k-\beta}^{s^*}}\right)$

$$\left|(A\bar{Q})_{k,1}\right| \leq_{cw} \frac{\|A\|_{s^*}}{\omega_k^{s^*}} \mathbb{I} \mathbb{I}_n |\bar{Q}_0| + \sum_{\beta=1}^{m-1} \|A\|_{s^*} \mathbb{I} \mathbb{I}_n \left(\frac{1}{\omega_{k-\beta}^{s^*}} + \frac{1}{\omega_{k+\beta}^{s^*}}\right) (|\bar{Q}_{\beta,1}| + |\bar{Q}_{\beta,2}|)$$

Multiplicando e dividindo por ω_k^s temos que

$$\left|(A\bar{Q})_{k,1}\right| \leq_{cw} \frac{\|A\|_{s^*}}{\omega_k^{s^*}} \frac{\omega_k^s}{\omega_k^s} \mathbb{I} \mathbb{I}_n |\bar{Q}_0| + \sum_{\beta=1}^{m-1} \frac{\|A\|_{s^*}}{\omega_k^s} \mathbb{I} \mathbb{I}_n \omega_k^s \left(\frac{1}{\omega_{k-\beta}^{s^*}} + \frac{1}{\omega_{k+\beta}^{s^*}}\right) (|\bar{Q}_{\beta,1}| + |\bar{Q}_{\beta,2}|)$$

Note que como $2 \leq s \leq s^*$, segue que $\frac{\omega_k^s}{\omega_k^{s^*}} \leq 1$. Além disso,

$$\begin{aligned} \omega_k^s \left(\frac{1}{\omega_{k-\beta}^{s^*}} + \frac{1}{\omega_{k+\beta}^{s^*}}\right) &= k^s \left(\frac{1}{(k-\beta)^{s^*}} + \frac{1}{(k+\beta)^{s^*}}\right) \leq \\ &\leq \frac{k^s}{(k-\beta)^{s^*}} + \frac{k^s}{(k+\beta)^{s^*}} \leq 1 + \frac{k^s}{(k-\beta)^s} \leq 1 + \left(\frac{M}{M-\beta}\right)^s \end{aligned}$$

de onde temos que

$$\left|(A\bar{Q})_{k,1}\right| \leq_{cw} \frac{\|A\|_{s^*}}{\omega_k^s} \mathbb{I} \mathbb{I}_n |\bar{Q}_0| + \sum_{\beta=1}^{m-1} \frac{\|A\|_{s^*}}{\omega_k^s} \mathbb{I} \mathbb{I}_n \omega_k^s \left[1 + \left(\frac{M}{M-\beta}\right)^s\right] (|\bar{Q}_{\beta,1}| + |\bar{Q}_{\beta,2}|)$$

Assim, definindo

$$W = \mathbb{I} \mathbb{I}_n |\bar{Q}_0| + \sum_{\beta=1}^{m-1} \left[1 + \left(\frac{M}{M-\beta}\right)^s\right] \mathbb{I} \mathbb{I}_n (|\bar{Q}_{\beta,1}| + |\bar{Q}_{\beta,2}|)$$

Finalmente, para $k \geq M$ temos que

$$|f_k(\bar{x})|_\infty \leq \frac{1}{\omega_k^s} \|A\|_{s^*} |W|_\infty$$

Além disso, considerando o Lema 2.5, podemos definir Y_M pela Equação 2.22.

$$Y_M = \frac{1}{M^{s+1}} \|A\|_{s^*} C_\Lambda |W|_\infty \quad (2.22)$$

2.6.2 Construção de Z

Recapitulando, a construção de Z consiste em obter $Z_k(r)$, $k \geq 0$ tal que seja válido que

$$\sup_{b_1, b_2 \in B(r)} \left| [DT(\bar{x} + b_1)b_2]_k \right| \leq_{cw} Z_k(r), \quad \forall k \geq 0$$

Como $b_1, b_2 \in B(r)$ será utilizada uma normalização que facilita os cálculos, escrevendo $b_1 = ru$ e $b_2 = rv$ em que $u, v \in B(1)$. Como u e v são seqüências infinitas de matrizes, temos a seguinte notação: $u = [u_0, u_1, \dots, u_k, \dots]$, em que cada $u_k = (u_{k,1}, u_{k,2}) \in M(n, \mathbb{R})^2$, de forma que a notação utilizada para as $Q_{k,1}$ e $Q_{k,2}$ será mantida, ou seja, $(u_{k,1}, u_{k,2}) = u_k$. O fato de que $u \in B(1)$ significa pela definição de $\|\cdot\|_s$ que $|u_{k,1}|, |u_{k,2}| \leq_{cw} \omega_k^{-s}$ para cada k . Como em geral, o fundamental é calcular $\max(|u_{k,1}|_\infty \omega_k^s, |u_{k,2}|_\infty \omega_k^s)$, por vezes $u_{k,1}$ e $u_{k,2}$ serão denotados apenas por u_k .

Como no caso de Y , é suficiente pelo Teorema 2.6 obter Z_k para $k = 0, 1, \dots, M-1$ e Z_M , de forma a caracterizar todos os Z_k conforme apresentado por esse Teorema. Para tanto, considere o operador $J^\dagger : \Omega^{s+1} \rightarrow \Omega^s$ definido como

$$(J^\dagger x)_k = \begin{cases} (Df^{(m)} \cdot x^m)_k & k = 0, \dots, m-1 \\ \Lambda_k x_k, & k \geq m, \end{cases} \quad (2.23)$$

Como $T(\bar{x} + ru) = \bar{x} + ru - Jf(\bar{x} + ru)$ temos que

$$\begin{aligned} DT(\bar{x} + ru)rv &= \lim_{t \rightarrow 0} \frac{[\bar{x} + ru + trv - Jf(\bar{x} + ru + trv)] - [\bar{x} + ru - Jf(\bar{x} + ru)]}{t} \\ &= \lim_{t \rightarrow 0} \frac{trv - J[f(\bar{x} + ru + trv) - f(\bar{x} + ru)]}{t} = rv - JDf(\bar{x} + ru)rv = [I - JDf(\bar{x} + ru)]rv \end{aligned}$$

Assim, considerando a definição de J^\dagger e somando e subtraindo $JJ^\dagger rv$ temos (2.24). É importante notar que as construções de J e J^\dagger são bastante semelhantes. Se J_m , aproximação para a derivada de $Df^{(m)}$ for suficientemente boa, $JJ^\dagger \approx I$. Em 2.24, $DT(\bar{x} + ru)rv$ é separado em duas parcelas, sendo que a primeira reflete a qualidade da aproximação da derivada de $Df^{(m)}$ e influenciará os Polinômios *Radii* e conseqüentemente o valor de r , raio da bola em Ω^s no qual o operador T é uma contração.

$$\begin{aligned} DT(\bar{x} + ru)rv &= [I - JDf(\bar{x} + ru)]rv \\ &= [I - JJ^\dagger]rv - J[Df(\bar{x} + ru) - J^\dagger]rv. \end{aligned} \quad (2.24)$$

Dessa forma, a definição de $Z(r)$ pode ser feita através das estimativas de $|[I - JJ^\dagger]rv|$ e $J[Df(\bar{x} + ru) - J^\dagger]rv$. Assim, defina as cotas intermediárias Z^0, Z^1, Z^2 de modo que sejam válidas as relações dadas pela Equações 2.25 e 2.26. A justificativa para a definição de Z_0, Z_1 e Z_2 é baseada no formato de f . A derivada de f calculada no ponto $\bar{x} + ru$ e aplicada no vetor rv não produz termos multiplicados por r^k com $k \geq 3$, como pode ser visto no Capítulo 3 que traz detalhes de implementação, entre eles o cálculo de $Df^{(m)}$, no qual esse fato fica claro.

$$|[I - JJ^\dagger]rv| \leq_{cw} Z^0 r, \quad \forall v \in B(1), \quad (2.25)$$

$$|[Df(\bar{x} + ru) - J^\dagger]rv| \leq_{cw} Z^1 r + Z^2 r^2, \quad \forall u, v \in B(1). \quad (2.26)$$

Com a Equação 2.24 podemos comparar elemento a elemento das matrizes de valores absolutos para cada componente k , o que implica que

$$\left| [DT(\bar{x} + ru)rv]_k \right| \leq_{cw} \left| [(I - JJ^\dagger)rv]_k \right| + \left| [J(Df(\bar{x} + ru) - J^\dagger)rv]_k \right|. \quad (2.27)$$

De modo que os elementos Z_k , para $k = 0, \dots, M - 1$ podem ser definidos por (2.28).

$$Z_k(r) = \begin{cases} Z_k^0 r + [J_m |(Z^1 r + Z^2 r^2)^m]_k, & k = 0, \dots, m - 1 \\ Z_k^0 r + |\Lambda_k^{-1}| (Z_k^1 r + Z_k^2 r^2), & k = m, \dots, M - 1. \end{cases} \quad (2.28)$$

Além disso, Z_M serão definidos de forma a satisfazer

$$\sup_{b_1, b_2 \in B(r)} |[DT(\bar{x} + b_1) b_2]_k|_\infty \leq \frac{M^s}{k^s} Z_M(r) \quad \text{para } k \geq M$$

Como $|v_k| \leq_{cw} w_k^{-s} \mathbb{I}$, defina Z^0 como

$$(Z^0)_k = \begin{cases} |[I - JJ^\dagger] \{w_j^{-s} \mathbb{I}\}_{j=0}^{m-1}]_k & k = 0, \dots, m - 1 \\ 0, & k \geq m \end{cases} \quad (2.29)$$

de modo que de fato temos a relação desejada.

$$|[I - JJ^\dagger]rv| \leq_{cw} Z^0 r.$$

Em 2.29 mostra de forma mais clara o fato de que r depende da qualidade da aproximação para a inversa de $Df^{(m)}$.

Uma vez apresentadas a cota para Z^0 deve-se obter a cota para Z^1 e Z^2 . Como $[(Df(\bar{x} + ru) - J^\dagger)rv]_k$ deve ser cotado por um polinômio quadrático em r , considere sua expansão como um polinômio quadrático em r .

Isso é possível na medida em que f é composta por produtos de matrizes. Como a derivada de f está avaliada no ponto $x + ru$ e aplicada a rv não podem aparecer termos multiplicados por r^p com $p > 2$.

$$[(Df(\bar{x} + ru) - J^\dagger)rv]_k = \sum_{i=1,2} c_{k,i} r^i. \quad (2.30)$$

Para calcular $c_{k,i}$ é importante lembrar a notação, uma vez que $u_k = (u_{k,1}, u_{k,2})$ para $k = 0, 1, 2, \dots$, inclusive $k = 0$, assim, $u_{0,1}$ tem o papel de R e $u_{0,2}$ tem o papel de Q_0 na notação original. Como a parte imaginária de Q_k é antissimétrica em relação a k , i.e $Q_{k,2} = -Q_{-k,2}$, e definindo $sg(l) = \text{sinal}(l)$, temos que:

$$c_{0,1} = \begin{bmatrix} 2 \sum_{k \geq m} v_k, \\ - \sum_{\substack{l+j=0 \\ |l| \geq m}} (A_{j,1} - sg(l)A_{j,2})v_{|l|} \end{bmatrix}, \quad c_{0,2} = \begin{bmatrix} 0 \\ u_0 v_0 + v_0 u_0 \end{bmatrix}, \quad (2.31)$$

para $k = 1, \dots, m-1$

$$c_{k,1} = - \sum_{\substack{l+j=k \\ |l| \geq m}} \begin{bmatrix} (A_{j,1} - sg(l)A_{j,2})v_{|l|} \\ (A_{j,2} + sg(l)A_{j,1})v_{|l|} \end{bmatrix}, \quad c_{k,2} = \begin{bmatrix} u_k v_0 + v_k u_0 \\ u_k v_0 + v_k u_0 \end{bmatrix}, \quad (2.32)$$

e para $k \geq m$

$$c_{k,1} = - \sum_{\substack{l+j=k \\ |l| \neq k}} \begin{bmatrix} (A_{j,1} - sg(l)A_{j,2})v_{|l|} \\ (A_{j,2} + sg(l)A_{j,1})v_{|l|} \end{bmatrix}, \quad c_{k,2} = \begin{bmatrix} u_k v_0 + v_k u_0 \\ u_k v_0 + v_k u_0 \end{bmatrix}. \quad (2.33)$$

Mas Z^1, Z^2 devem ser definidas de modo que $Z_k^1 \geq_{cw} |c_{k,1}|$ e $Z_k^2 \geq_{cw} |c_{k,2}|$. Mas como u e v tem norma unitária segue que $|u_k|, |v_k| \leq_{cw} w_k^{-s} \mathbb{I}_n$ e $|\pm Re(\mathcal{A}_j) \pm Im(\mathcal{A}_j)| \leq_{cw} |Re(\mathcal{A}_j)| + |Im(\mathcal{A}_j)|$. Note no entanto que $\mathbb{I}_n \mathbb{I}_n = n \mathbb{I}_n$, de forma que podemos definir Z^1 e Z^2 da seguinte forma

$$\begin{aligned} |c_{0,2}| \leq_{cw} 2n \begin{bmatrix} 0 \\ \mathbb{I}_n \end{bmatrix} &= Z_0^2, \\ |c_{k,2}| \leq_{cw} 2n w_k^{-s} \begin{bmatrix} \mathbb{I}_n \\ \mathbb{I}_n \end{bmatrix} &= Z_k^2, \quad k \geq 1. \end{aligned} \quad (2.34)$$

No entanto o argumento não está completo em termos numéricos, uma vez que $|c_{k,1}|$, está definido em termos de séries, que não podem ser calculadas no computador. Uma solução para esse problema é introduzir uma nova cota, o que é feito a seguir.

$$L_k > \max\{k, m\} \quad (2.35)$$

Como as séries foram interrompidas em uma quantidade finita de termos é necessário incluir as matrizes denotadas por H_k de modo que

$$\begin{aligned} |c_{0,1}| \leq_{cw} \left[\begin{array}{c} 2 \sum_{j=m}^{L_0} w_j^{-s} \mathbb{I}_n, \\ \sum_{\substack{l+j=0 \\ m \leq |l| \leq L_0}} (|Re(\mathcal{A}_j)| + |Im(\mathcal{A}_j)|) w_l^{-s} \mathbb{I}_n \end{array} \right] + H_0 &=: Z_0^1 \\ |c_{k,1}| \leq_{cw} \sum_{\substack{l+j=k \\ m \leq |l| \neq k, |l| \leq L_k}} \left[\begin{array}{c} (|Re(\mathcal{A}_j)| + |Im(\mathcal{A}_j)|) w_l^{-s} \mathbb{I}_n \\ (|Im(\mathcal{A}_j)| + |Re(\mathcal{A}_j)|) w_l^{-s} \mathbb{I}_n \end{array} \right] + H_k &=: Z_k^1, \quad k = 1, \dots, m-1 \end{aligned} \quad (2.36)$$

e analogamente para $k \geq m$. A definição de Z^1 é composta agora por uma soma finita que pode ser calculada e uma parte que não pode ser calculada no computador mas pode ser estimada analiticamente usando a teoria de séries. Defina

$$\zeta(M, s) = \frac{1}{(M+1)^s} + \frac{1}{(M+2)^s} + \frac{1}{s-1} \frac{1}{(M+2)^{s-1}},$$

e

$$H_0 = \begin{bmatrix} 2\zeta(L_0, s)\mathbb{I}_n \\ h_0\mathbb{I}_n \end{bmatrix}, \quad H_k = h_k \begin{bmatrix} \mathbb{I}_n \\ \mathbb{I}_n \end{bmatrix}, \quad (2.37)$$

em que para cada $k \geq 0$ temos

$$h_k = \frac{\sqrt{2}n\|\mathcal{A}\|_{s^*}}{(L_k + 1 - k)^{s^* - s}} \left(\zeta(L_k - k, 2s) + \zeta(L_k, 2s) \right).$$

Lema 2.6. *A fórmula 2.36 vale para H_0, H_k definida em 2.37.*

Demonstração. Note que para todo $M \geq 1$ e $s \geq 2$

$$\sum_{k=M+1}^{\infty} \frac{1}{k^s} < \zeta(M, s). \quad (2.38)$$

Devido ao fato de que $\sum_{k=M+1}^{\infty} \frac{1}{k^s} = \frac{1}{(M+1)^s} + \frac{1}{(M+2)^s} + \sum_{k=M+3}^{\infty} \frac{1}{k^s} < \frac{1}{(M+1)^s} + \frac{1}{(M+2)^s} + \int_{M+2}^{\infty} k^{-s} dk$. Portanto,

$$\left| 2 \sum_{j=L_0+1}^{\infty} w_j^{-s} \mathbb{I}_n \right| \leq_{cw} 2\zeta(L_0, s) \mathbb{I}_n.$$

Para o restante dos termos note que $(|\operatorname{Re}(\mathcal{A}_j)| + |\operatorname{Im}(\mathcal{A}_j)|) \leq_{cw} \sqrt{2}|\mathcal{A}_j| \leq_{cw} \sqrt{2} \frac{\|\mathcal{A}\|_{s^*}}{w_j^{s^*}} \mathbb{I}_n$, então, para todo $k \geq 0$, a cauda da serie pode ser limitada por

$$\begin{aligned} \left| \sum_{\substack{l+j=k \\ |l| \geq L_k+1}} (|\operatorname{Re}(\mathcal{A}_j)| + |\operatorname{Im}(\mathcal{A}_j)|) w_l^{-s} \mathbb{I}_n \right| &\leq_{cw} \sqrt{2} \|\mathcal{A}\|_{s^*} \sum_{l=L_k+1}^{\infty} \left(\frac{1}{w_{k-l}^{s^*}} + \frac{1}{w_{k+l}^{s^*}} \right) \mathbb{I}_n w_l^{-s} \mathbb{I}_n \\ &\leq_{cw} \frac{\sqrt{2}n\|\mathcal{A}\|_{s^*}}{(L_k + 1 - k)^{s^* - s}} \sum_{l=L_k+1}^{\infty} \left(\frac{1}{w_{l-k}^s} + \frac{1}{w_{k+l}^s} \right) w_l^{-s} \mathbb{I}_n. \end{aligned}$$

Na última passagem usamos o fato de que $L_k > k$, $s^* \geq s$ e a relação $\mathbb{I}_n \mathbb{I}_n = n \mathbb{I}_n$. O resultado segue aplicando 2.38 uma vez que a última série pode ser reescrita como

$$\begin{aligned} \sum_{l=L_k+1}^{\infty} \left(\frac{1}{w_{l-k}^s} + \frac{1}{w_{k+l}^s} \right) w_l^{-s} &= \sum_{l=L_k+1}^{\infty} \left(\frac{1}{w_{k+l}^s} \right) w_l^{-s} + \sum_{l=L_k-k+1}^{\infty} \left(\frac{1}{w_{k+l}^s} \right) w_l^{-s} \\ &\leq \sum_{l=L_k+1}^{\infty} w_l^{-2s} + \sum_{l=L_k-k+1}^{\infty} w_l^{-2s}. \end{aligned}$$

□

De [11], temos que

$$\sum_{\substack{\alpha+\beta=k \\ |\alpha| \neq k}} \frac{1}{w_\alpha^s w_\beta^s} \leq \frac{1}{w_k^s} \left[2 + 2 \sum_{l=1}^M \frac{1}{l^s} + \frac{2}{M^{s-1}(s-1)} + \eta_M - 1 - \frac{1}{w_{2k}^s} \right],$$

onde

$$\eta_k = 2 \left[\frac{k}{k-1} \right]^s + \left[\frac{4 \log(k-2)}{k} + \frac{\pi^2 - 6}{3} \right] \left[\frac{2}{k} + \frac{1}{2} \right]^{s-2}$$

Então para $k \geq M$ temos que

$$\begin{aligned}
|c_{k,1}|_\infty &\leq \sqrt{2}\|\mathcal{A}\|_{s^*} \sum_{\substack{l+j=k \\ |l| \neq k}} w_j^{-s} w_l^{-s} \\
&\leq \frac{\sqrt{2}\|\mathcal{A}\|_{s^*}}{w_k^s} \left[2 + 2 \sum_{l=1}^M \frac{1}{l^s} + \frac{2}{M^{s-1}(s-1)} + \eta_M - 1 \right] =: \frac{\sqrt{2}\|\mathcal{A}\|_{s^*}}{w_k^s} C_1, \\
|c_{k,2}|_\infty &\leq \frac{2n}{w_k^s}.
\end{aligned} \tag{2.39}$$

Como para $k \geq M$ o primeiro termo do lado direito da Equação 2.27 é nulo, vale que

$$\begin{aligned}
\left| [DT(\bar{x} + ru)rv]_k \right|_\infty &\leq_{cw} \left| [J(Df(\bar{x} + ru) - J^\dagger)rv]_k \right|_\infty \\
&\leq \|\Lambda_k^{-1}\|_\infty (|c_{k,1}|_\infty r + |c_{k,2}|_\infty r^2).
\end{aligned} \tag{2.40}$$

Assim, usando $\|\Lambda_k^{-1}\|_\infty \leq \frac{C_\Lambda}{k}$; $\forall k \geq m$ e que $k \geq M$, tem-se que $\|\Lambda_k^{-1}\|_\infty \leq \frac{C_\Lambda}{M}$, e pode-se definir Z_M como

$$Z_M = \frac{C_\Lambda}{M^{s+1}} (\sqrt{2}\|\mathcal{A}\|_{s^*} C_1 r + 2nr^2).$$

Capítulo 3

Detalhes de Implementação

O método apresentado no Capítulo 2 é composto basicamente por duas etapas: Obter uma solução aproximada $\bar{x} \in X$ e calcular $r > 0$ tal que exista $x^* \in B_{(\bar{x}, 0^\infty)}(r)$ com $f(x^*) = 0$. Neste Capítulo são apresentados de forma algorítmica todos os passos que compõem a técnica numérica rigorosa e os detalhes de sua implementação realizada com o software MATLAB[®].

3.1 O Algoritmo para Cálculo da Forma Normal de Floquet

Seja uma função $t \mapsto A(t)$ periódica de período τ . O método numérico descrito no Capítulo 2 retorna como resultado um número natural m , um ponto $\bar{x} = (\bar{R}, \bar{Q}_0, \bar{Q}_{1,1}, \bar{Q}_{1,2}, \bar{Q}_{2,1}, \bar{Q}_{2,2}, \dots) \in [M(n, \mathbb{R}) \times M(n, \mathbb{R})]^m$ e um número real $r > 0$ tal que existe um ponto $x^* \in B_{(\bar{x}, 0^\infty)}(r)$, $x^* = (R, Q_0, Q_{1,1}, Q_{1,2}, \dots)$ tal que

$$X(t) = \left[Q_0 + \sum_{k=-\infty}^{\infty} (Q_{k,1} + jQ_{k,2}) \right] e^{tR}$$

é uma matriz fundamental da equação diferencial $x' = A(t)x$, com $X(0) = I$.

O algoritmo para o cálculo da Série de Fourier da aplicação A será omitido por se tratar de um procedimento padrão e oferecido pelos pacotes computacionais utilizados neste trabalho, a saber Maple[®] 14 e MATLAB[®] versão 2008. Dessa forma, assume-se que a Série de Fourier de A já é conhecida. A técnica apresentada é composta essencialmente de duas grandes etapas: cálculo de \bar{x} e aplicação dos Polinômios *Radii* para determinação de r . A Figura 3.1 apresenta os principais passos da estratégia adotada para a solução do problema, desde a reformulação do problema original como um problema de obtenção de ponto fixo para um operador definido num espaço de dimensão infinita até a utilização da técnica de Polinômios *Radii*.

Para o cálculo de J_m no Algoritmo 1 utilizou-se o software Maple[®] 14, capaz de calcular analiticamente a inversa de $Df^{(m)}$. Deve-se observar, no entanto, que é necessário determinar o valor de m , caracterizando o espaço de projeção X^m . Ocorre que m deve ser tal que $m > K$, determinado após o cálculo de Λ_k , conforme descrito no Algoritmo 2. Mas, para o cálculo de K é necessário \bar{x} . Assim, é preciso implementar um processo iterativo de cálculo de soluções aproximadas para diferentes valores de m até que $m > K$. O cálculo dos Polinômios *Radii* o cálculo de Y_k , $k = 0, 1, 2, \dots, M-1$, Z_k , $k = 0, 1, 2, \dots, M-1$, Y_M e Z_M . Para determinar essas cotas é necessário uma série de cálculos preliminares, como apresentado no Algoritmo 2. Um resumo dos cálculos necessários para determinar Y , Z e conseqüentemente os Polinômios *Radii* é apresentado nas Seções 3.3.1 e 3.3.2. O processo iterativo $x_{k+1} = x_k - J_m f^{(m)}(x_k)$ do Algoritmo 1 requer um ponto inicial x_0 .

Algoritmo 1: Algoritmo para cálculo de $\bar{x} \in X^m$ tal que $f^{(m)}(\bar{x}) \approx 0$

Entrada:

$x_0 \leftarrow$ Ponto Inicial (Utilizando Algoritmo 3)

$\{A_k\}_{k \geq 0} \leftarrow$ Série de Fourier de $A(t)$

$m \leftarrow$ Dimensão do espaço de projeção

início

$A^m \leftarrow$ Aproximação para Inversa de $Df^{(m)}(x_0)$

$i \leftarrow 0$

enquanto *Condição de Parada não for satisfeita* **faça**

$x_{i+1} = x_i - J_m f^{(m)}(x_i)$

$i \leftarrow i + 1$

fim

fim

Saída:

$\bar{x} \leftarrow x_{i-1}$

Algoritmo 2: Algoritmo para cálculo da Forma Normal de Floquet

Entrada:

$\{A_k\}_{k \geq 0} \leftarrow$ Série de Fourier de $A(t)$

início

$s^* \leftarrow$ Determina s^* tal que $\|A\|_{s^*} < \infty$

$s \leftarrow$ Escolha de $2 \leq s \leq s^*$

$\|A\|_{s^*} \leftarrow$ Determina $\|A\|_{s^*}$ ou uma cota superior

$x_0 \leftarrow$ Ponto inicial $(R, Q_0, Q_{1,1}, Q_{1,2}, \dots, Q_{m-1,1}, Q_{m-1,2})$

$m \leftarrow$ Inicializa dimensão do espaço de projeção X^m

$\bar{x} \leftarrow$ **Calcula solução o aproximada tal que** $f^{(m)}(\bar{x}) \approx 0$

$K \leftarrow K \in \mathbb{N}$; $\Lambda_k = \frac{\partial f_k}{\partial Q_k}$ é diagonal secundária dominante $\forall k \geq K$

$C_\Lambda \leftarrow C_\Lambda$; tal que $\|\Lambda_k^{-1}\|_\infty \leq \frac{C_\Lambda}{k}$; $\forall k \geq K$

enquanto $K \geq m$ **faça**

$m \leftarrow m + 1$

$x_0 \leftarrow$ Ponto inicial $(R, Q_0, Q_{1,1}, Q_{1,2}, \dots, Q_{m-1,1}, Q_{m-1,2})$

$m \leftarrow$ Inicializa dimensão do espaço de projeção X^m

$\bar{x} \leftarrow$ **Calcula solução o aproximada tal que** $f^{(m)}(\bar{x}) \approx 0$

$K \leftarrow K \in \mathbb{N}$; $\Lambda_k = \frac{\partial f_k}{\partial Q_k}$ é diagonal secundária dominante $\forall k \geq K$

$C_\Lambda \leftarrow C_\Lambda$; tal que $\|\Lambda_k^{-1}\|_\infty \leq \frac{C_\Lambda}{k}$; $\forall k \geq K$

fim

$M \leftarrow$ Escolha de $M > m$

$Y_k \leftarrow$ Cálculo de Y_k para $k = 0, 1, \dots, M - 1$

$Y_M \leftarrow$ Cálculo de Y_M

$Z_k \leftarrow$ Cálculo de Z_k para $k = 0, 1, \dots, M - 1$

$Z_M \leftarrow$ Cálculo de Z_M

$\{p_k(r)\}_{0 \leq k \leq M} \leftarrow$ Polinomios Rádii usando Y_k, Z_k, Y_M e Z_M

$r \leftarrow$ Resultado de um algoritmo de busca para determinar r tal que $p_k(r) < 0 \forall 0 \leq k \leq M$

fim

Saída:

$\{\bar{x}, r\} \leftarrow$ **Retorna resultado do problema**

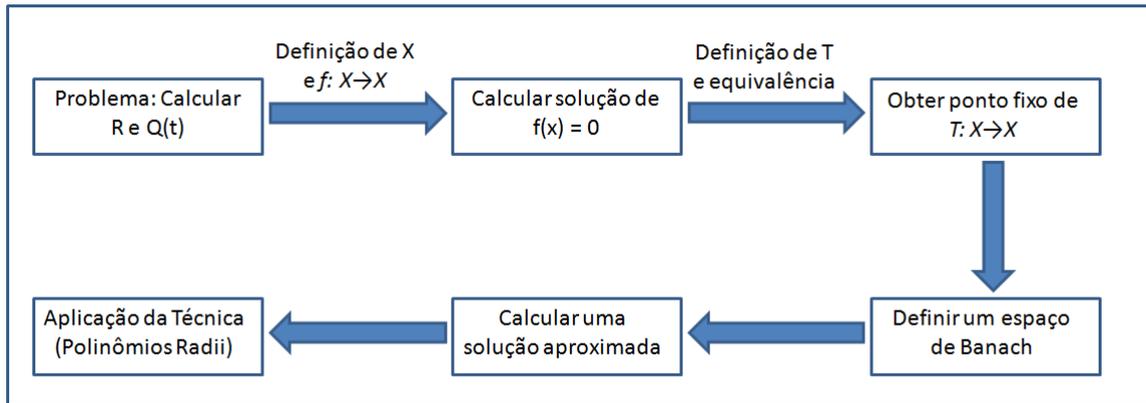


Figura 3.1: Resumo dos principais passos da estratégia de solução adotada

3.2 Cálculo do Ponto Inicial

Conforme apresentado no Algoritmo 1, a solução aproximada é obtida através de sucessivas iterações $x_{k+1} = x_k - J_m f^{(m)}(x_k)$. Dessa forma, é necessário determinar um ponto inicial x_0 . Para determinar um ponto x_0 adequado é importante lembrar que calcular a Forma Normal de Floquet consiste originalmente em determinar $t \mapsto Q(t) \in M(n, \mathbb{R})$ 2τ periódica e $R \in M(n, \mathbb{R})$ tais que $X(t) = Q(t)e^{tR}$ seja solução de $x' = A(t)x$ e $X(0) = I$, o que significa obter $Q(t)$ τ -periódica e R tais que

$$\begin{cases} Q'(t) = A(t)Q(t) - Q(t)R \\ Q(0) = I \end{cases} \quad (3.1)$$

3.2.1 Cálculo de $R_{inicial}$

Suponha que obtido $Y(t) \in M(n, \mathbb{R})$ solução da equação matricial $X' = A(t)X$ com $X(0) = I$. Assim, segue que $Y(2\tau) = Q(2\tau)e^{2\tau R} = e^{2\tau R}$. Dessa forma, a matriz R da Forma Normal de Floquet é dada pela Equação 3.2. A existência do logaritmo é garantida pelo fato de que $Y(2\tau)$ é não singular, o que implica como consequência da Proposição A.1 do Apêndice A, que $Y(2\tau)$ tem logaritmo em $M(n, \mathbb{C})$. Mais que isso, $Y(2\tau)^2$ tem logaritmo em $M(n, \mathbb{R})$. Dessa forma, como $Y(2\tau)^2 = (e^{2\tau R})^2 = e^{4\tau R}$, temos que

$$R = \frac{1}{4\tau} \log \left(Y(2\tau)^2 \right) \quad (3.2)$$

Embora $Y(2\tau)$ só possa ser determinado caso a equação tenha sido resolvida, é possível obter uma estimativa para $Y(2\tau)$ usando métodos numéricos clássicos para solução de equações diferenciais. O método escolhido para determinação de $Y(2\tau)$ foi o método de Runge-Kutta de quarta ordem. Mais detalhes da aplicação desse método para aproximação de $Y(2\tau)$ são apresentados no Apêndice D. Este método está implementado em MATLAB[®] e pode ser usado através da função `ode45`. O valor máximo do passo de integração utilizado foi $\Delta t = \frac{2\tau}{10000}$. Em [29] p.270 há uma comparação de desempenho de métodos numéricos de solução de equações diferenciais, justificando a popularidade do Método Runge-Kutta.

Como se trata de uma aproximação, não é possível garantir a existência de $\log(X(2\tau))$ em $M(n, \mathbb{R})$, sendo a existência do logaritmo em $M(n, \mathbb{C})$ garantida pelo Lema 3.1, que corresponde a uma parte do Teorema 1.27 de [14], cuja demonstração não será apresentada neste texto. A referência [14] contém ainda a descrição de métodos de cálculo do logaritmo de matrizes.

Lema 3.1. (Existência do Logaritmo Complexo) Toda matriz $G \in M(n, \mathbb{C})$ invertível possui um logaritmo, i.e. $G = e^E$ com $E \in M(n, \mathbb{C})$.

Por continuidade, se $X(2\tau)$ é uma boa aproximação de $Y(2\tau)$ então a parte complexa de $\frac{1}{4\tau} \log(X(2\tau)^2)$ deve ser pequena. Dessa forma, para a construção do ponto inicial do algoritmo será considerada apenas a parte real de $\frac{1}{4\tau} \log(X(2\tau)^2)$. Desse modo, o cálculo de R do ponto inicial será feita pela Equação

$$R_{inicial} = \text{Re} \left[\frac{1}{4\tau} \log \left(X(2\tau)^2 \right) \right] \quad (3.3)$$

3.2.2 Cálculo de $Q_{inicial}$

Usando a Equação 3.1 com $R = R_{inicial}$ calculado na anterior, a obtenção de $Q_{inicial}$ será feita como através de cálculo aproximado de $Q(t)$ solução da EDO 3.1 pelo método de Runge-Kutta de quarta ordem. Como $Q(t)$ é periódica de período 2τ será calculada uma aproximação para de $Q(t)$ usando $Q(t_k)$ para $t_0 = 0, t_1 = \Delta t, t_2 = 2\Delta t, \dots, t_k = k\Delta t$ e uma interpolação linear, produzindo $\tilde{Q}(t)$ conforme apresentado na Figura 3.2. A aproximação para a série de Fourier de $Q(t)$ que será utilizada na construção do ponto inicial é a série de Fourier de $\tilde{Q}(t)$. Como $Q(t) \in M(n, \mathbb{R})$, a Figura 3.2 é meramente ilustrativa, sendo a interpolação linear realizada em cada uma $(Q(t))_{ij}, i, j = 1, 2, \dots, n$. Os pontos $(t_k, Q(t_k))$ serão calculados através do método de Runge-

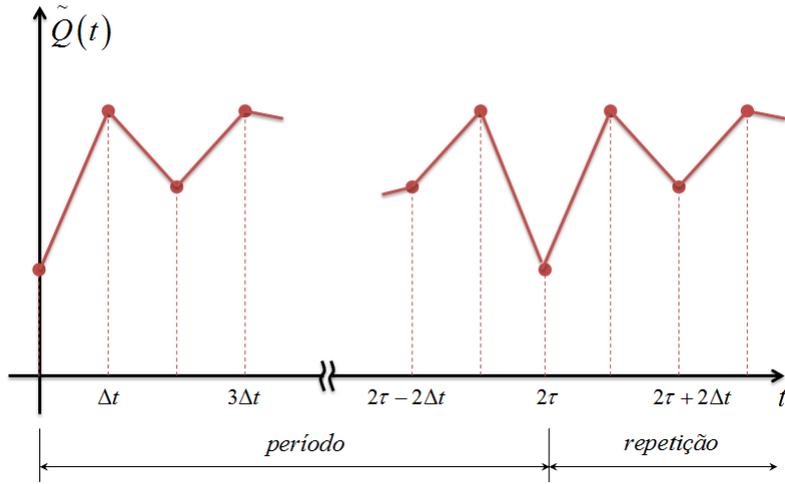


Figura 3.2: Aproximação de $Q(t)$ usando interpolação linear, $\tilde{Q}(t)$

Kutta de 4ª ordem, como para obter $X(2\tau)$, que segue a mesma ideia descrita na Seção 3.2.1, sendo utilizada a matriz R , aproximação calculada pelo procedimento daquela seção. O cálculo da série de Fourier de $Q(t)$ foi realizado utilizando a interpolação linear apresentada na Figura 3.2. Como em cada intervalo $[t_0, t_0 + \Delta t]$ $\tilde{Q}(t)$ é considerada linear da forma $a(t - t_0) + b$ temos que

$$\begin{aligned} I_k(t_0, \Delta t) &= \int_{t_0}^{t_0 + \Delta t} (a(t - t_0) + b) e^{-ik\frac{\pi}{\tau}t} dt = a \int_{t_0}^{t_0 + \Delta t} t e^{-ik\frac{\pi}{\tau}t} dt + (b - at_0) \int_{t_0}^{t_0 + \Delta t} e^{-ik\frac{\pi}{\tau}t} dt \\ &= a \left\{ \left[-\frac{\tau t}{k\pi} e^{-ik\frac{\pi}{\tau}t} \right]_{t_0}^{t_0 + \Delta t} + \int_{t_0}^{t_0 + \Delta t} \frac{\tau}{k\pi} e^{-ik\frac{\pi}{\tau}t} dt \right\} - \frac{\tau(b - at_0)}{k\pi} \left[e^{-ik\frac{\pi}{\tau}t} \right]_{t_0}^{t_0 + \Delta t} \\ &= e^{-ik\frac{\pi}{\tau}t_0} \left[\frac{\tau^2}{k^2\pi^2} + \frac{\tau b}{k\pi} \right] + e^{-ik\frac{\pi}{\tau}(t_0 + \Delta t)} \left[-a \frac{\tau(\Delta t)}{k\pi} - \frac{\tau(b)}{k\pi} - \frac{\tau^2}{k^2\pi^2} e^{-ik\frac{\pi}{\tau}(t_0 + \Delta t)} \right] \\ &= \frac{\tau}{k\pi} e^{-ik\frac{\pi}{\tau}t_0} \left[\frac{\tau}{k\pi} + b \right] + \frac{\tau}{k\pi} e^{-ik\frac{\pi}{\tau}(t_0 + \Delta t)} \left[-a(\Delta t) - b - \frac{\tau}{k\pi} \right] \end{aligned}$$

Assim, podemos calcular a Série de Fourier de $\tilde{Q}(t)$ utilizando a Equação 3.4 e conseqüentemente obter o ponto inicial para o algoritmo de cálculo da Forma Normal de Floquet.

$$\begin{cases} \tilde{Q}_{k,1} = \frac{1}{2\tau} \operatorname{Re} \left[\sum_{j=0}^{N-1} I_k \left(j \frac{2\tau}{N}, \frac{2\tau}{N} \right) \right] \\ \tilde{Q}_{k,2} = \frac{1}{2\tau} \operatorname{Im} \left[\sum_{j=0}^{N-1} I_k \left(j \frac{2\tau}{N}, \frac{2\tau}{N} \right) \right] \end{cases} \quad (3.4)$$

Uma vez que o ponto calculado utilizando a metodologia descrita nesta seção é apenas a primeira aproximação para a aplicação do esquema iterativo $x_{k+1} = x_k - J_m f^{(m)}(x_k)$, não se justifica a utilização de métodos mais rigorosos de controle de erro, o que tenderia a aumentar significativamente o algoritmo construído.

Algoritmo 3: Algoritmo para cálculo do Ponto Inicial

Entrada:

$\{A_k\}_{k \geq 0} \leftarrow$ Série de Fourier de $A(t)$

$m \leftarrow$ Inicializa dimensão do espaço de projeção X^m

$N \leftarrow$ Número de intervalos da partição de $[0, 2\tau]$

início

$\Delta t = \frac{2\tau}{N}$

$X(2\tau) \leftarrow$ Resultado da solução numérica de $\begin{cases} X' = A(t)X \\ X(0) = I \end{cases}$ (Runge - Kutta)

fim

$R_{inicial} \leftarrow \frac{1}{4\tau} \operatorname{Re} \left\{ \log \left(X(2\tau)^2 \right) \right\}$

$\{Q_{inicial}(k\Delta t)\}_{0 \leq k \leq N} \leftarrow$ Resultado da solução numérica de $\begin{cases} Q' = A(t)Q - QR_{inicial} \\ Q(0) = I \end{cases}$ (Runge - Kutta)

$\{Q_0, Q_1, \dots, Q_{m-1}\}_{inicial} \leftarrow$ Série de Fourier de $Q(t)$ calculado usando $Q_{inicial}(k\Delta t)_{0 \leq k \leq N}$

Saída: $R_{inicial}; (Q_0, Q_1, \dots, Q_{m-1})_{inicial}$

3.3 Cálculo dos Polinômios *Radii*

O Cálculo dos Polinômios *Radii* é realizado através da determinação de Cotas Y_k e Z_k . Nesta seção serão apresentados quais os cálculos necessários para a determinação das Cotas do Capítulo 2 e conseqüentemente, dos Polinômios *Radii*.

3.3.1 Cálculo de Y

A Equação 3.5, define Y_k para $k = 0, 1, \dots, M-1$. Dada a solução aproximada, \bar{x} , essa expressão utiliza J_m , já calculado para a obtenção de \bar{x} e Λ_k^{-1} , que depende apenas de uma quantidade finita de termos da série de Fourier de A que é uma das entradas do algoritmo.

$$Y_k = \begin{cases} \left| \left(J_m f^{(m)}(\bar{x}) \right)_k \right| & k = 0, 1, \dots, m-1 \\ \left| \Lambda_k^{-1} f_k(\bar{x}) \right| & k = m, m+1, \dots, M-1 \end{cases} \quad (3.5)$$

Para calcular Y_M utiliza-se a Equação 3.6. W depende apenas de \bar{x} , da dimensão de projeção m e de M escolhido conforme apresentado no Algoritmo 2. Além disso, C_Λ é calculado conforme indicado no Lema 2.5. Além disso, é suposto calculado $\|A\|_{s^*}$, ou uma cota superior, de modo que a definição de Y_M também está completa.

$$\begin{cases} Y_M = \frac{1}{M^{s+1}} \|A\|_{s^*} C_\Lambda |W|_\infty \\ W = \mathbb{I}_n |\bar{Q}_0| + \sum_{\alpha=1}^{m-1} \left(1 + \left(\frac{M}{M-\alpha} \right)^s \right) \mathbb{I}_n (|\bar{Q}_{i,1}| + |\bar{Q}_{i,2}|) \end{cases} \quad (3.6)$$

3.3.2 Cálculo de Z

Analogamente ao cálculo de Y , para determinar Z é necessário calcular Y_k para $k = 0, 1, 2, \dots, M-1$ e Y_M . Como a Equação 3.7 define Z_k para $k = 0, 1, \dots, M-1$ e já argumentamos sobre o cálculo de J_m e Λ_k^{-1} , resta analisar se o método está de fato completo em termos do cálculo de Z^0, Z^1 e Z^2 . O problema de definições em termos de séries foi resolvido com a introdução de H_k na Equação 2.36, que define Z^0 utilizando apenas uma quantidade finita de termos da série de Fourier de $A(t)$ e de H_0, H_k , definidos pela Equação 2.37. Assim, Z^1 pode ser calculado sem problemas. A Equação 2.34 define Z^2 utilizando apenas os valores de n e s .

$$Z_k(r) = \begin{cases} Z_k^0 r + \left[|J_m|(Z^1 r + Z^2 r^2) \right]_k, & k = 0, \dots, m-1 \\ Z_k^0 r + |\Lambda_k^{-1}|(Z_k^1 r + Z_k^2 r^2), & k = m, \dots, M-1. \end{cases} \quad (3.7)$$

A Equação 2.29 define Z^0 também utilizando uma quantidade finita de termos não nulos de A e A , ambos definidos por termos disponíveis para cálculo. Dessa forma, o cálculo de Z^0, Z^1 e Z^2 está bem definido em termos computacionais. Para calcular Z_M , utiliza-se a Equação 3.8, introduzida no Capítulo 2.

$$Z_M = \frac{C_\Lambda}{M^{s+1}} (\sqrt{2} \|A\|_{s^*} C_1 r + 2nr^2) \quad (3.8)$$

O termo C_1 presente na definição de Z_M está definido através da Equação 2.39, em cujo termo C_1 é dado por

$$\left[2 + 2 \sum_{l=1}^M \frac{1}{l^s} + \frac{2}{M^{s-1}(s-1)} + \eta_M - 1 \right]$$

sendo que η_M é dado por

$$\eta_M = 2 \left[\frac{M}{M-1} \right]^s + \left[\frac{4 \log(M-2)}{M} + \frac{\pi^2 - 6}{3} \right] \left[\frac{2}{M} + \frac{1}{2} \right]^{s-2}.$$

Assim, o cálculo de Z_M e portanto de Z está bem definido em termos computacionais. Esse argumento finaliza a análise de consistência da técnica numérica em termos de factibilidade dos cálculos a serem realizados.

3.4 Aritmética de Intervalos

Nesta seção será realizada uma breve introdução à aritmética de intervalos e como ela será aplicada para compor a técnica numérica rigorosa demonstrada no Capítulo 2 e discutida na Seção 3.1. O livro [18] é a referência utilizada neste trabalho. O Apêndice E apresenta uma introdução aos principais conceitos da Aritmética de Intervalos ou Aritmética Intervalar.

Uma vez que neste trabalho é estudado um método numérico rigoroso, é natural que alguns cálculos sejam realizados utilizando a aritmética de intervalos. A palavra alguns na última frase se justifica pelo fato de que os cálculos da solução aproximada \bar{x} não serão realizados utilizando a aritmética de intervalos. Pode-se destacar

pelo menos três razões para esta escolha: (i) Haveria um aumento de tempo computacional, uma vez que todas as operações usariam aritmética de intervalos, (ii) A utilização de aritmética de intervalos no processo iterativo $x_{k+1} = x_k - J_m f^{(m)}(x_k)$ por si só faz com que o comprimento dos intervalos considerados aumente quando k cresce, (iii) Um dos objetivos do método é fornecer uma solução \bar{x} de fácil aplicação, além de uma vizinhança na qual é garantida a presença do problema original, não matrizes de intervalos. No entanto, os cálculos utilizados para a determinação dos Polinômios *Radii* e consequentemente do valor de $r > 0$ tal que $T : B_{\bar{x}}(r) \rightarrow B_{\bar{x}}(r)$ é uma contração devem ser feitos utilizando a aritmética de intervalos, de modo que erros de truncamento não invalidem o resultado numérico rigoroso obtido. Assim, o próprio valor de r encontrado será um intervalo. Como o cálculo da solução aproximada no espaço de dimensão finita é uma aproximação e a etapa que envolve a demonstração do Capítulo 2 consiste do cálculo de $r > 0$, justifica-se a utilização obrigatória de aritmética de intervalos apenas nessa segunda etapa.

A Aritmética de Intervalos é aplicável em áreas como Análise Numérica, Engenharia e Projeto Assistido por computador, como introduzido por [18]. Várias conjecturas tem sido provadas na atualidade com o auxílio de Aritmética de Intervalos. O pacote computacional utilizado neste trabalho é o INTerval LABoratory[®], INTLAB[®] V6. Trata-se de uma toolbox livre para MATLAB[®] desenvolvida pelo *Institute for Reliable Computing - Hamburg University of Technology*. Uma descrição detalhada do pacote, suas funcionalidades e forma de utilização pode ser encontrada em [21].

Existem ainda alguns métodos eficientes de cálculo de soluções de equações $f(x) = 0$ usando aritmética de intervalos. Um desses métodos é baseado em intervalos encaixantes. Seja uma solução inicial $x^0 = (x_1^0, x_2^0, \dots, x_m^0)$ em que cada x_i^0 é uma matriz cujas entradas são intervalos. Dessa forma, aplicando o operador Krawczyk obtém-se $y^0 = (y_1^0, y_2^0, \dots, y_m^0)$ em que y_i^0 é uma matriz de intervalos. Assim, define-se cada x_i^1 como $x_i^1 = x_i^0 \cap y_i^1$, ou seja a matriz formada pela intersecção dos intervalos correspondentes de x_i^0 e y_i^0 . Em [18] são apresentados algoritmos referentes a essas técnicas de solução. Uma vantagem do método apresentado neste trabalho é que não há necessidade de definição de um intervalo inicial no qual devem estar as soluções.

3.5 Resultados experimentais

O método numérico apresentado baseia-se na ideia de calcular uma solução aproximada \bar{x} para a equação $f^{(m)}(x)$, e $r > 0$ tal que exista uma solução exata da Equação $f(x) = 0$ em $B_{\bar{x}}(r) \subset \Omega^s$. Uma importante questão é o fato de que $T : B_{\bar{x}}(r) \rightarrow B_{\bar{x}}(r)$ é uma contração. Sendo uma contração, ao realizar a iteração $x_{k+1} = T(x_k)$, obter-se-ia a solução exata do problema seja qual for o ponto inicial $x_0 \in B_{\bar{x}}(r)$. No entanto, não é possível implementar um processo iterativo com o operador T computacionalmente, apenas com sua aproximação, o operador $T^{(m)} : X^m \rightarrow X^m$. Assim, o resultado é tão melhor quanto menor for o valor de r , uma vez que isso significa que existe uma solução do problema original em $B_{\bar{x}}(r)$.

Nesta seção serão apresentados resultados experimentais das duas etapas do algoritmo, a saber, o cálculo da solução aproximada e a determinação de r usando o argumento baseado nos Polinômios *Radii*. A Equação 3.9 será utilizada no Capítulo 4 para a aplicação da técnica. A Figura 3.3 apresenta a evolução de $\|f(x_k)^m\|_s$, em que x_k^m são as iteradas do operador T^m definido por $T^m(x_k^m) = x_k^m - J_m f^m(x_k^m)$ com $s = 2$. Para a obtenção da Figura da esquerda foi utilizado o procedimento descrito na Seção 3.2 para obter x_0^m . Em seguida é utilizado o esquema $x_{k+1}^m = x_k^m - J_m f^m(x_k^m)$, em que $J_m \approx Df^{(m)}(x_0^m)$. Já no gráfico da esquerda foi utilizado um ponto inicial aleatório, de forma que houve divergência do algoritmo como pode ser observado.

$$A(t) = \begin{bmatrix} 0 & 1 \\ -\frac{13}{2} & -\frac{1}{4} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -\frac{9}{2} & 0 \end{bmatrix} \cos(2t) + \begin{bmatrix} 0 & 0 \\ -6 & 0 \end{bmatrix} \sin(2t) \quad (3.9)$$

Pela Figura 3.3 verifica-se uma rápida convergência, de forma que a norma-s de $f(\bar{x})$ atinge valores da ordem

de 10^{-15} , sendo a solução inicial x_0 obtida utilizando o Método de Runge-Kutta é tal que $f(\bar{x})$ é da ordem de 10^{-3} . Deve-se ressaltar que a solução aproximada foi obtida sem aritmética de intervalos, que será usada apenas na etapa rigorosa através do cálculo dos Polinômios *Radii*. Para a validação do método é preciso verificar se de

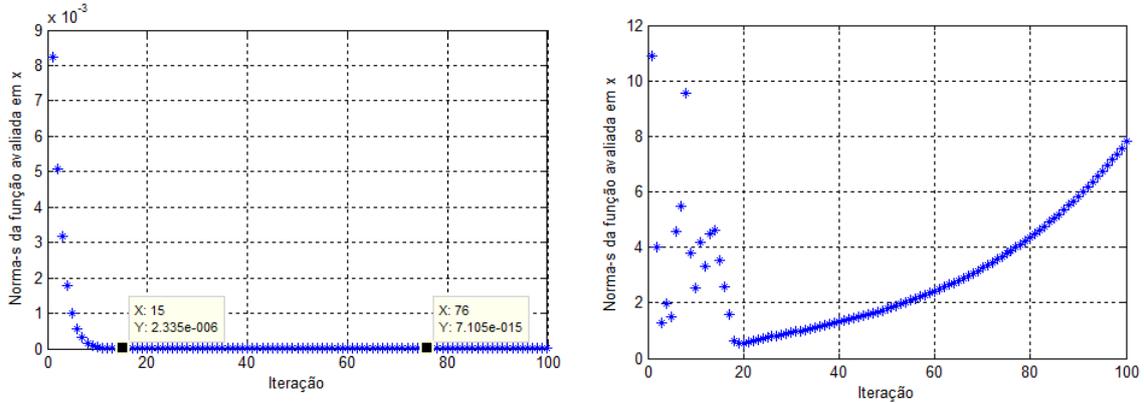


Figura 3.3: Resultados de convergência e não convergência para $A(t)$ da Equação 3.9

fato $X(t) = Q(t)e^{tR}$ é solução de $x' = A(t)x$. Para isso foi feita derivação numérica usando diferenças finitas e calculado Ψ pela Equação 3.10. Para uma solução exata do problema, $\Psi(t) = 0 \forall t \in \mathbb{R}$.

$$\Psi(t) = Q'(t) + Q(t)R - A(t)Q(t) \quad (3.10)$$

A Figura 3.4 apresenta o gráfico de $|\Psi(t)|_\infty$. Verifica-se que o $|\Psi(t)|_\infty$ é da ordem de 10^{-5} , pequeno se comparado com o $|A(t)|_\infty$, indicando que a solução aproximada possa ser aparentemente boa. Para uma análise mais precisa da qualidade da solução é necessário obter o valor de r tal que seja garantida a existência de uma solução exata em $B_{\bar{x}}(t)\Omega^s$, em que \bar{x} denota a solução aproximada completada de 0^∞ . Como a série de Fourier de $A(t)$ em

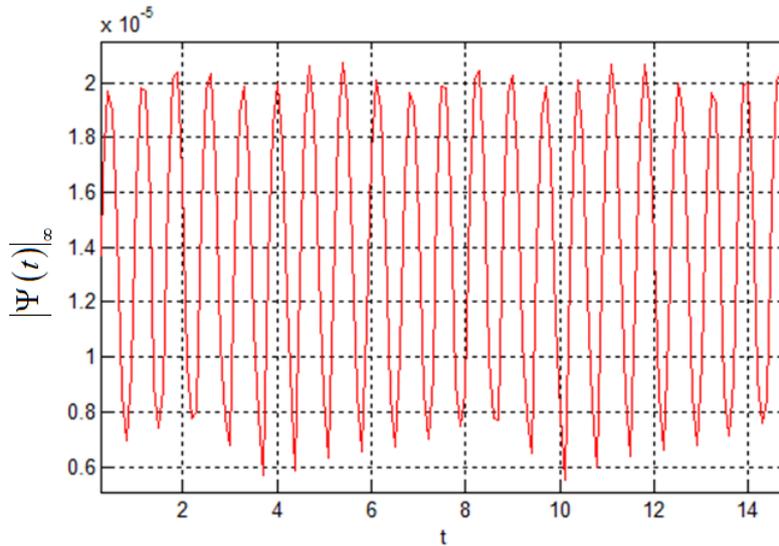


Figura 3.4: Gráfico de $|\Psi(t)|_\infty$ definido pela Equação 3.10

(3.9) tem uma quantidade finita de termos não-nulos, é possível realizar a etapa numérica rigorosa com o valor exato de $\|A\|_{s^*}$. No entanto, vale reforçar que para obter uma cota segura para r é necessário utilizar uma cota superior para $\|A\|_{s^*}$. A Tabela 3.1 apresenta os resultados da etapa rigorosa para equações $x' = A(t)x$ com $A(t)$ dado pela Equação 3.9, além dos problemas com $A(t)$ dados pela Equação 3.11. Como r é um intervalo,

em virtude da utilização da aritmética de intervalos, o resultado que interessa é o supremo do intervalo.

$$B(t) = \begin{bmatrix} 1 & 0 \\ 0 & \cos(t) \end{bmatrix} \quad C = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ \sum_{\substack{k=-\infty \\ k \neq 0}}^{\infty} \left(\frac{1}{k^2} e^{ikt}\right) & 0 & 1 \end{bmatrix} \quad (3.11)$$

Em termos práticos, o resultado mostrado na Tabela 3.1 garante a existência de uma solução exata do problema

#	A(t)	B(t)	C(t)
s^*	3	2	2
s	2	2	2
$\ A\ _{s^*}$	24	1	1
K	10	8	21
m	12	10	23
M	19	16	27
$\sup(C_\Lambda)$	6.836018990062298	9.096314936625827	1.929777070398553
$\sup(r)$	6.323592462254095e-008	9.133758561390195e-010	2.784982188670484e-007

Tabela 3.1: Resultados da execução do método numérico rigoroso para $A(t)$, $B(t)$ e $C(t)$ dadas pelas Equações 3.9 e 3.11

em $B_{\bar{x}}(r)$. Deve-se ressaltar, porém, que foram utilizadas cotas para a construção do algoritmo, de forma que é possível que existam cotas melhores, capazes de fornecer valores menores de r , garantindo assim melhores aproximações. Os valores de K tais que Λ_k é invertível e $\|\Lambda_k^{-1}\|_\infty < \frac{C_\Lambda}{k} \forall k \geq K$ são apresentados, bem como $m > K$ escolhido sempre igual a $K + 2$. É importante lembrar que em todos os casos foi possível obter um valor exato para $\|A\|_{s^*}$ e que, não sendo possível obtê-lo, uma cota superior é suficiente para os cálculos rigorosos. Além disso, foram escolhidos para teste funções cuja série de Fourier já fosse conhecida para facilitar a execução e evitar novos cálculos.

Capítulo 4

Aplicação da Forma Normal de Floquet no Controle Sistemas Dinâmicos Periódicos

Este Capítulo tem por objetivo aplicar a Forma Normal de Floquet ao controle de sistemas dinâmicos periódicos. Não é a primeira vez que essa técnica é utilizada, sendo os trabalhos [23], [24], [19], [25] e [8] relacionados ao mesmo assunto. No entanto, esses trabalhos utilizam a metodologia de [22] para a determinação da Forma Normal de Floquet. Neste trabalho é utilizada a técnica numérica rigorosa abordada nos Capítulos 2 e 3.

4.1 Controle Linear do Oscilador de Duffing

A Equação de Duffing, (4.1), modela um oscilador largamente utilizada na teoria de Sistemas Não-Lineares. A dinâmica do oscilador depende dos parâmetros ξ , α , β , γ e $\omega \in \mathbb{R}$ que tem interpretações físicas específicas tais como a elasticidade cúbica com coeficiente α que descreve a dureza em mecânica dos sólidos, conforme descrito em [13] e o termo ξ , que representa o amortecimento linear do oscilador. A equação é forçada pelo termo $\gamma \cos(\omega t)$. O coeficiente β controla a amplitude da força de restauração.

$$x'' + \xi x' - \beta x + \alpha x^3 = \gamma \cos(\omega t) \quad (4.1)$$

Seja $y : \mathbb{R} \rightarrow \mathbb{R}^2$ um caminho C^1 , τ -periódico fixado, em que $\omega = \frac{2\pi}{\tau}$. O objetivo desta seção é determinar uma aplicação $u : \mathbb{R} \times \mathbb{R}^2 \rightarrow \mathbb{R}$ para que dada uma solução $x(t)$ da Equação 4.2 tenhamos $\lim_{t \rightarrow \infty} (x(t) - y(t)) = 0$. Trata-se de um problema clássico de controle de sistemas dinâmicos, que será resolvido utilizando uma aproximação linear e a Forma Normal de Floquet.

$$x'' = -\xi x' + \beta x - \alpha x^3 + \gamma \cos(\omega t) + u(t, x, x') \quad (4.2)$$

Definindo a aplicação $f(t, x, x') = -\xi x' + \beta x - \alpha x^3 + \gamma \cos(\omega t)$ e u da forma $u(t, x', x'') = u_f(t) + u_L(t, x, x')$ com $u_f(t) = y''(t) - f(t, y(t), y'(t))$. Assim, vamos calcular u_L de modo que $\lim_{t \rightarrow \infty} (x(t) - y(t)) = 0$. Substituindo $u = u_f + u_L$ em (4.2),

$$x'' = f(t, x, x') + y'' - f(t, y, y') + u_L(t, x', x'') \quad (4.3)$$

Rearranjando os termos a equação fica

$$x'' - y'' = f(t, x, x') - f(t, y, y') + u_L(t, x', x'') \quad (4.4)$$

Definindo $e(t) = x(t) - y(t) \forall t \in \mathbb{R}$ temos que $x(t) = e(t) + y(t)$ Portanto

$$e'' = f(t, e + y(t), e' + y'(t)) - f(t, e + y(t), e' + y'(t)) + u_L(t, e + y(t), e' + y'(t))$$

Como f é diferenciável, temos

$$f(t, e + y(t), e' + y'(t)) - f(t, e + y(t), e' + y'(t)) = Df(t, y(t), y'(t))(e, e') + o(\|(e, e')\|)$$

com $\lim_{t \rightarrow \infty} \frac{o(\|(e, e')\|)}{\|(e, e')\|} = 0$. Definindo $\bar{u}_L(t, e, e') = u_L(t, e + y(t), e' + y'(t))$

$$e'' = Df(t, y(t), y'(t))(e, e') + o(\|(e, e')\|) + \bar{u}_L(t, e, e')$$

O termo $o(\|(e, e')\|)$ será desprezado sem maiores preocupações teóricas. Embora seja um procedimento bastante utilizado nas ciências aplicadas, trata-se de um argumento matematicamente incompleto. No entanto, esta seção tem por objetivo apresentar um exemplo de aplicação da Teoria de Floquet e não um trabalho rigoroso de controle de sistemas dinâmicos. Para simplificar a notação a variável e será utilizada mesmo após a simplificação. Assim, temos que

$$e'' = Df(t, y(t), y'(t))(e, e') + \bar{u}_L(t, e, e')$$

Dessa forma, definindo $e_1 = e$ e $e_2 = e'$ e notando que $Df(t, y(t), y'(t))(e_1, e_2) = \partial_2 f(t, y(t), y'(t))e_1 + \partial_3 f(t, y(t), y'(t))e_2$,

$$\underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{e'} = \underbrace{\begin{bmatrix} 0 & 1 \\ \partial_2 f(t, y(t), y'(t)) & \partial_3 f(t, y(t), y'(t)) \end{bmatrix}}_{A(t)} \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{\epsilon} + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_B \bar{u}_L(t, e_1, e_2) \quad (4.5)$$

O problema consiste em obter \bar{u}_L tal que $\lim_{t \rightarrow \infty} e(t) = 0$. No entanto, vamos nos restringir ao caso em que $\bar{u}_L(t, \epsilon) = F(t)\epsilon$ em que $F: \mathbb{R} \rightarrow M(1, 2, \mathbb{R})$ é 2τ -periódica. Assim, para determinar \bar{u}_L basta determinar F . Dessa forma, a Equação 4.5 pode ser reescrita como

$$e' = A(t)\epsilon + BF(t)\epsilon \quad (4.6)$$

com $A(t+\tau) = A(t) \forall t \in \mathbb{R}$. Seja $X(t) = Q(t)e^{tR}$ Forma Normal de Floquet da equação $x' = A(t)x$ e a mudança de coordenadas dependente de t dada por $\epsilon = Q(t)z$. Assim, substituindo na Equação 4.6

$$Q'(t)z + Q(t)z' = A(t)Q(t)z + BF(t)Q(t)z$$

Mas como $Q(t)e^{tR}$ é matriz fundamental de $x' = A(t)x$ temos que $Q'(t) + Q(t)R = A(t)Q(t)$, ou seja, $Q'(t) = A(t)Q(t) - Q(t)R$. Substituindo $Q'(t)$ na equação anterior, fica

$$Q(t)z' = Q(t)Rz + BF(t)Q(t)z$$

pré-multiplicando ambos os lados da equação anterior por $Q(t)^{-1}$ temos que a Equação 4.6 pode ser reescrita na variável z como

$$z' = Rz + Q(t)^{-1}BF(t)Q(t)z \quad (4.7)$$

Seja $K \in M(2, \mathbb{R})$ uma matriz arbitrária tal que todos os seus autovalores tem parte real negativa. Somando e subtraindo Kz do lado esquerdo da Equação 4.7 e reorganizando os termos fica

$$z' = \left\{ K + \underbrace{\left[Q(t)^{-1}BF(t)Q(t) - (K - R) \right]}_{\rho(t, F, K)} \right\} z$$

Deve-se observar que $K + \rho(t, F, K)$ é uma matriz 2τ -periódica. Dessa forma é suficiente que os multiplicadores de Floquet de $K + \rho(t, F, K)$ tenham módulo menor que 1 para que $\lim_{t \rightarrow \infty} z(t) = 0$ e consequentemente $\lim_{t \rightarrow \infty} \epsilon(t) = 0$. Como os multiplicadores de Floquet dependem continuamente da matriz, conforme enunciado da Proposição 1.5, se $\rho(t, F, K)$ for uma perturbação de norma pequena, os multiplicadores de Floquet de $K + \rho(t, F, K)$ também terão módulo menor que 1. Então, fixado K , para cada $t \in \mathbb{R}$, $F(t)$ é definido como sendo a matriz 1×2 que minimiza $\left\| Q(t)^{-1} BxQ(t) - (K - R) \right\|_{OP}$, ou seja,

$$F(t) = \arg \min_{x \in M(1,2,\mathbb{R})} \left\| Q(t)^{-1} BxQ(t) - (K - R) \right\|_{OP} \quad (4.8)$$

Deve-se destacar, no entanto, que esse processo não garante que $\lim_{t \rightarrow \infty} z(t) = 0$, uma vez que trata-se de um argumento de minimização, sendo que o ideal seria garantir que $\left\| Q(t)^{-1} BxQ(t) - (K - R) \right\|_{OP} = 0$. Assim, após o cálculo de F devem ser calculados os multiplicadores de Floquet de $z' = (K + \rho(t, F, K))z$, de forma a garantir que todos tenham parte real negativa para $F(t)$ calculado. Além disso, como não é possível garantir que $\rho(t, F, K) = 0$, também não será possível que a convergência das soluções de $z' = (K + \rho(t, F, K))z$ à origem sejam iguais à da equação $z' = Kz$. Para cada matriz K fixada será calculada uma aplicação F diferente. Não é possível com esse argumento garantir a existência de K tal que $F(t)$ calculado pela Equação 4.8 faça com que a equação 4.6 seja estável. O processo de escolha do par $(K, F(t))$ é comumente denominado na literatura técnica como *Sintonia do Sistema de Controle*.

O cálculo de $F(t)$ usando a Equação 4.8 é bastante simples e pode ser feito usando a pseudoinversa. A proposição 4.1 garante a existência da Pseudoinversa de Moore-Penrose de uma matriz. A demonstração desse fato e do Lema 4.1 pode ser obtida em [2].

Proposição 4.1 (Pseudoinversa de Moore-Penrose). *Seja $P \in M(n, m, \mathbb{C})$. Então, existe uma única matriz $P^+ \in M(m, n, \mathbb{C})$ tal que*

$$(i) \quad PP^+P = P$$

$$(ii) \quad P^+PP^+ = P^+$$

$$(iii) \quad (PP^+)^T = PP^+$$

$$(iv) \quad (P^+P)^T = P^+P$$

P^+ é chamada Pseudoinversa de Moore-Penrose de P .

Lema 4.1. *Sejam $P, B \in M(n, m, \mathbb{R})$, e $X = P^+B$. Então, para todo $Y \in \mathbb{R}^m$ vale que $\|PX - B\|_{OP} \leq \|PY - B\|_{OP}$.*

Dessa forma, podemos calcular $F(t)$ por

$$F(t) = \arg \min \left\| Q(t)^{-1} BxQ(t) - (K - R) \right\|_{OP} = \left[Q(t)^{-1} B \right]^+ (K - R) Q(t)^{-1} \quad (4.9)$$

Como exemplo prático será considerada a equação de Duffing com parâmetros $\xi = \frac{1}{4}$, $\beta = \alpha = \gamma = \omega = 1$. A trajetória desejada é $y(t) = 2 \cos(t) + \sin(t)$. Como $y(t) = 2 \cos(t) + \sin(t)$ então $y'(t) = -2\sin(t) + \cos(t)$. Calculando a matriz que representa a derivada de g com respeito à segunda variável na base canônica e substituindo $y(t) = 2 \cos(t) + \sin(t)$ temos que $A(t)$ é dado por (4.10).

$$A(t) = \begin{bmatrix} 0 & 1 \\ 1 - 3(2 \cos(t) + \sin(t))^2 & -0,25 \end{bmatrix} \quad (4.10)$$

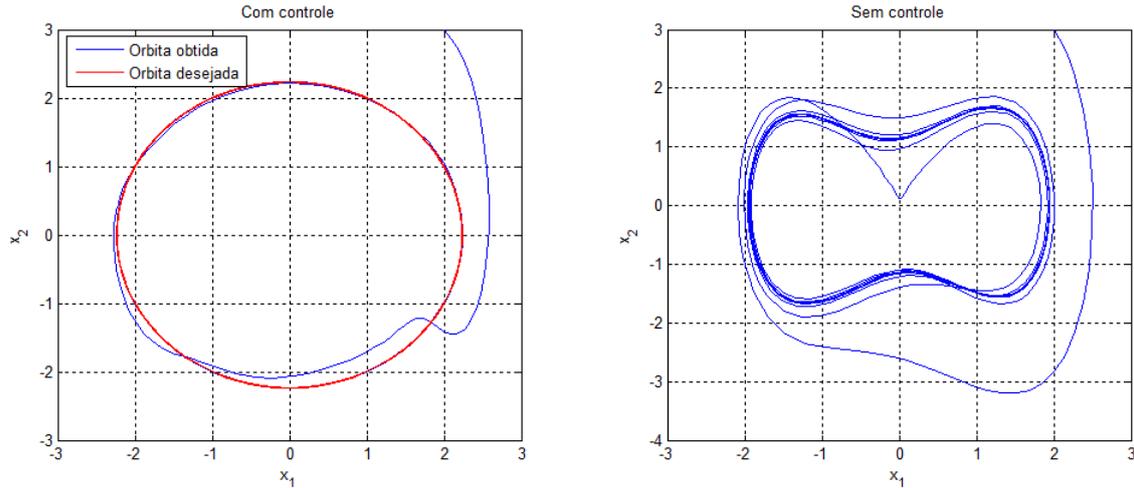


Figura 4.1: Solução da Equação de Duffing com parâmetros $\xi = \frac{1}{4}$, $\beta = \alpha = \gamma = \omega = 1$ com controle para $(x_1(0), x_2(0)) = (2, 3)$ e sem controle ($u = 0$).

Usando identidades trigonométricas básicas é fácil ver que

$$1 - 3(2 \cos(t) + \sin(t))^2 = -\frac{13}{2} - \frac{9}{2} \cos(2t) - 6 \sin(2t) \quad (4.11)$$

Assim temos que (4.12)

$$\epsilon'(t) = \begin{bmatrix} 0 & 1 \\ -\frac{13}{2} - \frac{9}{2} \cos(2t) - 6 \sin(2t) & -0,25 \end{bmatrix} \epsilon(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \bar{u}_L(t, \epsilon) \quad (4.12)$$

A Figura 4.1 apresenta trajetórias de sistema de Duffing com e sem ação de controle. Ambas as trajetórias foram obtidos com o ponto $x(0) = x_0 = (2, 3)$. Para obter os resultados desta figura, foi utilizado o controle $u = u_f + u_t$, em que u_f foi calculado com os parâmetros exatos do oscilador de Duffing. A matriz R dada por (4.13) foi obtida usando o método numérico do Capítulo 2 e será utilizada no cálculo de $u(t)$.

$$R = \begin{bmatrix} -0,205661 & -0,972286 \\ -2,088110 & -0,108474 \end{bmatrix} \quad (4.13)$$

Além disso, é necessário escolher a matriz K . Para que $\lim_{t \rightarrow \infty} \epsilon(t) = 0$ escolheu-se K dada por (4.14), que tem valores característicos $\lambda_1 = -2$ e $\lambda_2 = -3$.

$$K = \begin{bmatrix} -3,2408 & 0,5490 \\ -0,5442 & -1,7592 \end{bmatrix} \quad (4.14)$$

A Figura 4.2 apresenta a ação de controle quando esse é implementado conhecendo-se exatamente os parâmetros. Observa-se que para t grande a ação sobre x_1 aproxima-se de zero, enquanto a ação sobre $x_2 = x'(t)$ faz com que a órbita seja mantida. Isso ocorre porque a órbita escolhida não é uma órbita do sistema sem controle. Caso a órbita escolhida fosse uma órbita do sistema sem o controle a ação de controle teria baixa amplitude, com o objetivo de fazer com que eventuais desvios fossem corrigidos. A Figura 4.3 apresenta a órbita obtida $x(t)$ e a órbita desejada $y(t)$ para o Oscilador Duffing controlado. Observa-se que o erro tende a zero à medida que t cresce, conforme desejado. Em termos práticos é usual estimar os parâmetros de um sistema dinâmico, não sendo conhecidos os parâmetros exatos do sistema e por vezes nem mesmo a dinâmica do mesmo. Assim, utiliza-se o conjunto de parâmetros estimados para a construção da estratégia de controle.

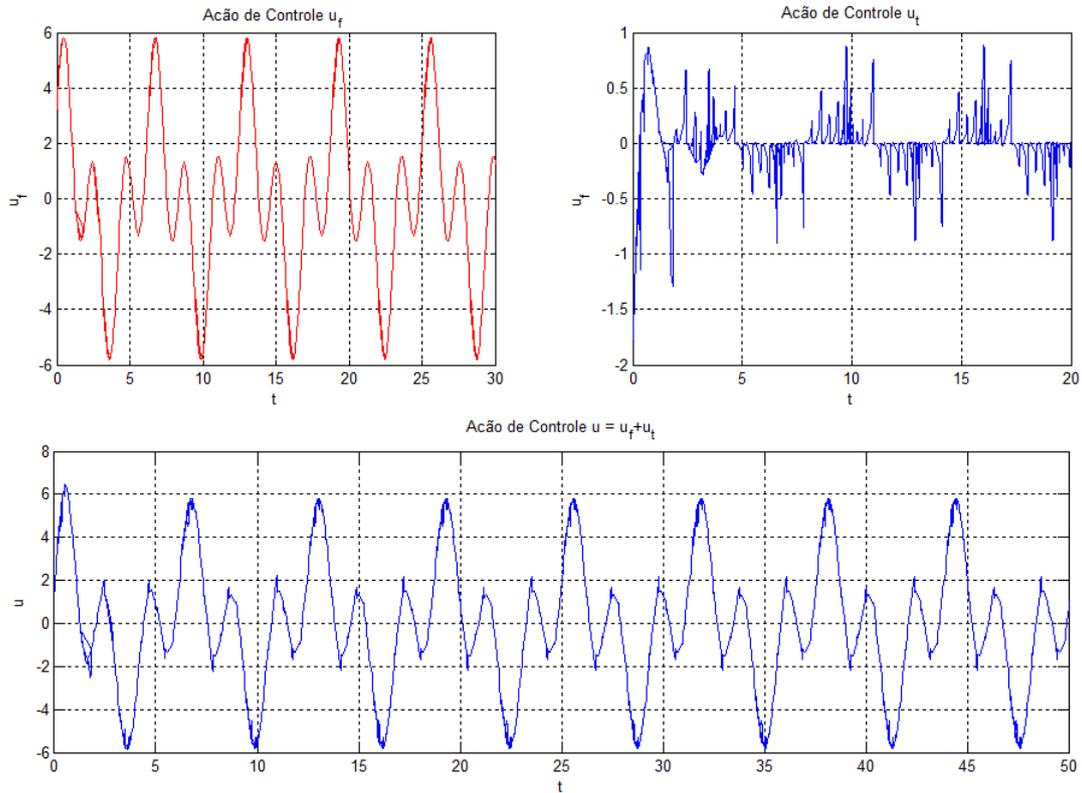


Figura 4.2: Ação de controle para o Oscilador de Duffing para controle implementado conhecendo-se exatamente os parâmetros

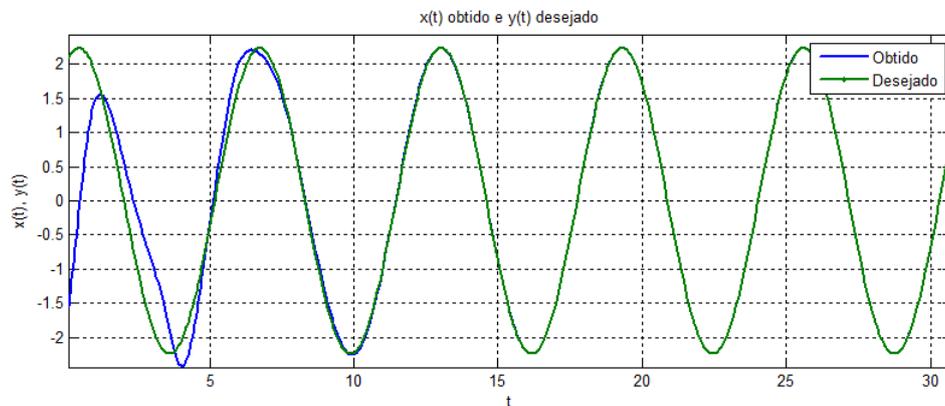


Figura 4.3: Gráficos de $x(t)$ e $y(t)$ para o Oscilador Duffing com $x(0) = -1,5$ e $x'(0) = 2$

4.2 Controle Linear do Pêndulo Forçado Periodicamente

A Figura 4.4 apresenta um esquema do pêndulo forçado verticalmente de forma periódica. Trata-se de uma massa m acoplada a uma haste de massa nula e comprimento l . A extremidade oposta da haste move-se periodicamente com $d(t) = H \cos(\omega t)$. Usando o conceito de Lagrangeano da Mecânica Clássica e o desenvolvimento apresentado em [19], obtém-se a equação (4.15), que rege o comportamento do pêndulo forçado periodicamente. Uma análise detalhada do pêndulo forçado é apresentada em [1].

$$\frac{d^2\theta}{dt^2} + \frac{2}{l} \frac{d\theta}{dt} + \frac{H\omega^2}{l} \cos(\omega t) \sin(\theta) + \frac{g}{l} \sin(\theta) = 0 \quad (4.15)$$

Definindo $p = \frac{H\omega^2}{l}$, $\beta = \frac{2}{l}$ e $g = l$ obtém-se a Equação 4.16 que será utilizada nesta seção.

$$\theta'' + \beta\theta' + [p \cos(\omega t) + 1] \sin(\theta) = 0 \quad (4.16)$$

Como no caso do Oscilador de Duffing, o objetivo é adicionar termo de controle $u : \mathbb{R} \times \mathbb{R}^2 \rightarrow \mathbb{R}$ à Equação 4.16, dando origem à Equação 4.17. Deve-se notar que o termo de controle $u(t, \theta, \theta')$ age sobre a segunda derivada da posição angular. Em termos físicos, u age sobre a aceleração do pêndulo, na direção do movimento.

$$\theta'' = -\beta\theta' - [p \cos(\omega t) + 1] \sin(\theta) + u(t, \theta, \theta') \quad (4.17)$$

Os parâmetros utilizados nesta aplicação são $\beta = 0, 1$, $p = 2$ e $\omega = 1$. A órbita desejada neste caso será

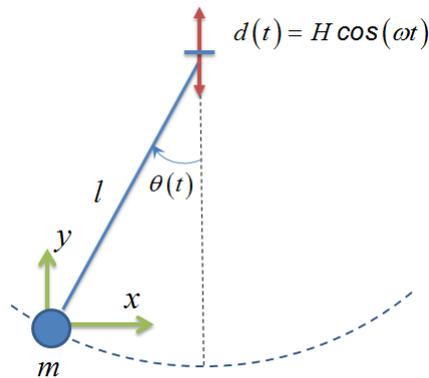


Figura 4.4: Diagrama esquemático de um pêndulo forçado verticalmente

$y(t) = 2 \cos(t)$. Seguindo o mesmo desenvolvimento feito para o Oscilador de Duffing, define-se $x_1(t) = \theta(t)$, $x_2(t) = \theta'(t)$, $y_1(t) = y(t) = 2 \cos(t)$, $y_2(t) = y'(t) = -2 \sin(t)$ e $(e_1, e_2) = (x_1 - y_1, x_2 - y_2)$, a matriz $A(t)$ dada por (4.18).

$$A(t) = \begin{bmatrix} 0 & 1 \\ -(1 + 2 \cos(t)) \cos(2 \cos(t)) & -0, 1 \end{bmatrix} \quad (4.18)$$

Dessa forma, o problema se resume a calcular \bar{u}_L utilizando a metodologia apresentada para o Oscilador Duffing, ou seja calcular $F(t)$ pela Equação 4.9, já que $\bar{u}_L(t, \epsilon) = -F(t)\epsilon$.

$$\epsilon' = \begin{bmatrix} 0 & 1 \\ -(1 + 2 \cos(t)) \cos(2 \cos(t)) & -0, 1 \end{bmatrix} \epsilon + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \bar{u}_L(t, \epsilon)$$

A Figura 4.6 apresenta a solução para o pêndulo controlado e sem controle ($u = 0$). Os valores característicos de K neste caso foram -10 e -12 . A Figura 4.5 apresenta o resultado do controle do pêndulo forçado com implementação realizada utilizando valores exatos para os parâmetros, com K possuindo valores característicos em -5 e -6 . As Figuras 4.7 apresenta as ações de controle u_f , u_L e $u = u_f + u_L$ na situação descrita. Como o objetivo desta seção é apenas apresentar uma possibilidade de aplicação do método demonstrado no Capítulo 2 não serão apresentados mais estudos referentes ao controle dos sistemas dinâmicos aqui abordados. Em um trabalho que fosse feito no campo de Controle, haveria algumas análises adicionais tais como: Rejeição de perturbações, Comportamento no caso de erro de parâmetros (sistema não precisamente conhecido), considerando particularmente perturbações na periodicidade de $Q(t)$. Além disso, uma análise relevante seria a comparação com outras técnicas de controle como PID, por exemplo.

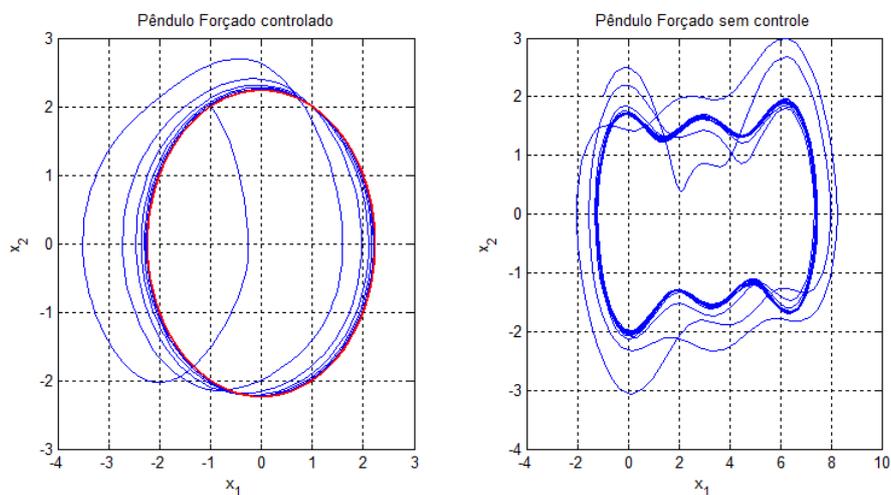


Figura 4.5: Órbita desejada e órbita obtida com o controle implementado para pólos em -5 e -6 para parâmetros $\beta = 0, 1$, $p = 2$ e $\omega = 1$.

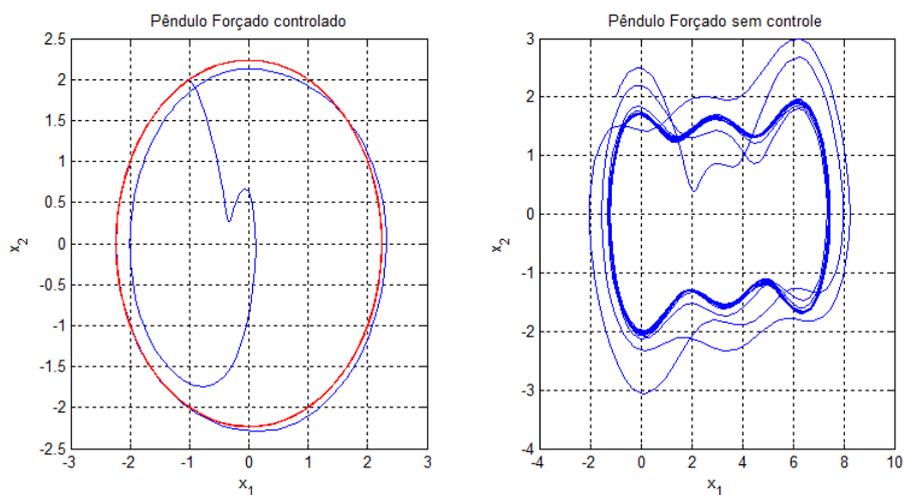


Figura 4.6: Órbita desejada e órbita obtida com o controle implementado para valores característicos em -10 e -12

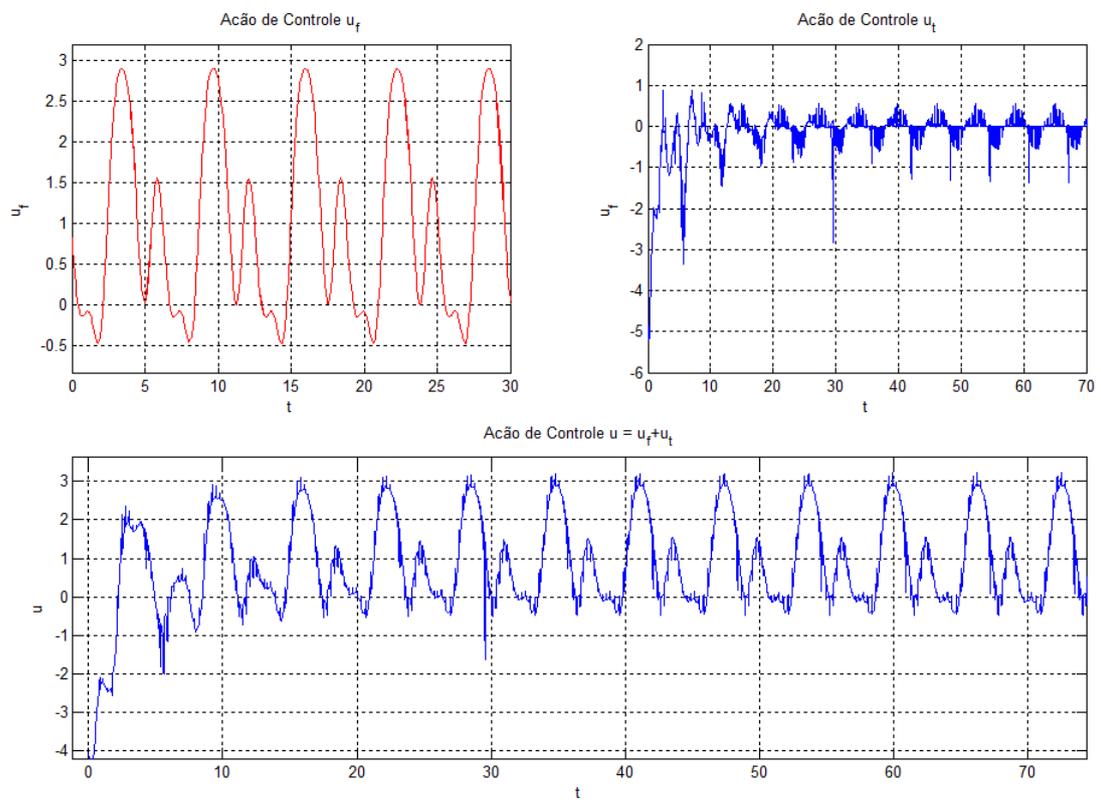


Figura 4.7: Ação de controle para o pêndulo com o controle implementado para pólos em -5 e -6

Conclusões

Neste trabalho foi apresentada uma metodologia de cálculo numérico rigoroso da Forma Normal de Floquet, $X(t) = Q(t)e^{tR}$ das soluções de uma Equação Diferencial Ordinária com coeficientes periódicos $x' = A(t)x$. O método foi baseado na obtenção da série de Fourier de $Q(t)$ e da matriz constante R de modo a fazer com que $X(t)$ satisfaça a EDO e a condição inicial, o que conduziu a uma equação $f(x) = 0$ em que $f : \Omega^s \rightarrow \Omega^s$, Ω^s um espaço de Banach. Definiu-se uma dimensão m para a construção de um espaço de dimensão finita para projeção X^m , o que conduziu à definição um operador $T : \Omega^s \rightarrow \Omega^s$ cujos pontos fixos estão em correspondência com as soluções de $f(x) = 0$.

Foi obtida uma solução aproximada para $f(x) = 0$, denotada por $(\bar{x}, 0^\infty)$ em que $\bar{x} \in X^m$ é tal que $f^{(m)}(\bar{x}) \approx 0$ e X^m tem dimensão finita, $f^{(m)} : X^m \rightarrow X^m$. O passo final consiste na aplicação dos Polinômios *Radii* para calcular $r > 0$ tal que $T : B_{\bar{x}, 0^\infty}(r) \subset \Omega^s \rightarrow B_{(\bar{x}, 0^\infty)}(r)$ seja uma contração. Assim, usando o Teorema do Ponto Fixo de Banach é possível garantir a existência de $x \in B_{(\bar{x}, 0^\infty)}(r) \in \Omega^s$ tal que $f(x) = 0$, que corresponde à Forma Normal de Floquet da equação $x' = A(t)x$.

Como aplicação foi apresentada uma técnica de controle de sistemas dinâmicos τ -periódicos, da forma $x' = f(t, x)$ em que $f(t, x) = f(t + \tau, x) \forall t \in \mathbb{R}$, através de uma estratégia de controle linear. Foi possível fazer com que soluções de $x' = f(t, x)$ sejam dirigidas para uma órbita periódica ou um ponto estacionário. O sistema foi controlado através da inserção de um termo $u : \mathbb{R} \rightarrow \mathbb{R}^n$ resultado na equação $x' = f(t, x) + u$.

O método numérico abordado neste trabalho pode ser aplicado aos mais diversos tipos de problema em ciência aplicada, nos quais a precisão nos resultados seja essencial. Embora o procedimento seja altamente complexo e computacionalmente caro, atividades de projeto usualmente não requerem execução em tempo real. Dessa forma, a metodologia descrita pode ser utilizada, por exemplo, em complexos problemas de Engenharia, largamente abordados nos dias atuais.

Um trabalho futuro interessante é comparar o desempenho em termos computacionais da técnica numérica rigorosa apresentada e a metodologia utilizada em [19], introduzida por [22], que é baseada na decomposição de $Q(t)$ em uma base de polinômios ortogonais. Do ponto de vista computacional, um passo importante é melhorar o cálculo de J^m , já que o método numérico apresenta melhores resultados (menor valor de r dada uma solução aproximada \bar{x} , indicando que essa solução é boa aproximação) quando $J^m Df^{(m)} \approx I$ é uma boa aproximação.

A utilização da Teoria de Floquet generalizada em controle de sistemas não-lineares conduz também a um trabalho futuro interessante. Trata-se de um resultado matemático potencialmente valioso para a área de controle de sistemas dinâmicos. A construção de um método numérico para cálculo da forma Normal de Floquet pode ser também um problema interessante do ponto de vista matemático e aplicado. Não foi avaliada a possibilidade de utilização de uma técnica numérica baseada em Polinômios *Radii* para esse problema, embora seja possível cogitar sua aplicação nesse problema.

Apêndice A

Logaritmo de Matrizes

Neste Apêndice será apresentada uma sequência de resultados para demonstração da Proposição 1.3. A k -ésima derivada de uma aplicação $f : \mathbb{R} \rightarrow M(n, \mathbb{C})$ é denotada por $f^{(k)}$ como é usual na literatura. Resultados básicos sobre exponencial não serão demonstrados e serão utilizados seguindo a abordagem das Seções 2.1 e 2.2 de [7].

Lema A.1. *Sejam $Z : \mathbb{R} \rightarrow M(n, \mathbb{C})$ uma aplicação diferenciável tal que $Z(t)$ e $Z'(t)$ comutam para todo $t \in \mathbb{R}$ e $f : \mathbb{R} \rightarrow M(n, \mathbb{C})$ definida por $f(t) = e^{Z(t)}$. Então vale que*

$$f^{(k)}(t) = \sum_{i=0}^{k-1} \binom{k-1}{i} f^{(i)}(t) Z^{(k-i)}(t) \quad (\text{A.1})$$

Demonstração. Inicialmente, é importante notar que $(e^{Z(t)})' = e^{Z(t)} Z'(t)$, que corresponde ao enunciado da Proposição 2.2 de [7]. A demonstração deste Lema será feita por indução sobre k . O resultado vale para $k = 1$, já que

$$\sum_{i=0}^{k-1} \binom{k-1}{i} f^{(i)}(t) Z^{(k-i)}(t) = \sum_{i=0}^0 \binom{0}{i} f^{(i)}(t) Z^{(1-i)}(t) = f(t) Z'(t)$$

Assumindo que o resultado seja válido para $k = r$, vamos provar que vale para $k = r + 1$. Como o resultado vale para $k = r$,

$$f^{(r)}(t) = \sum_{i=0}^{r-1} \binom{r-1}{i} f^{(i)}(t) Z^{(r-i)}(t)$$

desse modo

$$\begin{aligned} f^{(r+1)}(t) &= \left(f^{(r)}(t) \right)' = \sum_{i=0}^{r-1} \binom{r-1}{i} \left[f^{(i+1)}(t) Z^{(r-i)}(t) + f^{(i)}(t) Z^{(r-i+1)}(t) \right] = \\ &= \sum_{i=0}^{r-1} \binom{r-1}{i} f^{(i+1)}(t) Z^{(r-i)}(t) + \sum_{i=0}^{r-1} \binom{r-1}{i} f^{(i)}(t) Z^{(r+1-i)}(t) \end{aligned}$$

Definindo $j = i + 1$ na parcela da esquerda da última equação, temos que

$$f^{(r+1)}(t) = \sum_{j=1}^r \binom{r-1}{j-1} f^{(j)}(t) Z^{(r+1-j)}(t) + \sum_{i=0}^{r-1} \binom{r-1}{i} f^{(i)}(t) Z^{(r+1-i)}(t)$$

separando as parcelas de $j = r$ e $i = 0$ fica

$$\begin{aligned} f^{(r+1)}(t) &= f^{(r)}(t) Z^{(1)}(t) + \sum_{j=1}^{r-1} \binom{r-1}{j-1} f^{(j)}(t) Z^{(r+1-j)}(t) + \\ &+ f^{(0)}(t) Z^{(r+1)}(t) + \sum_{i=1}^{r-1} \binom{r-1}{i} f^{(i)}(t) Z^{(r+1-i)}(t) \end{aligned}$$

juntando os dois somatórios, fica

$$f^{(r+1)}(t) = f^{(r)}(t) Z^{(1)}(t) + f^{(0)}(t) Z^{(r+1)}(t) + \sum_{i=1}^{r-1} \left[\binom{r-1}{i} + \binom{r-1}{i-1} \right] f^{(i)}(t) Z^{(r+1-i)}(t)$$

Usando a igualdade $\binom{r-1}{i} + \binom{r-1}{i-1} = \binom{r}{i}$ e que $\binom{r}{r} = \binom{r}{0} = 1$ temos que

$$\begin{aligned} f^{(r+1)}(t) &= \binom{r}{r} f^{(r)}(t) Z^{(1)}(t) + \binom{r}{0} f^{(0)}(t) Z^{(r+1)}(t) + \sum_{i=1}^{r-1} \binom{r}{i} f^{(i)}(t) Z^{(r+1-i)}(t) \\ &= \sum_{i=0}^r \binom{r}{i} f^{(i)}(t) Z^{(r+1-i)}(t) \end{aligned}$$

o que completa a demonstração. \square

O Lema a seguir mostra como podemos calcular o logaritmo de uma matriz $I + N$ em que N é nilpotente

Lema A.2. *Seja $N \in M(n, \mathbb{C})$ uma matriz Nilpotente. Então, definindo $G = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} N^k$, vale que $e^G = I + N$.*

Demonstração. Definamos $M : \mathbb{R} \rightarrow M(n, \mathbb{C})$ por $M(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} (tN)^k$ e $f(t) = e^{M(t)}$. Pela definição de exponencial de matrizes temos que existem $a_0, a_1, \dots, a_k, \dots \in \mathbb{C}$ tais que $e^{M(t)} = a_0 I + a_1 tN + a_2 t^2 N^2 + \dots + a_k t^k N^k + \dots$. De fato, essa soma tem uma quantidade finita de termos não nulos já que N é uma matriz nilpotente. Além disso, cabe lembrar que $M(t)$ é um polinômio em tN , logo comuta com $M'(t)$. Como $f(0) = e^{M(0)} = e^{0I} = I$, temos que $a_0 = 1$. Além disso, $f'(0) = \left[\frac{d}{dt} (e^{M(t)}) \right]_{t=0} = e^{M(0)} M'(0) = e^{0I} N = N$. Por outro lado, $f'(0) = a_1 N$, o que faz com que $a_1 = 1$. Queremos provar que $a_k = 0 \forall k \geq 2$. Afirmamos que é suficiente mostrar que $f^{(k)}(0) = 0 \forall k \geq 2$. De fato, $f^{(k)}(0) = (k!) a_k N^k$. Para isso, vamos utilizar um argumento de indução sobre k . Inicialmente, vejamos que

$$f''(t) = \frac{d^2}{dt^2} (e^{M(t)}) = \frac{d}{dt} (e^{M(t)} M'(t)) = e^{M(t)} [M'(t)]^2 + e^{M(t)} M''(t)$$

Como $M(0) = 0$, $M'(0) = N$ e $M''(0) = -N^2$ temos que $f''(0) = 0$. O passo de indução consiste em supor que $f^{(k)}(0) = 0$ para $k = 2, 3, \dots, m$. Vamos provar então que $f^{(m+1)}(0) = 0$. Usando o Lema A.1 temos que

$$f^{(m+1)}(0) = \sum_{i=0}^m \binom{m}{i} f^{(i)}(0) M^{(m+1-i)}(0)$$

temos que

$$f^{(m+1)}(0) = \binom{m}{0} f^{(0)}(0) M^{(m+1-0)}(0) + \binom{m}{1} f^{(1)}(0) M^{(m+1-1)}(0) + \sum_{i=2}^m \binom{m}{i} f^{(i)}(0) M^{(m+1-i)}(0)$$

Usando o fato de que pela hipótese de indução $f^{(i)}(0) = 0$,

$$f^{(m+1)}(0) = f^{(0)}(0) M^{(m+1)}(0) + m f^{(1)}(0) M^{(m)}(0)$$

Como $f^{(0)}(0) = f(0) = I$, $f^{(1)}(0) = f'(0) = N$, fica

$$f^{(m+1)}(0) = M^{(m+1)}(0) + mNM^{(m)}(0)$$

Além disso, pela definição de $M(t)$, temos que $M^{(m+1)}(0) = (-1)^{m+2}(m!)N^{m+1}$ e $M^{(m)}(0) = (-1)^{m+1}(m-1)!N^m$. Desse modo, $f^{(m+1)}(0) = 0$, o que implica que $e^{M(t)} = I + tN$. Definindo $G = M(1)$ temos que $e^G = I + N$, como queríamos. \square

A seguir fixaremos a notação utilizada para referência à forma canônica de Jordan. Sejam $r, d \in \mathbb{N}$ e as matrizes nilpotentes N e \bar{N} definidas por

$$N = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{r \times r} \quad \bar{N} = \begin{bmatrix} 0 & I_2 & 0 & 0 & 0 \\ 0 & 0 & I_2 & 0 & 0 \\ 0 & 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \cdots & 0 & I_2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{2d \times 2d}$$

consideremos a seguinte notação: $J(\lambda, r) = \lambda I_r + N$ para $\lambda \in \mathbb{C}$. Além disso, para $\gamma \in \mathbb{R}$ definimos

$$F(\gamma, r) = \gamma I_r + N \text{ e para } \alpha, \beta \in \mathbb{R} \text{ a matriz } R(\alpha, \beta, d) = \text{diag} \left(\underbrace{\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}, \dots, \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}}_{d \text{ vezes}} \right) + \bar{N}.$$

Lema A.3. *Sejam $\lambda \in \mathbb{C} \setminus \{0\}$, $\gamma \in \mathbb{R} \setminus \{0\}$, $(\alpha, \beta) \in \mathbb{R}^2 \setminus \{0\}$ e $r, d = 1, 2, \dots$ então*

(i) *Existe $A \in M(r, \mathbb{C})$ tal que $e^A = J(\lambda, r)$*

(ii) *Existem $B_1 \in M(r, \mathbb{C})$ e $B_2 \in M(r, \mathbb{R})$ tais que $e^{B_1} = F(\gamma, r)$ e $e^{B_2} = [F(\gamma, r)]^2$.*

(iii) *Existe $D \in M(2d, \mathbb{R})$ tal que $e^D = R(\alpha, \beta, d)$.*

Demonstração. Como $J = J(\lambda, r) = \lambda I + N$ com N nilpotente temos que $J = \lambda I(I + \frac{1}{\lambda}N)$. Mas como $\frac{1}{\lambda}N$ é nilpotente, o Lema A.2 garante a existência de $G \in M(r, \mathbb{C})$ tal que $e^G = I + \frac{1}{\lambda}N$. Definindo $A = \log \lambda I + G$, temos que $e^A = [e^{\log \lambda I}][e^G] = (\lambda I)(I + \frac{1}{\lambda}N) = \lambda I + N$, completando a prova do ítem (i). A existência de B_1 no ítem (ii) é garantida pela existência de A no ítem (i), uma vez que se $\mathbb{R} \subset \mathbb{C}$.

Para garantir a existência de B_2 basta notar que $F^2 = [F(\gamma, r)]^2 = (\gamma I + N)^2 = \gamma^2 I + 2\gamma N + N^2$. Definindo $L = 2\gamma N + N^2$ temos que $F^2 = \gamma^2 I + L = \gamma^2 I(I + \frac{1}{\gamma^2}L)$. Como $\frac{1}{\gamma^2}L$ é nilpotente, a definição de Logaritmo no Lema A.2 garante a existência de uma matriz real $G \in M(r, \mathbb{R})$ (é real pelo fato de que é definida como uma soma de matrizes reais) tal que $e^G = I + \frac{1}{\gamma^2}L$. Assim, definindo $B_2 = \log \gamma^2 I + G$ temos a demonstração do ítem (ii).

Para provar o ítem (iii) denotemos $K = \text{diag} \left(\underbrace{\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}, \dots, \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}}_{d \text{ vezes}} \right)$. Como $\alpha^2 + \beta^2 \neq 0$ temos que K é

invertível e $K^{-1} = \frac{1}{\sqrt{\alpha^2 + \beta^2}} \text{diag} \left(\underbrace{\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, \dots, \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}}_{d \text{ vezes}} \right)$ Dessa forma, $R = R(\alpha, \beta, d) = K(I + K^{-1}\bar{N})$.

Como \bar{N} e K^{-1} comutam, $K^{-1}\bar{N}$ é nilpotente, o que garante que $\log(I + K^{-1}\bar{N}) \in M(2d, \mathbb{R})$. Precisamos então garantir que $\log(K) \in M(2d, \mathbb{R})$. Mas como

$$\log(K) = \text{diag} \left(\underbrace{\log \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}, \dots, \log \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}}_{d \text{ vezes}} \right)$$

precisamos calcular $\log \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}$. É fácil ver que existe $\theta \in \mathbb{R}$ tal que

$$\begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} = \sqrt{\alpha^2 + \beta^2} \begin{bmatrix} \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} & -\frac{\beta}{\sqrt{\alpha^2 + \beta^2}} \\ \frac{\beta}{\sqrt{\alpha^2 + \beta^2}} & \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \end{bmatrix} = \sqrt{\alpha^2 + \beta^2} \begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix}$$

Desse modo, temos que

$$\log \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix} = \log(\sqrt{\alpha^2 + \beta^2}) I_2 + \log \left(\begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix} \right)$$

Como $\log(\sqrt{\alpha^2 + \beta^2}) \in \mathbb{R}$, temos que provar que $\log \left(\begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix} \right) \in M(2, \mathbb{R})$. O cálculo desse logaritmo será motivado pela Fórmula de Euler para números complexos. Escrevendo

$$\begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix} = (\cos \theta) I + (\text{sen} \theta) \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

e notando que a matriz $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ é tal que $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^2 = -I_2$ (propriedade semelhante ao número complexo

i), é natural que uma tentativa para o logaritmo seja $\begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}$. De fato,

$$e^{\begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}} = \begin{bmatrix} 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots & -\theta + \frac{\theta^3}{3!} - \frac{\theta^5}{5!} + \dots \\ \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots & 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots \end{bmatrix} = \begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix}$$

ou seja,

$$\log \begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix} = \begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}$$

o que completa a demonstração. \square

Na demonstração anterior foi utilizada uma analogia entre números complexos e matrizes 2×2 . Na realidade não se trata de uma analogia, mas sim de um isomorfismo entre os números complexos e o seguinte subgrupo de $M(2)$,

$$\Sigma = \left\{ \begin{bmatrix} \alpha & -\beta \\ \beta & \alpha \end{bmatrix}; \alpha, \beta \in \mathbb{R} \right\}$$

esse isomorfismo é trivial e é dado por

$$\chi(a + bi) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

é fácil ver que trata-se de um isomorfismo $\chi: \mathbb{C} \rightarrow \Sigma$. Um bom livro sobre o assunto [26].

Proposição A.1. *Se $C \in M(n, \mathbb{C})$ é uma matriz não singular, existe uma matriz $B \in M(n, \mathbb{C})$ tal que $e^B = C$. Além disso se $C \in M(n, \mathbb{R})$ é não-singular então existe uma matriz não singular $B \in M(n, \mathbb{R})$ tal que $e^B = C^2$.*

Demonstração. Seja $T \in M(n, \mathbb{C})$ Forma Canônica de Jordan de C , tal que $T = SCS^{-1}$ para alguma matriz $S \in M(n, \mathbb{C})$ não-singular. Se formos capazes de obter $\bar{B} \in M(n, \mathbb{C})$ tal que $e^{\bar{B}} = T$ então $C = S^{-1}TS = S^{-1}e^{\bar{B}}S = e^{S^{-1}\bar{B}S}$. Assim, definindo $B = S^{-1}\bar{B}S$ teremos que $e^B = C$. Como $T = \text{diag}(T_1, \dots, T_k)$ onde cada T_i é um bloco de Jordan, e pelo Lema A.3 temos que para cada $i = 1, 2, \dots, k$ existe uma matriz complexa \bar{B}_i tal que $e^{\bar{B}_i} = T_i$, temos que

$$T = \text{diag} \left(e^{\bar{B}_1}, e^{\bar{B}_2}, \dots, e^{\bar{B}_k} \right) = e^{\text{diag}(\bar{B}_1, \bar{B}_2, \dots, \bar{B}_k)}$$

Definindo a matriz complexa $\bar{B} = \text{diag}(\bar{B}_1, \bar{B}_2, \dots, \bar{B}_k)$, temos que $e^{\bar{B}} = T$, logo $B = S^{-1}\bar{B}S$ e $e^B = C$ como queríamos.

Para provar a segunda parte, suponhamos que $C \in M(n, \mathbb{R})$. Então, pelo Teorema da Forma Canônica de Jordan, existe $P \in M(n, \mathbb{C})$ não-singular tal que $Y = PCP^{-1}$. Assim, $Y^2 = PC^2P^{-1}$. Como Y é uma matriz diagonal por blocos F e R do Lema A.3, Y^2 admite logaritmo real. Assim, usando a mesma demonstração anterior temos o resultado. \square

Observação A.1. *Embora seja possível garantir a existência do logaritmo, ele pode não ser único. Seja $\alpha \in \mathbb{R}$ e, como exemplo, a matriz C definida por*

$$C = \begin{bmatrix} \cos(\alpha) & -\text{sen}(\alpha) \\ \text{sen}(\alpha) & \cos(\alpha) \end{bmatrix}$$

que é tal que $C = e^{B_n} \forall n \in \mathbb{Z}$, em que B_n é dado por

$$B_n = (\alpha + 2n\pi) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Além disso,

$$C^2 = \begin{bmatrix} \cos(2\alpha) & -\text{sen}(2\alpha) \\ \text{sen}(2\alpha) & \cos(2\alpha) \end{bmatrix}$$

O logaritmo de C^2 não é único, pois $C^2 = e^{B_n} \forall n \in \mathbb{Z}$, em que B_n é dado por

$$B_n = (2\alpha + 2n\pi) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Apêndice B

Teoria de Floquet Generalizada

O Teorema da Forma Normal de Floquet, (Teor. 1.2), pode ser generalizado para uma classe de EDO's não-lineares. A abordagem utilizada neste Apêndice segue o exposto em [30]. Sejam as EDO's definidas pela Equações B.1 e B.2 em que $f : \mathbb{R} \times \mathbb{R}^n$ e $\phi : \mathbb{R} \times \mathbb{R}^n$ são funções de classe C^1 .

$$x' = f(t, x) \tag{B.1}$$

$$y' = \phi(t, y) \tag{B.2}$$

Definição B.1. *Seja uma função contínua $H : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. Então H é dita função homeomórfica τ -periódica se satisfaz as condições:*

(i) $H(t + \tau, x) = H(t, x) \forall x \in \mathbb{R}^n \text{ e } \forall t \in \mathbb{R}$

(ii) *Para cada $t \in \mathbb{R}$ fixo, a aplicação $H(t, \bullet) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ é um homeomorfismo.*

Definição B.2 (Equivalência topológica para sistemas periódicos). *Seja $x : \mathbb{R} \rightarrow \mathbb{R}^n$ solução de (B.1). Suponha que exista uma função homeomórfica periódica τ -periódica $H(t, x)$ tal que $y(t) = H(t, x(t))$ é solução de (B.2). Então, (B.1) é dito topologicamente equivalente a (B.2), e $H(t, x)$ é a função de equivalência entre (B.1) e (B.2).*

A Definição B.2 é uma generalização natural de equivalência topológica definida para sistemas autônomos. Além disso, definindo $G(t, \bullet) = H(t, \bullet)^{-1}$ faz com que $G(t, y(t))$ seja solução de (B.1) desde que $y(t)$ seja solução de (B.2). Denote $X(t, t_0, x_0)$ a solução de $x' = f(t, x)$ em que $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ é uma função contínua e $f(t + \tau, x) = f(t, x) \forall t \in \mathbb{R} \text{ e } \forall x \in \mathbb{R}^n$, e que $X(t_0, t_0, x_0) = x_0$. Nesta seção $X(t, t_0, x_0)$ será a solução de (B.1) e $Y(t, t_0, y_0)$ será a solução de (B.2).

Lema B.1. *Para todo $t, \theta \in \mathbb{R}$ e $x, y \in \mathbb{R}^n$, vale que $X(t + \tau, \theta + \tau, x) = X(t, \theta, x)$ e $Y(t + \tau, \theta + \tau, y) = Y(t, \theta, y)$*

Demonstração. Este fato é bastante intuitivo. Defina $\gamma(t) = X(t + \tau, \theta + \tau, x)$. Por sua definição, $\gamma(t)$ satisfaz $\gamma'(t) = X'(t + \tau, \theta + \tau, x) = f(t + \tau, X(t + \tau, \theta + \tau, x)) = f(t, \gamma(t))$ Além disso $\gamma(\theta) = x = X(t, \theta, x)$. Pelo teorema de existência e unicidade de soluções, uma vez que $f \in C^1$ temos que $X(t + \tau, \theta + \tau, x) = \gamma(t) = X(t, \theta, x)$, $\forall t \in \mathbb{R}$. O mesmo vale para Y pois as hipóteses sobre ϕ são as mesmas de f . \square

Além disso, suponha que $J : \mathbb{R}^n \rightarrow \mathbb{R}^n$ seja um homeomorfismo. Então, defina H e G por:

$$\begin{aligned} H(t, x) &= Y(t, 0, JX(0, t, x)) \\ G(t, y) &= X(t, 0, J^{-1}(0, t, y)) \end{aligned} \tag{B.3}$$

Lema B.2. *Para cada $t \in \mathbb{R}$ e $x, y \in \mathbb{R}^n$, temos que:*

$$\begin{aligned} H(t, G(t, y)) &= y \\ G(t, H(t, x)) &= x \end{aligned}$$

Demonstração. Como $H(t, G(t, y)) = H(t, 0, JX(0, t, G(t, y)))$, vamos calcular $X(0, t, G(t, y))$. Note que $X(0, t, G(t, y))$ é a solução da EDO avaliada em 0 com condição inicial de que em $X(t, t, G(t, y)) = G(t, y)$. Mas $G(t, y)$ é a solução da EDO avaliada em t com condição inicial $X(0, 0, J^{-1}Y(0, t, y)) = J^{-1}Y(0, t, y)$. Assim, $X(0, t, G(t, y)) = J^{-1}Y(0, t, y)$, o que implica $H(t, G(t, y)) = H(t, 0, Y(0, t, y)) = y$. O mesmo raciocínio pode ser aplicado para obter a outra igualdade. \square

Lema B.3. *Seja $J : \mathbb{R}^n \rightarrow \mathbb{R}^n$ um homeomorfismo. Se para cada $x \in \mathbb{R}^n$, $JX(\tau, 0, x) = Y(\tau, 0, Jx)$, então para cada $t \in \mathbb{R}$ e $x, y \in \mathbb{R}^n$, temos que*

$$\begin{aligned} G(t + \tau, y) &= G(t, y) \\ H(t + \tau, x) &= H(t, x) \end{aligned}$$

Demonstração. $H(t + \tau, x) = Y(t + \tau, 0, JX(0, t + \tau, x))$. Se em $t = 0$ $Y = JX(0, t + \tau, x)$, em $t = \tau$ $Y = Y(\tau, 0, JX(0, t + \tau, x))$. Logo $H(t + \tau, x) = Y(t + \tau, \tau, Y(\tau, 0, JX(0, t + \tau, x)))$. Usando a hipótese com $x = X(0, t + \tau, x)$ temos que $H(t + \tau, x) = Y(t + \tau, \tau, JX(\tau, 0, X(0, t + \tau, x)))$. Usando o mesmo argumento do Lema B.2, $H(t + \tau, x) = Y(t + \tau, \tau, JX(t, t + \tau, x))$. Usando o deslocamento no tempo de τ , uma vez que f é periódica, segue que $H(t + \tau, x) = Y(t, 0, JX(0, t, x)) = H(t, x)$. A outra igualdade pode ser obtida seguindo o mesmo procedimento. \square

Lema B.4. *Para cada $t_0 \in \mathbb{R}$ e $x_0, y_0 \in \mathbb{R}^n$.*

- (i) $H(t, X(t, t_0, x_0))$ é uma solução de (B.2), se $X(t, t_0, x_0)$ é uma solução de (B.1).
- (ii) $G(t, Y(t, t_0, y_0))$ é uma solução de (B.1), se $Y(t, t_0, y_0)$ é uma solução de (B.2)

Demonstração. Como por definição $H(t, X(t, t_0, x_0)) = Y(t, 0, JX(0, t, X(t, t_0, x_0)))$ e $X(0, t, X(t, t_0, x_0)) = X(0, t_0, x_0)$, $H(t, X(t, t_0, x_0)) = Y(t, 0, JX(0, t_0, x_0))$, que é solução de (B.2). \square

Teorema B.1. *Suponha que $f, \phi \in C(\mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$, então os sistemas (B.1) e (B.2) são topologicamente conjugados se e somente se existe um homeomorfismo $J : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que $JX(\tau, 0, x) = Y(\tau, 0, Jx)$*

Demonstração. (\Leftarrow) Suponha que exista o tal homeomorfismo $J : \mathbb{R}^n \rightarrow \mathbb{R}^n$. As funções G e H definidas por (B.3) são contínuas e periódicas com respeito à variável t . Fixado $t \in \mathbb{R}$, vamos provar que a aplicação $H(t, \bullet) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ é uma bijeção. De fato, a injetividade é garantida pelo Teorema de Existência e Unicidade de Soluções de Equações Diferenciais, pois $H(t, x) = Y(t, 0, \tau X(0, t, x))$ e sendo t fixado se $H(t, x_1) = H(t, x_2)$ então $Y(t, 0, JX(0, t, x_1)) = Y(t, 0, JX(0, t, x_2))$ que, por J ser um homeomorfismo, implica $x_1 = x_2$. A sobrejetividade é garantida pelo Lema B.2.

(\Rightarrow) Se os sistemas (B.1) e (B.2) são equivalentes com função de equivalência $H(t, x)$ então $H(t, X(t, 0, x_0))$ é uma solução de B.2, o que significa que existe $y_0 \in \mathbb{R}^n$ tal que $H(t, X(t, 0, x_0)) = Y(t, 0, y_0)$. Substituindo $t = 0$ fica $H(0, X(0, 0, x_0)) = H(0, x_0) = y_0$, ou seja, $H(t, X(t, 0, x_0)) = Y(t, 0, H(0, x_0))$. Defina então $J = H(0, \bullet)$. Assim, $JX(\tau, 0, x_0) = Y(\tau, 0, Jx_0)$, o que completa a demonstração. \square

Teorema B.2 (de Floquet Generalizado). *Suponha que $X(\tau, 0, x) = Cx$ para alguma matriz C , ou seja, $X(\tau, 0, x)$ é uma função linear de x . Então $x' = f(t, x)$ é topologicamente equivalente ao sistema linear autônomo (B.4), em que $B = \frac{1}{\tau} \log(C)$ e $H(t, x) = e^{tB} X(0, t, x)$ é a função de equivalência.*

$$y' = Bx \quad (\text{B.4})$$

Demonstração. Como $Y(\tau, 0, x) = e^{\tau B} x$ e $B = \frac{1}{\tau} \log(C)$, temos que $Y(\tau, 0, x) = Cx = X(\tau, 0, x)$. Tomando então J como a aplicação identidade de \mathbb{R}^n e aplicando o Teorema B.1, temos que os sistemas são conjugados. Além disso, $H(t, x) = Y(t, 0, X(0, t, x)) = e^{tB} X(0, t, x)$. \square

O Teorema B.2 é uma generalização do Teorema de Floquet. De fato, se $A(t)$ é τ -periódica, considere a Forma Normal de Floquet $X(t) = Q(t)e^{tR}$. Assim, $X(t, 0, x) = Q(t)e^{tR}x$, fazendo com que $X(\tau, 0, x) = Q(\tau)e^{\tau R}x \forall x \in \mathbb{R}^n$. Assim, a matriz C está bem definida para todos os sistemas do tipo $x' = A(t)x$ com $A(t)$ periódica. Para demonstrar que não é necessário que $f(t, x)$ seja linear em x considere o Exemplo B.1, que trata de uma função $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ bastante simples e não-linear.

Exemplo B.1. *Um exemplo apresentado em [30] é a equação $x' = f(t, x)$ em que $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ definida por (B.5).*

$$f(t, x) = \begin{cases} x(x-1)\cos(t); & x \in (0, 1) \\ 0; & x \in (-\infty, 0] \cup [1, \infty) \end{cases} \quad (\text{B.5})$$

Como a EDO B.5 é separável é possível encontrar sua solução usando integração. Assim, a solução encontrada é dada por B.6. Como o período de f é 2π temos que $X(2\pi, 0, x) = x \forall x \in \mathbb{R}$. Assim, f satisfaz as hipóteses do Teorema B.2 com $C = 1$. Assim, a equação $x' = f(t, x)$ pe topologicamente conjugada no sentido apresentado pela Definição B.2 à equação $y' = By$ em que $B = \frac{1}{2\pi i} \log(C) = 0$, ou seja, $y' = 0$.

$$X(t, t_0, x_0) = \begin{cases} \frac{x_0}{x_0 + (1 - x_0)e^{\text{sen}(t) - \text{sen}(t_0)}}; & x_0 \in (0, 1) \\ x_0; & x_0 \in (-\infty, 0] \cup [1, \infty) \end{cases} \quad (\text{B.6})$$

A Função de equivalência é dada por $H(t, x) = X(0, t, x)$.

Embora a Teoria de Floquet Generalizada seja mais recente e mais forte que a Teoria Clássica de Floquet, todo o restante do trabalho será dedicado ao cálculo e aplicações da Forma Normal de Floquet para sistemas Lineares Periódicos, em razão da dificuldade de cálculo de C para funções mais complicadas que a do Exemplo B.1 e em dimensões maiores.

Apêndice C

Séries de Fourier

Este Apêndice contém uma coletânea de resultados elementares sobre Série de Fourier.

Teorema C.1 (da Base Ortonormal). *Sejam $G = \{e_i\}_{i \in \mathbb{N}}$ um conjunto ortonormal de um Espaço com produto interno E e $x_i = \langle x, e_i \rangle$ para cada $x \in E$. São equivalentes:*

(i) Para todo $x \in E$, temos que $x = \sum_{i \in \mathbb{N}} x_i e_i$.

(ii) Para todo $x, y \in E$, $\langle x, y \rangle = \sum_{i \in \mathbb{N}} x_i \bar{y}_i$, em que \bar{y}_i denota o conjugado complexo de y_i .

(iii) $\|x\| = \sqrt{\sum_{i \in \mathbb{N}} |x_i|^2}$.

(iv) O espaço gerado pelas combinações lineares finitas de G é denso em E . Ou seja, para todo $x \in E$ e $\epsilon > 0$ existe N tal que $\left\| x - \sum_{i=0}^N x_i e_i \right\| < \epsilon$.

A demonstração do Teorema da Base Ortonormal pode ser encontrada em [20]. A Equação C.1 define um produto interno para os Espaços de funções complexas e reais de variável real. Denote por $\|\bullet\|_{L^2} = \sqrt{\langle \bullet, \bullet \rangle}$ a norma gerada por esse produto interno.

$$\langle f, g \rangle_{L^2} = \frac{1}{2\tau} \int_0^{2\tau} f(x) \bar{g}(x) dx \quad (\text{C.1})$$

Definição C.1 (Espaço L^2). *Seja $[0, 2\tau] \subset \mathbb{R}$. Defina os conjuntos*

$$L^2([0, 2\tau], \mathbb{K}) = \{f : [0, 2\tau] \rightarrow \mathbb{K}; \|f\|_{L^2} < \infty\}$$

em que \mathbb{K} denota \mathbb{R} ou \mathbb{C} .

Proposição C.1 (Bases Ortonormais de L^2). *Os conjuntos $L^2([0, 2\tau], \mathbb{R})$ e $L^2([0, 2\tau], \mathbb{C})$ munidos do produto interno definido pela Equação C.1 formam Espaços de Hilbert. O conjunto B_1 e B_2 formam bases ortonormais para os Espaço de Hilbert $L^2([0, 2\tau], \mathbb{C})$ e $L^2([0, 2\tau], \mathbb{R})$, respectivamente.*

$$B_1 = \left\{ \frac{1}{\sqrt{2\tau}} e^{ik \frac{2\pi}{2\tau} t} \right\}_{k \in \mathbb{Z}} \quad B_2 = \left\{ \frac{1}{\sqrt{2\tau}}, \frac{1}{\sqrt{2\tau}} \text{sen} \left(k \frac{2\pi}{2\tau} \right), \frac{1}{\sqrt{2\tau}} \cos \left(k \frac{2\pi}{2\tau} \right) \right\}_{k \in \mathbb{N} \setminus \{0\}}$$

No caso deste trabalho o interesse é por funções $g : \mathbb{R} \rightarrow \mathbb{R}$ contínuas e periódicas de período 2τ . Dessa forma, considere $f : [0, 2\tau] \rightarrow \mathbb{R}$ definida como a restrição $f = g|_{[0, 2\tau]}$ temos que $f \in L^2([0, 2\tau], \mathbb{C})$ e em particular a $L^2([0, 2\tau], \mathbb{R})$. Assim, vale a igualdade de (C.2).

$$f(t) =_{L^2} \sum_{k=-\infty}^{\infty} \left\langle f(t), \frac{1}{\sqrt{2\tau}} e^{ik\frac{2\pi}{2\tau}t} \right\rangle_{L^2} \frac{1}{\sqrt{2\tau}} e^{ik\frac{2\pi}{2\tau}t} = \sum_{k=-\infty}^{\infty} f_k e^{ik\frac{2\pi}{2\tau}t} \quad (\text{C.2})$$

em que

$$f_k = \frac{1}{2\tau} \left\langle f(t), e^{ik\frac{2\pi}{2\tau}t} \right\rangle_{L^2} = \frac{1}{2\tau} \int_0^{2\tau} f(t) e^{-ik\frac{2\pi}{2\tau}t} dt \quad (\text{C.3})$$

Como g consiste da extensão periódica de f a série de Fourier de f , define-se a *Série de Fourier* de g como sendo $\sum_{k=-\infty}^{\infty} f_k e^{ik\frac{2\pi}{2\tau}t}$. O símbolo $=_{L^2}$ indica que temos uma igualdade entre funções em L^2 , ou seja, a igualdade pode não valer em todo $t \in [0, 2\tau]$. De fato, o conjunto em que não vale a igualdade tem medida de Lebesgue nula.

Seja $g : \mathbb{R} \rightarrow \mathbb{R}$ uma função 2τ -periódica e tal que $\int_0^{2\tau} |g(t)|^2 dt < \infty$ e defina $f : [0, 2\tau] \rightarrow \mathbb{R}$ a restrição de g ao intervalo $[0, 2\tau]$. Assim, é possível representar a Série de Fourier g decompondo f na base B_1 ou B_2 . Vamos mostrar que elas são equivalentes. De fato, considere a decomposição de f na base B_1

$$f(t) = \sum_{k=-\infty}^{\infty} f_k e^{ik\frac{2\pi}{2\tau}t}$$

Usando a fórmula de Euler temos que

$$f(t) = f_0 + \sum_{k=1}^{\infty} f_k \cos\left(k\frac{2\pi}{2\tau}t\right) + i \sum_{k=-1}^{\infty} f_k \text{sen}\left(k\frac{2\pi}{2\tau}t\right)$$

Como f é uma função real então $f_k = C(f_{-k})$

$$f(t) = f_0 + \sum_{k=1}^{\infty} 2 \text{Re}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) - 2 \text{Im}(f_k) \text{sen}\left(k\frac{2\pi}{2\tau}t\right)$$

O restante desta seção será dedicado à convergência uniforme da série de Fourier. Para tanto, considere a soma parcial da série de Fourier de uma função $H : \mathbb{R} \rightarrow \mathbb{R}$.

$$S_N(t) = H_0 + \sum_{n=1}^N H_n \cos\left(\frac{n\pi}{\tau}t\right) + \sum_{n=1}^N H_n \text{sen}\left(\frac{n\pi}{\tau}t\right) \quad (\text{C.4})$$

Proposição C.2 (Convergência Uniforme da Série de Fourier). *Seja $H : \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua e periódica de período $2L$, com derivada primeira integrável e absolutamente integrável. Então, a série de Fourier de f converge uniformemente para f .*

A Proposição C.2 pede apenas que a derivada primeira de H seja integrável e absolutamente integrável. Se H for de classe C^1 ela atenderá à essa condição. A demonstração completa da proposição C.2 está desenvolvida em [9]. Os dois resultados seguintes demonstram esse caso particular da Proposição C.2.

Proposição C.3. *Se H for uma função tal que $H' \in CP([-L, L])$, então, a série de Fourier de H' pode ser obtida através da série de Fourier de H mediante a derivação termo a termo.*

Demonstração. A demonstração é simples, segue da observação de que A_n pode ser calculado por

$$\begin{aligned} A_n &= \frac{1}{L} \int_{-L}^L H(t) \cos\left(\frac{n\pi}{L}t\right) dt = \left[\frac{1}{L} \frac{L}{n\pi} H(t) \cos\left(\frac{n\pi}{L}t\right) \right]_{-L}^L - \\ &- \int_{-L}^L H'(t) \frac{L}{n\pi} \text{sen}\left(\frac{n\pi}{L}t\right) dt = -\frac{L}{n\pi} \int_{-L}^L H'(t) \text{sen}\left(\frac{n\pi}{L}t\right) dt \end{aligned} \quad (\text{C.5})$$

Deve-se notar no entanto que $\int_{-\tau}^{\tau} H'(t) \operatorname{sen}\left(\frac{n\pi}{L}t\right) dt$ é o termo B_n da série de Fourier de $H'(t)$, já que existe a série de Fourier de H' por pertencer a $CP[-L, L]$. O mesmo pode ser feito para A_n , completando a demonstração. \square

Proposição C.4. *Seja $H \in C([-L, L])$ com $H(-L) = H(L)$. Suponha que $H' \in C([-L, L])$. Então $S_N \rightarrow H$ uniformemente, isto é, existe $N_0 \in \mathbb{N}$ tal que*

$$N \geq N_0 \Rightarrow \sup_{t \in [-L, L]} |H(t) - S_N(t)| < \varepsilon$$

Demonstração. Utilizando a desigualdade de Cauchy-Schwarz,

$$\begin{aligned} |S_N(t) - S_M(t)| &= \left| \sum_{n=M+1}^N \left[A_n \cos\left(\frac{n\pi}{\tau}t\right) + B_n \sin\left(\frac{n\pi}{\tau}t\right) \right] \right| \\ &\leq \sum_{n=M+1}^N \left| A_n \cos\left(\frac{n\pi}{\tau}t\right) + B_n \sin\left(\frac{n\pi}{\tau}t\right) \right| \\ &\leq \sum_{n=M+1}^N \sqrt{A_n^2 + B_n^2} \\ &\leq \sum_{n=M+1}^N \sqrt{\left(\frac{\tau}{n\pi}(A_n)'\right)^2 + \left(\frac{\tau}{n\pi}(B_n)'\right)^2} \\ &= \frac{\tau}{\pi} \sum_{n=M+1}^N \frac{1}{n} \sqrt{[(A_n)']^2 + [(B_n)']^2} \end{aligned} \tag{C.6}$$

Pela desigualdade de Bessel temos que $\sum_{n=M+1}^N \left([(A_n)']^2 + [(B_n)']^2 \right)^{\frac{1}{2}}$ é limitada. Além disso $\sum_{n=1}^{\infty} \left(\frac{1}{n^2}\right)$ converge, fazendo com que S_N seja uma seqüência uniformemente de Cauchy em um espaço completo, logo converge uniformemente. \square

O Lema C.1 é um resultado fundamental para a demonstração do Teorema C.2 e para a derivação de funções definidas em termos de séries de Fourier que serão realizadas no Capítulo 2. Embora sua demonstração seja simples ela requer outros resultados. Uma demonstração completa é apresentada em [16], p. 160.

Lema C.1 (Derivação Termo a Termo). *Seja (f_n) uma seqüência de funções de classe C^1 no intervalo $[0, 2\tau]$. Se para $t \in [0, 2\tau]$ a seqüência de números reais $(f_n(t))$ converge e se as derivadas (f_n') convergem uniformemente em $[0, 2\tau]$ então $(\lim f_n)' = \lim (f_n')$.*

Teorema C.2 (Decaimento dos Coeficientes da Série de Fourier). *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função 2τ -periódica e $\sum_{k=-\infty}^{\infty} f_k e^{ik\frac{2\pi}{\tau}t}$ sua Série de Fourier. Se $f \in C^{r-1}$ e a r -ésima derivada, $f^{(r)}$ é contínua por partes então existe $C > 0$ tal que para cada $k \neq 0$ vale*

$$|f_k| \leq \frac{C}{|k|^r}$$

Reciprocamente, se existe $C > 0$ tal que $|f_k| \leq \frac{C}{|k|^{r+\alpha}}$ com $k > 0$, $\alpha > 0$ então $f \in C^{k-1}$

Demonstração. A demonstração deste fato é uma consequência imediata da demonstração da Proposição C.3. O resultado é válido para $r = 1$. Suponha que seja válido para $r = p$. Então se $f \in C^{p-1}$ com $f^{(p)} \in CP([0, 2\tau])$ existe $C > 0$ tal que para cada $k \neq 0$ $|f_k| \leq \frac{C}{|k|^p}$. Seja $f \in C^p$ com $f^{(p+1)} \in CP([0, 2\tau])$. Pela Proposição C.3 temos que

$$\left| (f^{(p)})_k \right| = \frac{D}{k} \left| (f^{(p-1)})_k \right| \leq \frac{D}{k} \frac{C}{k^p} = \frac{DC}{k^{p+1}}$$

completando a primeira parte da demonstração.

Considere a série de Fourier de f , $f(t) = \sum_{k=-\infty}^{\infty} f_k e^{ik\frac{2\pi}{2\tau}t}$. Para cada $t \in [0, 2\tau]$ considere a sequência numérica

$$f_n(t) = \sum_{k=-\infty}^n f_k e^{ik\frac{2\pi}{2\tau}t} = f_0 + \sum_{k=1}^n \left[2 \operatorname{Re}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) - 2 \operatorname{Im}(f_k) \operatorname{sen}\left(k\frac{2\pi}{2\tau}t\right) \right]$$

assim, temos que

$$f'_n(t) = k\frac{2\pi}{2\tau} \sum_{k=1}^n \left[-2 \operatorname{Re}(f_k) \operatorname{sen}\left(k\frac{2\pi}{2\tau}t\right) - 2 \operatorname{Im}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) \right]$$

Como $|f_k| \leq \frac{C}{k^{1+\alpha}}$ com $\alpha > 0$ então $(f'_n(t))_{n \in \mathbb{N}}$ converge uniformemente em $[0, 2\tau]$. Assim,

$$\lim (f'_n(t)) = (\lim (f_n(t)))' = f'(t)$$

Dessa forma temos que

$$f'(t) = k\frac{2\pi}{2\tau} \sum_{k=1}^{\infty} \left[-2 \operatorname{Re}(f_k) \operatorname{sen}\left(k\frac{2\pi}{2\tau}t\right) - 2 \operatorname{Im}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) \right]$$

Fazendo o mesmo procedimento até $k-2$ chegamos a função

$$f^{(r-2)}(t) = \left(\frac{2\pi}{2\tau}\right)^{r-2} k^{r-2} \sum_{k=1}^{\infty} \left[\pm 2 \operatorname{Re}(f_k) \operatorname{sen}\left(k\frac{2\pi}{2\tau}t\right) \pm 2 \operatorname{Im}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) \right]$$

Definindo novamente para cada $t \in [0, 2\tau]$ a sequência

$$f_n^{(r-2)}(t) = \left(\frac{2\pi}{2\tau}\right)^{r-2} k^{r-2} \sum_{k=1}^n \left[\pm 2 \operatorname{Re}(f_k) \operatorname{sen}\left(k\frac{2\pi}{2\tau}t\right) \pm 2 \operatorname{Im}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) \right]$$

temos que sua derivada

$$f^{(r-1)}(t) = \left(\frac{2\pi}{2\tau}\right)^{r-1} k^{r-1} \sum_{k=1}^n \left[\pm 2 \operatorname{Re}(f_k) \operatorname{sen}\left(k\frac{2\pi}{2\tau}t\right) \pm 2 \operatorname{Im}(f_k) \cos\left(k\frac{2\pi}{2\tau}t\right) \right]$$

converge uniformemente pela hipótese de que $|f_k| \leq \frac{C}{k^{r+\alpha}}$. Veja a necessidade de $\alpha > 0$, uma vez que $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. Isso completa a demonstração. \square

Seja $A : \mathbb{R} \rightarrow M(n, \mathbb{R})$ uma aplicação periódica de período τ , tal que $a_{ij} : \mathbb{R} \rightarrow \mathbb{R}$ possui série de Fourier

$$a_{ij}(t) = \sum_{k=-\infty}^{\infty} a_{ij}^k e^{ik\frac{2\pi}{\tau}t}$$

com $a_{ij}^k \in \mathbb{C} \forall k \in \mathbb{Z}$. Então, a matriz A pode ser escrita como

$$A(t) = \sum_{k=-\infty}^{\infty} A_k e^{ik\frac{2\pi}{\tau}t} \quad (\text{C.7})$$

em que A_k é a matriz em que o termo da linha i e da coluna j é a_{ij}^k .

Definição C.2 (Série de Fourier de uma aplicação matricial). A Equação C.7 é dita *Série de Fourier de A*.

Proposição C.5. Sejam $A(t) = \sum_{k=-\infty}^{\infty} (A_{k,1} + iA_{k,2}) e^{ik\frac{2\pi}{\tau}t}$ e $Q(t) = \sum_{k=-\infty}^{\infty} (Q_{k,1} + iQ_{k,2}) e^{ik\frac{2\pi}{2\tau}t}$ as séries de Fourier de período 2τ das aplicações $A, Q : \mathbb{R} \rightarrow M(n, \mathbb{R})$ periódicas, de períodos τ e 2τ , respectivamente. Então valem as seguintes propriedades:

$$(i) (AQ)_k = \sum_{r=-\infty}^{\infty} A_r Q_{k-r}$$

$$(ii) Q_{k,1} = Q_{-k,1} \text{ e } Q_{k,2} = -Q_{-k,2} \quad \forall k \in \mathbb{Z}.$$

Demonstração. Para demonstrar o ítem (i) calculamos os coeficientes da série de Fourier da aplicação $t \mapsto A(t)Q(t)$

$$(AQ)_k = \frac{1}{2\tau} \int_0^{2\tau} A(t) Q(t) e^{-ik\frac{2\pi}{2\tau}t} dt = \frac{1}{2\tau} \int_0^{2\tau} \left(\sum_{r=-\infty}^{\infty} A_r e^{ir\frac{2\pi}{2\tau}t} \right) Q(t) e^{-ik\frac{2\pi}{2\tau}t} dt$$

reescrevendo a série como limite de somas parciais e rearranjando os termos

$$(AQ)_k = \frac{1}{2\tau} \int_0^{2\tau} \left(\lim_{M \rightarrow \infty} \sum_{r=-M}^M A_r e^{ir\frac{2\pi}{2\tau}t} \right) Q(t) e^{-ik\frac{2\pi}{2\tau}t} dt = \frac{1}{2\tau} \int_0^{2\tau} \lim_{M \rightarrow \infty} \sum_{r=-M}^M A_r e^{i(r-k)\frac{2\pi}{2\tau}t} Q(t) dt$$

Usando o Teorema da Convergência dominada para cada entrada das aplicações matriciais e, além disso, trocando as posições da integral com a soma finita,

$$(AQ)_k = \frac{1}{2\tau} \lim_{M \rightarrow \infty} \sum_{r=-M}^M \int_0^{2\tau} A_r e^{i(r-k)\frac{2\pi}{2\tau}t} Q(t) dt = \frac{1}{2\tau} \lim_{M \rightarrow \infty} \sum_{r=-M}^M A_r \int_0^{2\tau} e^{i(r-k)\frac{2\pi}{2\tau}t} Q(t) dt$$

Desse modo,

$$(AQ)_k = \frac{1}{2\tau} \lim_{M \rightarrow \infty} \sum_{r=-M}^M A_r \int_0^{2\tau} e^{i(r-k)\frac{2\pi}{2\tau}t} Q(t) dt = \lim_{M \rightarrow \infty} \sum_{r=-M}^M A_r \underbrace{\frac{1}{2\tau} \int_0^{2\tau} e^{i(r-k)\frac{2\pi}{2\tau}t} Q(t) dt}_{Q_{k-r}} = \sum_{r=-\infty}^{\infty} A_r Q_{k-r}$$

completando a demonstração do ítem (i).

Para demonstrar o ítem (ii) temos que:

$$\bar{A}_k = A_{k,1} - iA_{k,2} = \int_{-\infty}^{\infty} A(t) e^{ik\frac{2\pi}{2\tau}t} dt = A_{-k} = A_{-k,1} + iA_{-k,2}$$

Dessa forma, $A_{k,1} = A_{-k,1}$ e $A_{k,2} = -A_{-k,2}$ como queríamos. \square

Apêndice D

Método de Runge-Kutta para solução aproximada de EDOs

Os métodos de Runge-Kutta formam uma família consagrada de métodos numéricos para aproximação de soluções de Equações Diferenciais Ordinárias. Um membro dessa família largamente utilizado é o Método de Runge-Kutta de quarta ordem, frequentemente denominado apenas Método de Runge-Kutta, que será apresentado neste Apêndice. A base teórica utilizada como referência é [15]. A equação a ser resolvida é $X' = A(t)X$ com $X(0) = I$. Como o objetivo é obter uma aproximação $X(2\tau)$ vamos tomar a seguinte partição de $[0, 2\tau]$, $\Gamma = \{t_0 = 0, \Delta t, 2\Delta t, 3\Delta t, \dots, 2\tau = t_N\}$ em que $\Delta t = \frac{2\tau}{N}$. Dessa forma, vamos definir o seguinte esquema iterativo

$$\begin{cases} X_{k+1} = X_k + \frac{\Delta t}{6} (F_1 + 2F_2 + 2F_3 + F_4) \\ X_0 = I \\ t_k = k(\Delta t) \end{cases} \quad (\text{D.1})$$

em que os termos F_1, F_2, F_3 e F_4 são definidos por

$$\begin{cases} F_1 = A(t_k) X_k \\ F_2 = A\left(t_k + \frac{\Delta t}{2}\right) \left(X_k + \frac{\Delta t}{2} F_1\right) \\ F_3 = A\left(t_k + \frac{\Delta t}{2}\right) \left(X_k + \frac{\Delta t}{2} F_2\right) \\ F_4 = A(t_k + \Delta t) (X_k + \Delta t F_2) \end{cases}$$

Dessa forma, uma aproximação para $\tilde{X}(k\Delta t)$ de $X(k\Delta t)$ é dada por $\tilde{X}(k\Delta t) = X_k$, $0 \leq k \leq N$. Assim, uma aproximação para $X(2\tau)$ é X_N .

Originalmente encontra-se na literatura os métodos de Runge-Kutta para solução de equações do tipo $x' = f(t, x)$ com $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$. Se definirmos $f(t, x) = A(t)x$ com $x : \mathbb{R} \rightarrow \mathbb{R}^n$ é fácil ver que a forma com que esses métodos são apresentados na literatura podem ser facilmente adaptados para equações do tipo $X' = A(t)X$

com $X : \mathbb{R} \rightarrow \in M(n, \mathbb{R})$ já que $A(t)$ atua sobre cada coluna de $X(t)$ em separado, ou seja

$$\begin{bmatrix} x'_1 & x'_2 & \dots & x'_n \end{bmatrix} = A(t) \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \Leftrightarrow \begin{cases} x'_1 = A(t) x_1 \\ x'_2 = A(t) x_2 \\ \vdots \\ x'_n = A(t) x_n \end{cases}$$

A outra equação cuja solução é necessária para a obtenção do ponto inicial do algoritmo rigoroso e que foi resolvida utilizando o método de Runge-Kutta é dada por

$$\begin{cases} Q' = A(t) Q - QR \\ Q(0) = I \end{cases}$$

Além do cálculo do ponto inicial, as soluções das EDOs que representam o Oscilador Duffing e o Pêndulo Forçado Periodicamente, apresentadas nas Figuras 4.1, 4.1, 4.5 e 4.6 foram obtidas através do método de Runge-Kutta de 4ª ordem. Assim, como as aplicações do método neste trabalho não requerem extrema precisão, não serão apresentados resultados sobre os erros de truncamento associados ao método de Runge-Kutta. Deve-se ressaltar que, por se tratar de um método numérico aproximado, não é possível garantir nem mesmo que o método resulte uma solução periódica para $Q(t)$ na equação anterior. Para ilustrar esse caso, consideremos a equação do pêndulo sem amortecimento dada por

$$\frac{d^2\theta}{dt^2} + \omega^2 \sin \theta = 0$$

Definindo $x = \theta$ e $y = \theta'$ temos que

$$\begin{bmatrix} x \\ y \end{bmatrix}' = \begin{bmatrix} y \\ -\omega^2 \sin x \end{bmatrix}$$

As trajetórias dessa equação devem satisfazer $\frac{1}{2}y^2 + \omega^2(1 - \cos(x)) = C_0$, onde $C_0 \in \mathbb{R}$ é obtida através do ponto inicial $(x(0), y(0))$, conforme descrito no capítulo 9 de [3] p.275. Tomando o ponto inicial $(x(0), y(0)) = (1, 1)$ e $\omega = 1$, temos que a figura a seguir apresenta a solução obtida utilizando o método de Runge-Kutta de 4ª ordem e $(C(i) - C_0)$, onde $C(i) = \frac{1}{2}y(i)^2 + \omega^2(1 - \cos(x(i)))$, sendo $(x(i), y(i))$ os valores obtidos pelo método numérico. Idealmente $(C(i) - C_0) = 0, \forall i \in \mathbb{R}$. No entanto, quanto maior o valor de i verifica-se uma tendência de aumento dessa diferença, em virtude de haver um erro em cada iteração do algoritmo de Runge-Kutta.

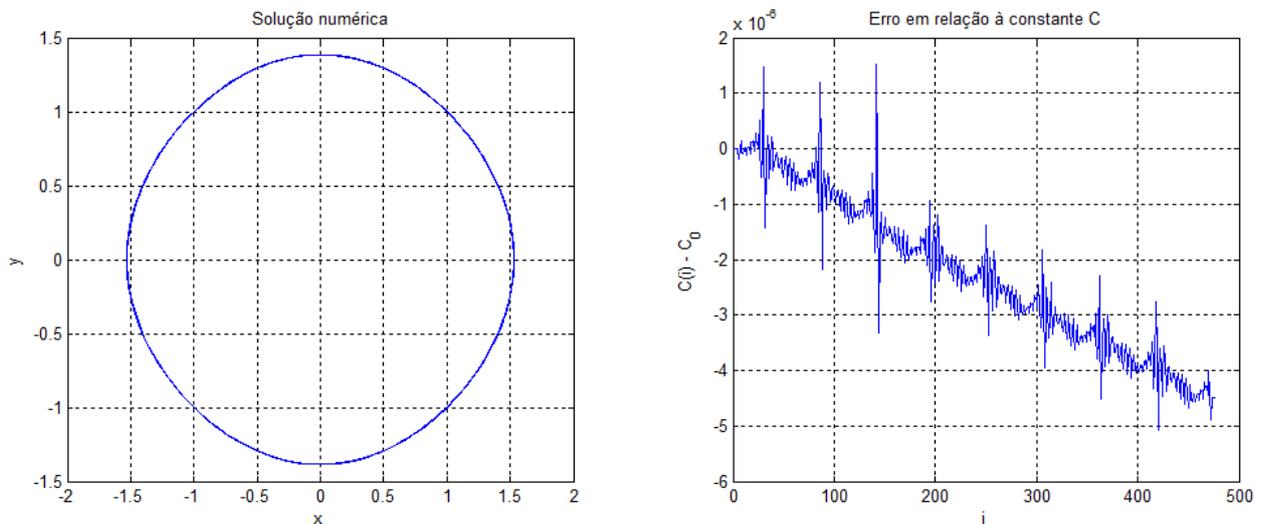


Figura D.1: Solução pelo método de Runge-Kutta. Diferença entre $C(i)$ e C_0

Apêndice E

Introdução à Aritmética Intervalar

A Aritmética Intervalar (AI), também conhecida como Aritmética de Intervalos consiste essencialmente na realização de cálculos numéricos utilizando intervalos ao invés de números Reais. A Aritmética de Ponto Flutuante (APF) utilizada por algoritmos computacionais possui precisão limitada pela quantidade de bits utilizados para a representação de um número. Dessa forma, é impossível representar fielmente um número irracional através da APF devido aos erros de arredondamento, produzindo resultados incorretos. Essa deficiência é solucionada utilizando a AI, através da representação de números Reais por intervalos. Como exemplo, consideremos que a representação computacional disponível para a realização de cálculos numéricos para solução de um determinado problema tem precisão de 14 casas decimais. Assim a representação de π é $3,14159265358979$. Como π é um número irracional há um erro de truncamento na representação utilizada. No entanto, é possível garantir que $\pi \in [3,14159265358978, 3,14159265358980]$. Dessa forma, utiliza-se a representação de π como esse intervalo, sendo necessário redefinir as operações entre números Reais como operações entre intervalos de números racionais.

A ideia de representar números Reais como intervalos foi introduzida na década de 1950 por diversas pessoas. No entanto uma abordagem mais precisa sobre a Análise Intervalar foi apresentada por Moore em 1966. A AI pode ser aplicada em diversas áreas da Análise Numérica e utilizada em diversas aplicações práticas, tais como problemas de Engenharia, Projeto Assistido por Computador e principalmente Demonstrações Assistidas por Computador. Diversas conjecturas tem sido recentemente provadas utilizando a AI, sendo a mais famosa delas a *Conjectura de Keppler*, que permaneceu em aberto por cerca de 400 anos. Os conceitos apresentados neste trabalho são baseados nas referências [18] e [21].

E.1 Aritmética Intervalar Real

Para Aritmética Intervalar utiliza-se comumente intervalos fechados com a seguinte notação:

$$\mathbf{x} = [x, \bar{x}] = \{x \in \mathbb{R}; \underline{x} \leq x \leq \bar{x}\}$$

Dessa forma, o conjunto de todos os intervalos de números Reais é denotado por \mathbb{IR}

$$\mathbb{IR} = \{[x, \bar{x}]; x, \bar{x} \in \mathbb{R}; \underline{x} \leq \bar{x}\}$$

Antes de definir as operações para intervalos Reais é necessário definir a notação utilizada. O ponto médio de um intervalo

$$\text{mid}(\mathbf{x}) = \tilde{x} = \frac{1}{2} (\underline{x} + \bar{x})$$

O raio de um intervalo é definido por

$$\text{rad}(\mathbf{x}) = \frac{1}{2} (\bar{x} - \underline{x})$$

Um intervalo \mathbf{x} é dito *fino* quando $\text{rad}(\mathbf{x}) = 0$, ou seja, quando $\underline{x} = \bar{x}$. O valor absoluto de um intervalo

$$|\mathbf{x}| = (\mathbf{x}) = \max \{|x|; x \in \mathbf{x}\}$$

A mignitude de um intervalo

$$|\mathbf{x}| = \text{mig}(\mathbf{x}) = \min \{|x|; x \in \mathbf{x}\}$$

A distância entre dois intervalos \mathbf{x} e \mathbf{y} é definida por

$$d(\mathbf{x}, \mathbf{y}) = \max(|\underline{x} - \underline{y}|, |\bar{x} - \bar{y}|)$$

Definição E.1 (Relações de Ordem em \mathbb{IR}). *Definimos as seguintes relações de ordem em \mathbb{IR} :*

$$\begin{aligned} \mathbf{x} < \mathbf{y} &\Leftrightarrow \bar{x} < \underline{y} \\ \mathbf{x} \leq \mathbf{y} &\Leftrightarrow \bar{x} \leq \underline{y} \\ \mathbf{x} > \mathbf{y} &\Leftrightarrow \underline{x} > \bar{y} \\ \mathbf{x} \geq \mathbf{y} &\Leftrightarrow \underline{x} \geq \bar{y} \end{aligned} \tag{E.1}$$

Definição E.2 (Definição das operações básicas). *Define-se as operações elementares para intervalos através das seguintes expressões*

$$\left\{ \begin{array}{l} \mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \bar{x} + \bar{y}] \\ \mathbf{x} - \mathbf{y} = [\underline{x} - \bar{y}, \bar{x} - \underline{y}] \\ \mathbf{x} \times \mathbf{y} = [\min\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}, \max\{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}] \\ \mathbf{1}/\mathbf{x} = [1/\bar{x}, 1/\underline{x}] \text{ se } \underline{x} > 0 \text{ ou } \bar{x} < 0 \\ \mathbf{x}/\mathbf{y} = \mathbf{x} \times \mathbf{1}/\mathbf{y} \end{array} \right. \tag{E.2}$$

Se $0 \in [\underline{x}, \bar{x}]$ definimos \mathbf{x}/\mathbf{y} por

$$\frac{\mathbf{x}}{\mathbf{y}} = \left\{ \begin{array}{ll} [\bar{x}/\infty] & \text{se } \bar{x} \leq 0 \text{ e } \bar{y} = 0 \\ [-\infty, \bar{x}/\bar{y}] \cup [\bar{x}/\underline{y}, \infty] & \text{se } \bar{x} \leq 0 \text{ e } \underline{y} < 0 < \bar{y} \\ [-\infty, \bar{x}/\bar{y}] & \text{se } \bar{x} \leq 0 \text{ e } \underline{y} = 0 \\ [-\infty, \infty] & \text{se } \underline{x} < 0 < \bar{x} \\ [-\infty, \underline{x}/\underline{y}] & \text{se } \underline{x} \geq 0 \text{ e } \bar{y} = 0 \\ [-\infty, \underline{x}/\underline{y}] \cup [\underline{x}/\bar{y}, \infty] & \text{se } \underline{x} \geq 0 \text{ e } \underline{y} < 0 < \bar{y} \\ [\underline{x}/\bar{y}, \infty] & \text{se } \underline{x} \geq 0 \text{ e } \underline{y} = 0 \end{array} \right. \tag{E.3}$$

Como o objetivo é definir diversas aplicações basando-se nas operações básicas definidas anteriormente, é necessário garantir que a adição e a multiplicação intervalar tenham propriedades similares às propriedades elementares da adição e multiplicação de números Reais. Isso é feito nas proposições que seguem, cuja demonstração é bastante simples e não será apresentada.

Proposição E.1 (Propriedades da adição intervalar). *Sejam $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{IR}$. A adição intervalar definida pela Definição E.2 tem as seguintes propriedades:*

- **Fechamento:** $\mathbf{x}, \mathbf{y} \in \mathbb{IR} \implies \mathbf{x} + \mathbf{y} \in \mathbb{IR}$
- **Associatividade:** $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$
- **Comutatividade:** $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
- **Possui elemento neutro:** $([0, 0])$.

Proposição E.2 (Propriedades da multiplicação intervalar). *Sejam $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{IR}$. A multiplicação intervalar definida pela Definição E.2 tem as seguintes propriedades*

- **Fechamento:** $\mathbf{x}, \mathbf{y} \in \mathbb{IR} \implies \mathbf{x} \times \mathbf{y} \in \mathbb{IR}$
- **Associatividade:** $\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) = (\mathbf{x} \times \mathbf{y}) \times \mathbf{z}$
- **Comutatividade:** $\mathbf{x} \times \mathbf{y} = \mathbf{y} \times \mathbf{x}$
- **Possui elemento neutro:** $([1, 1])$
- **Subdistributividade:** $\mathbf{x} \times (\mathbf{y} + \mathbf{z}) \subseteq \mathbf{x} \times \mathbf{y} + \mathbf{x} \times \mathbf{z}$.

Dada uma função $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$, é preciso definir como aplicar a função a um intervalo. A rigor seria necessário definir uma extensão da função f . Neste trabalho vamos admitir um abuso de notação e definir a avaliação de uma função de variável real em um intervalo.

Definição E.3 (Avaliação de uma função em um intervalo). *Seja $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$. Define-se a avaliação da função f em um intervalo como sendo $f(\mathbf{x}) = \{f(x); x \in \mathbf{x}\}$*

Uma vez definidas as operações elementares e a avaliação de uma função em um intervalo pode-se construir algoritmos implementando as principais funções utilizadas em Matemática tais como seno, cosseno, exponencial, etc. O pacote computacional INTLAB[®] contém diversas funções já implementadas internamente, disponíveis para utilização.

E.2 Aritmética Intervalar Complexa

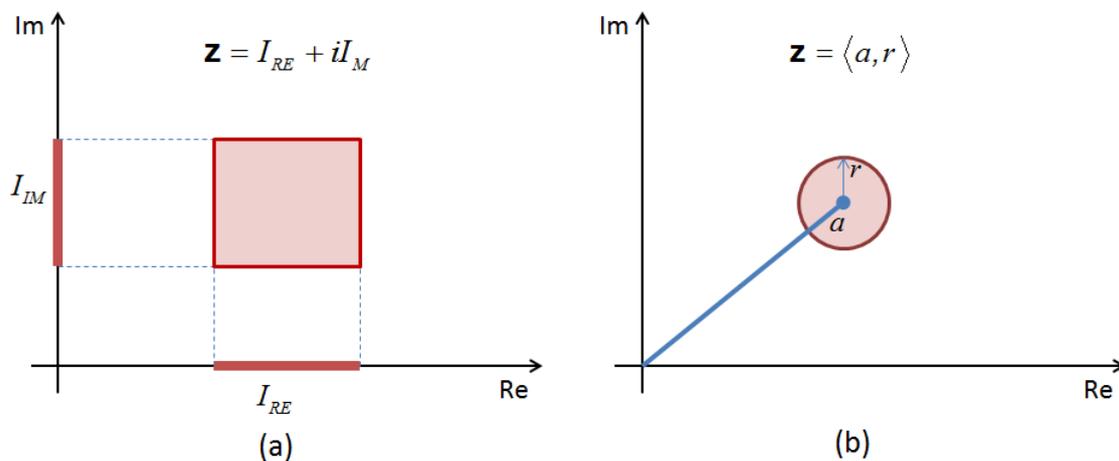


Figura E.1: Representação de intervalos complexos na forma retangular (a) e circular (b)

Vamos utilizar um abuso de linguagem ao fazer referência aos chamados intervalos complexos, que tem objetivos análogos aos intervalos Reais: eliminar possíveis erros numéricos de truncamento. Um intervalo complexo corresponde, formalmente, a uma região do plano complexo, mais precisamente um conjunto fechado e simplesmente conexo do plano complexo. Vamos analisar apenas dois tipos de intervalos complexos, representados da Figura E.1. O tipo de intervalo representado na Figura E.1 (a) pode ser definido como

$$\mathbf{z} = \mathbf{x} + i\mathbf{y} = \{x + iy; x \in \mathbf{x}; y \in \mathbf{y}\} \quad (\text{E.4})$$

Utilizando a definição de um intervalo complexo como apresentado na equação acima leva trivialmente à definição da multiplicação de intervalos complexos dada pela equação a seguir, em que $\mathbf{z}_1 = \mathbf{x}_1 + i\mathbf{y}_1$ e $\mathbf{z}_2 = \mathbf{x}_2 + i\mathbf{y}_2$.

$$\mathbf{z}_1 \times \mathbf{z}_2 = \mathbf{x}_1 \times \mathbf{x}_2 - \mathbf{y}_1 \times \mathbf{y}_2 + i(\mathbf{x}_1 \times \mathbf{y}_2 + \mathbf{x}_2 \times \mathbf{y}_1) \quad (\text{E.5})$$

Analogamente, define-se a multiplicação de dois intervalos complexos como

$$\mathbf{z} = ([1, 2] + i[1, 2]) \times ([3, 4] + i[3, 4])$$

que resulta no seguinte intervalo complexo

$$z = [-5, 5] + i[6, 16]$$

No entanto, o conjunto $\{ab; a \in [1, 2] + i[1, 2]; b \in [3, 4] + i[3, 4]\}$ não é um retângulo no plano complexo, mas sim a área representada na Figura E.2, que está contida na região definida de acordo com a operação de multiplicação, que é o retângulo $[-5, 5] \times [6, 16]$. Dessa forma, há uma perda de precisão do método desnecessariamente, o que prejudica a técnica de aritmética de intervalos complexa. O segundo tipo de intervalo

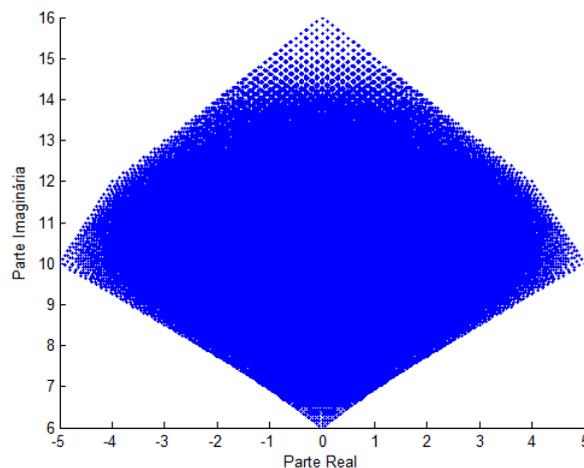


Figura E.2: Região correspondente à multiplicação de dois intervalos complexos

complexo apresentado na Figura E.1 (b), denominado intervalo complexo circular, consiste de uma região circular de raio r centrada num ponto a do plano complexo, ou seja,

$$\mathbf{z} = \langle a, r \rangle = \{z \in \mathbb{C}; |z - a| \leq r\} \quad (\text{E.6})$$

Existem diversas definições de multiplicação de intervalos complexos circulares. Sejam $\mathbf{z}_1 = \langle a_1, r_1 \rangle$ e $\mathbf{z}_2 = \langle a_2, r_2 \rangle$ dois intervalos complexos circulares. Uma das definições mais simples para o produto $\mathbf{z}_1 \times \mathbf{z}_2$ foi introduzida por Gargantini e Henrici, que corresponde à seguinte expressão

$$\mathbf{z}_1 \times \mathbf{z}_2 = \langle a_1 a_2, |a_1| r_2 + |a_2| r_1 + r_1 r_2 \rangle \quad (\text{E.7})$$

Proposição E.3. *A definição da multiplicação de intervalos circulares pela Equação (E.7) atende à propriedade de que a multiplicação de quaisquer dois números complexos pertencentes aos intervalos pertence ao intervalo gerado pela multiplicação dos mesmos, ou seja*

$$\mathbf{z}_1 \times \mathbf{z}_2 \subset \{pq; p \in z_1; q \in z_2\}$$

Demonstração. Como queremos estimar o $|pq - a_1a_2|$, temos que

$$|pq - a_1a_2| = |a_1(q - a_2) + a_2(p - a_1) + (p - a_1)(q - a_2)|$$

Aplicando a desigualdade triangular fica

$$|pq - a_1a_2| \leq |a_1||q - a_2| + |a_2||p - a_1| + |p - a_1||q - a_2|$$

Mas $|p - a_1| \leq r_1$ e $|q - a_2| \leq r_2$, logo

$$|pq - a_1a_2| \leq |a_1|r_2 + |a_2|r_1 + r_1r_2$$

que completa a demonstração. □

E.3 Representação de Vetores e Matrizes

A representação de vetores e matrizes através de aritmética de intervalos é imediata a partir da representação de números Reais e complexos. Dessa forma, um intervalo matricial consiste simplesmente de uma matriz cujas entradas são intervalos Reais ou Complexos.

Referências Bibliográficas

- [1] M. BARTUCCELLI, G. GENTILE, and K. GEORGIOU. On the dynamics of a vertically driven damped planar pendulum. *Proceedings of the Royal Society A*, pages 3007–3022, 2001.
- [2] A. BEN-ISRAEL. *Generalized Inverses*. Springer-Verlag, 1^a edition, 2003.
- [3] W. BOYCE and R DiPRIMA. *Equações Diferenciais Elementares e Problemas de Valores de Contorno*. LTC, 8th edition, 2006.
- [4] H. BREZIS. *Analyse Fonctionnelle - Théorie et Application*. Masson, 1983.
- [5] R. CASTELLI and J.P. LESSARD. Rigorous numerics in Floquet theory: computing stable and unstable bundles of periodic orbits. *SIAM Journal on Applied Dynamical Systems*, 12:204–245, 2013.
- [6] C. CHICONE. *Ordinary Differential Equations with Applications*. Springer, 2006.
- [7] C. DOERING and A LOPES. *Equações Diferenciais Ordinárias*. IMPA, 4th edition, 2010.
- [8] A. DÁVID. Bifurcation control of nonlinear systems with time-periodic coefficients. *Journal of Dynamic Systems, Measurement and Control ASME*, 125,:541–558, 2003.
- [9] D. G. FIGUEIREDO. *Análise de Fourier e Equações Diferenciais Parciais*. IMPA, 1^a edition, 1977.
- [10] G. FLOQUET. Sur les Équations différentielles linéaires à coefficients périodiques. *Annales de l'École Normale Supérieure*, 12:47–88, 1883.
- [11] M. GAMEIRO and J.P. LESSARD. Analytic estimates and rigorous continuation for equilibria of higher-dimensional pdes. *Journal of Differential Equations*, 249(9), pages 2237–2268, 2010.
- [12] S. GERSCHGORIN. Über die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Academie des Sciences de l'URSS*, pages 749–754, 1931.
- [13] J. GUCKENHEIMER and J. G. HOLMES. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Springer, 2nd edition, 1993.
- [14] N. J. HIGHAM. *Functions of Matrices. Theory and Computation*. Society for Industrial and Applied Mathematics, Paris.
- [15] J. D. HOFFMAN. *Numerical Methods for Engineers and Scientists*. Marcel Dekker, Inc., 2nd edition, 2001.
- [16] E. LIMA. *Análise Real*. IMPA, 10^a edition, 2008.
- [17] E. LIMA. *Curso de Análise*. Sociedade Brasileira de Matemática - SBM, 2010.

- [18] R. MOORE, R. KEARFOTT, and M. CLOUD. *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics - SIAM, 2009.
- [19] N. J. PERUZZI. *Dinâmica Não Linear e Controle de Sistemas Ideais e Não-Ideais Periódicos*. PhD thesis, Universidade Estadual de Campinas, 2005.
- [20] W. RUDIN. *Functional Analysis*. McGraw-Hill, 2nd edition, 1991.
- [21] S.M. Rump. INTLAB - INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. <http://www.ti3.tu-harburg.de/rump/>.
- [22] S. C. SINHA and E. A. BUTCHER. Symbolic computation of fundamental solution matrices for linear time-periodic dynamical systems. *Journal of Sound and Vibration*, 206(1):61–85, 1997.
- [23] S. C. SINHA and V. S. DESHMUKH. Control of dynamic systems with time-periodic coefficients via the Lyapunov-Floquet transformation and backstepping technique. *Journal of Vibration and Control*, 10:1517–1533, 2004.
- [24] S. C. SINHA and A. DÁVID. Control of chaos in nonlinear systems with time-periodic coefficients. *Phil. Trans. Royal Society A*, 364:2417–2432, 2006.
- [25] S. C. SINHA, E. GOURDON, and Y. ZHANG. Control of time-periodic systems via symbolic computation with application to chaos control. *Communications in Nonlinear Science and Numerical Simulation*, 10:835–854, 2005.
- [26] M SOARES. *Cálculo em Uma Variável Complexa*. IMPA, 5th edition, 2012.
- [27] J. B. VAN DER BERG, LESSARD. J. P, and K. MISCHAIKOW. Global smooth solution curves using rigorous branch following. *Math. Comp.*, 79(271):1565–1584, 2010.
- [28] S. WIGGINS. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer, 2nd. edition, 2003.
- [29] W. YANG, W. CAO, T. CHUNG, and J. MORRIS. *Applied Numerical Methods Using Matlab*. Wiley Interscience, 2005.
- [30] C. ZOU and J. SHI. Generalization of the Floquet theory. *Nonlinear Analysis*, 71:1100–1107, 2009.