

MARCIO EISENCRAFT

**CONTRIBUIÇÕES DA TEORIA DA ESTIMAÇÃO PARA
MODULAÇÕES DIGITAIS QUE UTILIZAM SINAIS CAÓTICOS**

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Engenharia Elétrica.

São Paulo
2006

MARCIO EISENCRAFT

**CONTRIBUIÇÕES DA TEORIA DA ESTIMAÇÃO PARA
MODULAÇÕES DIGITAIS QUE UTILIZAM SINAIS CAÓTICOS**

Tese apresentada à Escola Politécnica
da Universidade de São Paulo para
obtenção do título de Doutor em
Engenharia Elétrica.

Área de Concentração:
Sistemas Eletrônicos

Orientador:
Prof. Dr. Luiz Antonio Baccalá

São Paulo
2006

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 20 de abril de 2006

Marcio Eisencraft

Prof. Dr. Luiz Antonio Baccalá

Eisencraft, Marcio

Contribuições da teoria da estimação para modulações digitais que utilizam sinais caóticos.

São Paulo, 2006.

190p.

Edição Revisada

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Telecomunicações e Controle.

1. Caos (sistemas dinâmicos) 2. Comunicações digitais I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II.t

Ao Professor Max Gerken (in memoriam)

AGRADECIMENTOS

Ao orientador Prof. Luiz A. Baccalá pelo apoio e orientação na elaboração deste trabalho mesmo nos momentos mais difíceis.

Ao Prof. Max Gerken pelos conselhos e exemplos a serem seguidos e pelas orientações de iniciação científica, mestrado e da fase inicial do doutorado.

Aos colegas e amigos professores Maria D. S. Miranda e Magno T. M. Silva pelo contínuo incentivo e colaboração.

Aos professores Leonardo A. B. Torres, Elbert E. N. Macau, Iberê L. Caldas e José R. C. Piqueira pelas contribuições dadas quando da defesa desta tese.

Ao meu tio-avô Prof. Moyses Szajnbok pela leitura meticulosa da tese e pelas valiosas discussões.

A meus pais Nelson e Helena pelo apoio e paciência, a minha avó Regina pelos exemplos de coragem e dedicação e aos demais familiares.

À minha noiva Danielle Lueth e meus grandes amigos Guilherme Araújo e Marcio Manako pelo companheirismo e compreensão.

Ao Laboratório de Comunicações e Sinais da Escola Politécnica da Universidade de São Paulo pela oportunidade de realização deste doutorado.

À Escola de Engenharia da Universidade Presbiteriana Mackenzie por acreditar em minha capacidade e apoiar a concretização deste sonho.

Aos professores que, quer seja em salas de aula, quer seja em livros, acompanharam-me ao longo de meus vinte e nove anos.

A todos os colegas, amigos e alunos que direta ou indiretamente colaboraram na realização deste trabalho.

“Vivemos neste mundo para nos esforçar a aprender sempre, para esclarecer uns aos outros, por meio de troca de idéias, e para nos aplicar todos os dias e cada vez mais às Ciências e às Artes”

W. A. Mozart

“Não falaremos aqui das vantagens da Matemática Teórica em todos os ramos da atividade humana. Digamos apenas que, na Engenharia, os práticos sem teoria poderão fazer o que ontem se fazia, mas serão incapazes de criar o que se fará amanhã.”

J. L. A. Lisboa

“A função de um especialista não é estar mais certo do que as outras pessoas, mas estar errado por motivos mais sofisticados.”

D. Buttler

Resumo

Neste trabalho investiga-se o emprego de técnicas de estimação em sistemas de modulação digital que utilizam sinais caóticos. Inicialmente, aspectos básicos das teorias de sistemas não-lineares e de modulações digitais são revisitados seguidos de técnicas recentemente propostas de modulações digitais caóticas com receptores por correlação coerente, não-coerente e diferencial: o CSK (*Chaos Shift Keying*), o DCSK (*Differential Chaos Shift Keying*) e algumas de suas variantes, em especial o FM-DCSK (*Frequency Modulated DCSK*). Nessa descrição, utiliza-se a notação de equivalente passa-baixas de tempo discreto para facilitar a comparação com modulações digitais convencionais. Deduz-se o limite de Cramér-Rao para a estimação da condição inicial de órbitas caóticas em função de propriedades estatísticas do mapa que as gerou e descrevem-se dois estimadores para elas: o MLE (*Maximum Likelihood Estimator*) que se aplica a mapas com densidade invariante uniforme e o algoritmo de Viterbi para o qual se apresenta uma generalização a fim de aplicá-lo a uma classe maior de mapas. Por apresentar ganho de estimação maior na faixa de relação sinal-ruído de interesse, este último é utilizado em propostas de sistemas de modulação digital que utilizam estimação de órbitas para detectar o símbolo enviado: o ML-CSK (*Maximum Likelihood CSK*) modificado para poder usar mapas com densidade invariante não-uniforme, empregando um ou dois mapas e o ML-DCSK (*Maximum Likelihood DCSK*). Por simulação, avaliou-se o desempenho em termos de taxa de erro desses sistemas sob ruído branco aditivo gaussiano.

Abstract

In this work, we investigate the use of estimation techniques to digital modulation systems that use chaotic signals. Initially, basic aspects of nonlinear systems and digital modulation theory are reviewed followed by currently proposed techniques of chaotic digital modulation with coherent, noncoherent and differential correlation receivers: CSK (Chaos Shift Keying), DCSK (Differential Chaos Shift Keying) and some of its variants in special FM-DCSK (Frequency Modulated DCSK). These systems are described using a discrete-time lowpass equivalent model to facilitate comparison with conventional digital modulation systems. We derive Cramér-Rao lower bounds for the estimation of the initial condition of chaotic orbits as a function of the statistical properties of the chaos generating map and describe two chaotic orbits estimators: the MLE (Maximum Likelihood Estimator) that applies only to maps with uniform invariant density and the Viterbi algorithm for which a generalization is presented that allows its application to a broader class of maps. Because of the larger estimation gains attained in the signal-to-noise ratio range of interest, the latter is used in proposed digital modulation systems that use orbit estimation to detect the transmitted symbol: ML-CSK (Maximum Likelihood CSK) modified to allow maps with nonuniform invariant density using one map or two maps and ML-DCSK (Maximum Likelihood DCSK). The performance of these systems in terms of symbol error rate is accessed via simulation under additive white gaussian noise perturbations.

Sumário

Lista de Figuras	13
Lista de Tabelas	19
Lista de Abreviaturas	21
Lista de Símbolos	23
1 Introdução	26
1.1 Sinais caóticos	26
1.2 Aplicações de sinais caóticos em comunicações	30
1.2.1 Espalhamento espectral por seqüência direta	31
1.2.2 Modulação analógica	32
1.2.3 Modulação digital	32
1.3 Algumas linhas de pesquisa atuais	36
1.4 Objetivos	38
1.5 Organização da tese	38
2 Sinais caóticos em tempo discreto: natureza e caracterização	40
2.1 Conceitos e definições básicas	41
2.2 Estabilidade	42
2.2.1 Estabilidade no sentido de Lagrange	42
2.2.2 Estabilidade de órbitas estacionárias	42
2.2.3 Estabilidade de órbitas genéricas	43
2.2.4 Conjuntos limites	45

2.2.5	Estabilidade estrutural	45
2.3	Número e expoente de Lyapunov	46
2.4	Sinais caóticos	49
2.5	Exemplos de mapas caóticos	51
2.5.1	Mapa tenda $f_T(\cdot)$	51
2.5.2	Mapas tenda inclinada $f_I(\cdot)$	52
2.5.3	Mapa quadrático $f_Q(\cdot)$	54
2.6	Densidade invariante e o operador de Frobenius-Perron	55
2.7	Conclusões	62
3	Elementos básicos de comunicação digital	65
3.1	Estrutura utilizada	66
3.1.1	O modelo de canal	67
3.2	Modelos equivalentes de tempo discreto	69
3.2.1	Sistemas passa-banda	69
3.2.2	Modelo equivalente em banda base de tempo discreto	70
3.3	Modulação e demodulação	71
3.4	Configurações de receptores	72
3.5	Exemplo: o PSK binário	74
3.6	Conclusões	77
4	Modulação digital usando portadoras caóticas	78
4.1	O chaveamento caótico (CSK)	79
4.1.1	Modulação CSK com um mapa	81
4.1.2	Demodulação do CSK com um mapa	84
4.1.3	Modulação CSK com dois mapas	86
4.1.4	Demodulação por correlação coerente do CSK com dois mapas	87
4.2	O chaveamento caótico diferencial (DCSK)	89
4.2.1	Modulação DCSK	90
4.2.2	Demodulação DCSK	91
4.3	O problema da variabilidade da energia por símbolo	94

4.4	O FM-DCSK	96
4.5	Comparação qualitativa do desempenho em canal AWGN	100
4.5.1	Receptor por correlação não-coerente	101
4.5.2	Receptor por correlação coerente e a sincronização caótica	103
4.5.3	Receptor diferencial	105
4.6	Curvas de desempenho em canal AWGN	106
4.6.1	Curvas de SER para o CSK com recepção não-coerente	107
4.6.2	Curvas de SER para o DCSK e FM-DCSK com recepção diferencial	111
4.7	Conclusões	114
5	Estimação de sinais caóticos e de suas condições iniciais	116
5.1	Limites de desempenho para estimadores de condições iniciais	117
5.1.1	O limite inferior de Cramér-Rao (CRLB)	118
5.1.2	O CRLB na estimação da condição inicial de mapas caóticos	122
5.1.3	Exemplos	126
5.1.4	Comentários	132
5.2	O MLE	133
5.2.1	O MLE - Definição e propriedades	133
5.2.2	Estimação da condição inicial com o MLE	134
5.2.3	Estimação de órbitas com o MLE	137
5.3	O algoritmo de Viterbi	141
5.3.1	Partição do espaço de estados	141
5.3.2	O algoritmo de Viterbi para estimar órbitas caóticas	143
5.3.3	Simulações e resultados numéricos	147
5.4	Comparações entre o algoritmo de Viterbi e o MLE	156
5.5	Conclusões	157
6	Modulação digital utilizando estimação de sinais caóticos	159
6.1	O ML-CSK modificado	160
6.1.1	O ML-CSK com duas funções de base	160
6.1.2	O ML-CSK com uma função de base	164

6.2	O ML-DCSK	166
6.3	Curvas de desempenho em canal AWGN	167
6.3.1	Curvas de desempenho do ML-CSK modificado	167
6.3.2	Curvas de desempenho do ML-DCSK	171
6.4	Conclusões	172
7	Conclusões e trabalhos futuros	174
7.1	Contribuições	175
7.2	Conclusões	178
7.2.1	Modulações digitais usando sinais caóticos baseadas em correlação . . .	179
7.2.2	Estimação de condições iniciais de sinais caóticos	180
7.2.3	Estimação de sinais caóticos	181
7.2.4	Sistemas de modulação envolvendo estimação de sinais caóticos	182
7.3	Trabalhos Futuros	182
	Referências Bibliográficas	184

Lista de Figuras

1.1	Órbitas do mapa quadrático da Eq. (1.1) com diferentes condições iniciais: (a) $s_1(n)$ com $s_1(0) = 0,7$; (b) $s_2(n)$ com $s_2(0) = 0,70000001$; (c) módulo da diferença entre as duas órbitas.	27
1.2	Características espectrais e de covariância de órbitas do mapa $f_Q(\cdot)$ da Eq. (1.1). (a) Estimativa da densidade espectral de potência (DEP) da órbita $s_1(n)$ com condição inicial $s_1(0) = 0,7$; (b) função de autocorrelação de $s_1(n)$; (c) função de correlação cruzada entre $s_1(n)$ e a órbita $s_2(n)$ com condição inicial $s_2(0) = 0,4$. $C_{s_1s_1}$ foi normalizada de forma que $C_{s_1s_1}(0) = 1$	30
1.3	Diagrama de blocos de um sistema de comunicação digital caótico com uma única função de base.	33
2.1	Diagrama esquemático do comportamento de uma órbita próxima (a) de um ponto fixo atrator (b) de um ponto fixo estável segundo Lyapunov.	43
2.2	(a) Mapa tenda $f_T(s)$; (b) órbita com condição inicial $s_0 = \sqrt{2}/2$	52
2.3	(a) Mapa tenda inclinada $f_I(s)$; (b) órbita com condição inicial $s_0 = 0$ para $\alpha = 0,8$	53
2.4	Número de Lyapunov de uma órbita do mapa $f_I(\cdot)$ em função do parâmetro α	53
2.5	(a) Mapa quadrático $f_Q(\cdot)$; (b) órbita com condição inicial $s_0 = 0,7$	54
2.6	Órbitas e histogramas do mapa $f_Q(\cdot)$ obtidos com $N = 10000$ pontos, $N_S = 40$ intervalos e condições iniciais (a) $s_1(0) = 0,7$ e (b) $s_2(0) = 0,4$	56
2.7	Órbitas e histogramas de $f_Q(\cdot)$ sem comportamento caótico. (a) $s_0 = -\sqrt{\frac{3}{4}}$; (b) $s_0 \approx 0,809016994$. Foram usadas $N = 10000$ amostras e $N_S = 40$ subintervalos.	57

2.8	Contra-imagem do intervalo $[-1, s]$ por $f_Q(\cdot)$ representada por linhas grossas no eixo das abscissas com $\beta = \sqrt{\frac{1-s}{2}}$	58
2.9	(a) Evolução da densidade constante $p(s) = \frac{1}{2}$ por $f_Q(\cdot)$; (b) Evolução da densidade $p(s) = \frac{3}{2}s^2$ por $f_T(\cdot)$. Em azul $p(s)$, em verde $\mathcal{P}p(s)$, em vermelho $\mathcal{P}^2p(s)$ e em preto a densidade invariante, $p_*(s)$	61
3.1	Diagrama de blocos de um sistema de comunicações digitais.	66
3.2	Modelo de canal de comunicações usado nas simulações.	68
3.3	Exemplos de curvas SER para modulações digitais convencionais em AWGN.	68
3.4	Espectro de magnitude típico de sinal gerado por um sistema passa-banda.	69
3.5	Geração de $x_m(n)$ a partir das funções de base, caso $N_b = 2$	72
3.6	Diagrama de blocos de um receptor de correlação com $N_b = 2$	73
3.7	Diagrama de blocos de um sistema PSK binário com detecção coerente: (a) transmissor; (b) receptor.	75
3.8	Histograma da componente observada z_{m1} com $E_b = 1$ (a) na ausência de ruído no canal e (b) com ruído no canal com $E_b/N_0 = 15\text{dB}$. Os histogramas foram calculados em $N_S = 40$ subintervalos tendo sido transmitidos 10000 símbolos. Em tracejado é indicado o nível de decisão.	77
4.1	Geração de $x_m(n)$ a partir das seqüências de base caóticas, caso $N_b = 2$	80
4.2	Obtenção do vetor de observação \mathbf{z}_m por correlação a partir do sinal transmitido, caso $N_b = 2$	81
4.3	Modulador CSK com uma função de base.	82
4.4	Sinais CSK transmitidos para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$: (a) CSK unipolar com $k = 0,5$; (b) COOK; (c) CSK bipolar. Em cada caso, $N = 50$ amostras e $E_b = 1$	83
4.5	Receptor coerente para o CSK usando uma seqüência de base.	84
4.6	Valor esperado dos vetores de observação dos sistemas CSK com uma função de base e $E_b = 1$. (a) CSK unipolar com $k = 0,5$; (b) COOK; (c) CSK bipolar. A linha tracejada indica, em cada caso, o nível de decisão.	84
4.7	Receptor não-coerente para sinal CSK com uma função de base.	85

4.8	Histogramas de z_{m1} para (a) COOK e (b) CSK unipolar com $E_b = 1$ e $k = 0,5$: caso sem ruído.	86
4.9	Receptor coerente para modulação CSK com duas funções de base.	87
4.10	Valores esperados dos vetores de observação para demodulação coerente do CSK com duas seqüências de base e $E_b = 1$. A linha tracejada indica o limite teórico de decisão.	89
4.11	Estrutura dos sinais $s_i(n)$ usados no DCSK.	90
4.12	Diagrama de blocos de um modulador DCSK.	91
4.13	Sinal DCSK transmitido para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$ com $N = 50$ amostras e $E_b = 1$	91
4.14	Diagrama de blocos de um receptor DCSK diferencial.	92
4.15	Valores esperados para os vetores de observação para um receptor DCSK difer- encial.	93
4.16	Saída do correlator DCSK diferencial para os símbolos $\{1, 1, 0, 1, 0, 0, 1, 0\}$. No- vamente $E_b = 1$, $N = 50$ e o mapa tenda $f_T(\cdot)$ foi utilizado como gerador do sinal caótico.	93
4.17	Histograma de z_{m1} para um receptor DCSK diferencial. Caso sem ruído e $E_b = 1$.	94
4.18	Histograma da energia de $s_1(n)$:(a) periódico; (b) órbita do mapa tenda normal- izado com $N = 10$ e (c) órbita do mapa tenda normalizado com $N = 100$. Em cada caso foram considerados 10000 realizações.	95
4.19	Média e desvio-padrão da estimação de E_{s1} em função de N para o mapa tenda normalizado.	96
4.20	Diagrama de blocos de um modulador FM-DCSK em tempo contínuo.	97
4.21	Diagrama de blocos de um modulador FM-DCSK em tempo discreto.	99
4.22	Parte real e parte imaginária do sinal FM-DCSK transmitido para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$ com $N = 50$ amostras e $E_b = 1$	99
4.23	Diagrama de blocos genérico de um demodulador por correlação utilizando sinais caóticos com $N_b = 1$	101

4.24	Histogramas da componente observada: (a) COOK e (b) CSK unipolar ($k = 0,5$) na ausência de ruído; (c) COOK e (d) CSK unipolar ($k = 0,5$) com SNR = 10dB. Em cada caso, $E_b = 1$ e 10000 símbolos foram transmitidos.	103
4.25	Taxas de erro de símbolo do CSK unipolar para $k = 0,5$	107
4.26	Taxas de erro de símbolo do sistema COOK.	108
4.27	(a) Histograma de z_{m1} para o COOK com $E_b/N_0 = 25$ dB; (b) idem para $E_b/N_0 = 40$ dB; (c) nível de decisão em função da SNR no canal para $E_b = 1$ e $N = 100$	110
4.28	Taxas de erro de símbolo para o sistema CSK unipolar em função da constante k . A E_b/N_0 no canal foi mantida em 20dB.	110
4.29	Taxas de erro de símbolo do sistema DCSK.	111
4.30	Taxas de erro de símbolo do sistema FM-DCSK.	113
4.31	Taxas de erro de símbolo em canal AWGN de sistemas de comunicação digital usando sinais caóticos para $N = 10$	113
5.1	Gráficos de $p(s'(0); s_0)$ para alguns valores de s_0	119
5.2	CRLB na estimação da condição inicial de uma órbita de $f_T(\cdot)$ em AWGN. Para comparação, também é mostrado o CRLB na estimação de um sinal constante em AWGN.	127
5.3	CRLB na estimação da condição inicial da órbita $s(n, 0)$ do mapa $f_I(\cdot)$ em função de α para $\sigma_r = 1$ e diversos valores de N . A aproximação da desigualdade (5.26) é mostrada em tracejado.	128
5.4	CRLB na estimação da condição inicial da órbita $s(n, 0,6)$ do mapa $f_I(\cdot)$ em função de α para diversos valores de N e $\sigma_r = 1$. A aproximação da desigualdade (5.26) é mostrada em tracejado.	128
5.5	CRLB na estimação da condição inicial da órbita $s(n, s_0)$ do mapa $f_I(\cdot)$ com (a) $\alpha = 0,8$ e (b) $\alpha = -0,8$ em função de s_0 para diversos valores de N e $\sigma_r = 1$. A aproximação da desigualdade (5.26) é mostrada em tracejado.	129
5.6	CRLB na estimação da condição inicial da órbita $s(n, s_0)$ do mapa $f_Q(\cdot)$ para $N = 3$ e $\sigma_r = 1$	130

5.7	CRLB na estimação da condição inicial da órbita $s(n, s_0)$ do mapa $f_Q(\cdot)$ para diversos valores de N e $\sigma_r = 1$. A aproximação da desigualdade (5.26) é mostrada em tracejado.	131
5.8	Exemplo de função de verossimilhança.	134
5.9	Ganho de estimação do MLE para condição inicial do mapa $f_T(\cdot)$. Em tracejado o limite para o ganho dado por (5.45).	136
5.10	Ganho de estimação do MLE para condição inicial do mapa $f_I(\cdot)$ em função de α para (a) $\text{SNR}_{in} = 20\text{dB}$ e (b) $\text{SNR}_{in} = 90\text{dB}$. Em tracejado os limites teóricos da desigualdade (5.47).	138
5.11	Ganho de estimação do MLE para órbitas do mapa $f_T(\cdot)$ para diversos valores de N . Os limites dados pela expressão (5.49) são indicados em tracejado. . . .	139
5.12	SNR na entrada e na saída do MLE para órbitas do mapa $f_T(\cdot)$	140
5.13	Exemplo de estrutura para a decodificação usando o algoritmo de Viterbi. Neste caso, $N_S = 4$	145
5.14	Subintervalos para o calculo da matriz de transição de estados para o mapa $f_T(\cdot)$ com $N_S = 5$	148
5.15	Subintervalos para o calculo da matriz de transição de estados para o mapa $f_Q(\cdot)$ com $N_S = 5$	149
5.16	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para mapas da família tenda inclinada para diversos valores de α , $N_S = 20$ e $N = 10$. A limitação da Eq. (5.76) é mostrada em tracejado.	150
5.17	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa $f_T(\cdot)$ e diversos comprimentos N da seqüência sendo estimada. $N_S = 20$. A limitação da Eq. (5.76) é mostrada em tracejado.	151
5.18	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa $f_T(\cdot)$ para diversos valores de N_S , o número de subintervalos em que $[-1, 1]$ foi dividido. As limitações da Eq. (5.76) são mostradas em tracejado.	152
5.19	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi utilizando a conjugação para o mapa quadrático para diversos comprimentos N da seqüência estimada e $N_S = 20$. A limitação da Eq. (5.76) é mostrada em tracejado. . . .	154

5.20	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi utilizando a conjugação para o mapa quadrático para diversos valores de N_S e $N = 10$. As limitações da Eq. (5.76) são mostradas em tracejado.	154
5.21	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa quadrático para diversos valores de N sem aplicação do mapa de conjugação. A limitação da Eq. (5.76) é mostrada em tracejado.	155
5.22	(a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa quadrático para diversos valores de N_S sem aplicação do mapa de conjugação. As limitações da Eq. (5.76) são mostradas em tracejado.	155
5.23	Curvas de ganho para o MLE e para o algoritmo de Viterbi para estimação de uma órbita do mapa tenda $f_T(\cdot)$	156
6.1	Diagrama de blocos do receptor ML-CSK com duas funções de base.	160
6.2	Construção do mapa $f_2(\cdot)$ para $f_1(\cdot) = f_T(\cdot)$	161
6.3	Tentativa de construção do mapa $f_2(\cdot)$ para $f_1(\cdot) = f_Q(\cdot)$ usando a regra da Eq. (6.4). Em destaque o ponto fixo superatrator que aparece nesse caso.	162
6.4	Construção do mapa $f_2(\cdot)$ utilizado nas simulações para $f_1(\cdot) = f_Q(\cdot)$	163
6.5	Sinais ML-CSK com duas funções de base para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$: (a) mapa tenda; (b) mapa quadrático. Em ambos os casos $N = 50$ e $E_b = 1$. Os símbolos “1” e “0” são transmitidos por $x_1(n)$ e $x_2(n)$ respectivamente.	164
6.6	Diagrama de blocos do receptor ML-CSK com uma função de base.	165
6.7	Sinais ML-CSK com uma função de base para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$: (a) mapa tenda; (b) mapa quadrático. Em ambos os casos $N = 50$ e $E_b = 1$. Os símbolos “1” e “0” são transmitidos por $x_1(n)$ e $x_2(n)$ respectivamente.	165
6.8	Diagrama de blocos do receptor ML-DCSK. O estimador de Viterbi fornece uma estimativa dos primeiros $N/2$ pontos de $x'_m(n)$	166
6.9	Taxas de erro de símbolo do ML-CSK com duas funções de base utilizando $f_1(\cdot) = f_T(\cdot)$	168
6.10	Taxas de erro de símbolo do ML-CSK com duas funções de base utilizando $f_1(\cdot) = f_Q(\cdot)$	169

6.11	Taxas de erro de símbolo do ML-CSK com uma função de base utilizando $f_1(.) = f_T(.)$	170
6.12	Taxas de erro de símbolo do ML-CSK com uma função de base utilizando $f_1(.) = f_Q(.)$	170
6.13	Taxas de erro de símbolo para os sistemas ML-CSK testados ($N = 10$).	171
6.14	Taxas de erro de símbolo do ML-DCSK utilizando o mapa $f_T(.)$ como gerador e $N_S = 100$	172

Lista de Tabelas

1.1	Algumas modulações digitais caóticas baseadas no CSK e no DCSK.	35
2.1	Teoremas do Capítulo 2	63
2.2	Principais definições do Capítulo 2	64
4.1	Problemas das modulações caóticas estudadas no capítulo.	114
5.1	Valores máximos de SNR_{in} para ganho de estimação positivo.	153
5.2	Teoremas do Capítulo 5	158
7.1	Principais contribuições da tese	176

Lista de Abreviaturas

ASK	<i>Amplitude Shift Keying</i> (Chaveamento de amplitude)
AWGN	<i>Additive White Gaussian Noise</i> (Ruído gaussiano branco aditivo)
CRLB	<i>Cramér-Rao Lower Bound</i> (Limite inferior de Cramér-Rao)
CDSK	<i>Correlation Delay Shift Keying</i> (Chaveamento com atraso e correlação)
COOK	<i>Chaotic On-Off-Keying</i> (Chaveamento liga-desliga caótico)
CSK	<i>Chaos Shift Keying</i> (Chaveamento caótico)
DCI	Dependência sensível às condições iniciais
DCSK	<i>Differential Chaos Shift Keying</i> (Chaveamento caótico diferencial)
DEP	Densidade Espectral de Potência
DPSK	<i>Differential Phase Shift Keying</i> (Chaveamento de fase diferencial)
DS-SS	<i>Direct Sequence Spread Spectrum</i> (Espalhamento espectral por seqüência direta)
FM-DCSK	<i>Frequency Modulated DCSK</i> (DCSK com modulação em frequência)
FSK	<i>Frequency Shift Keying</i> (Chaveamento de frequência)
M-DCSK	<i>Modified DCSK</i> (DCSK modificado)
ML	<i>Maximum Likelihood</i> (Máxima verossimilhança)
MLE	<i>Maximum Likelihood Estimator</i> (Estimador de máxima verossimilhança)
ML-CSK	<i>Maximum Likelihood CSK</i> (CSK com estimação de máxima verossimilhança)
ML-DCSK	<i>Maximum Likelihood DCSK</i> (DCSK com estimação de máxima verossimilhança)
MVU	<i>Minimum Variance Unbiased Estimator</i> (Estimador não-enviesado de mínima variância)
PSK	<i>Phase Shift Keying</i> (Chaveamento de fase)

RF	Radiofrequência
QCSK	<i>Quadrature CSK</i> (CSK em quadratura)
SCSK	<i>Symmetric CSK</i> (CSK simétrico)
SNR	<i>Signal-to-Noise Ratio</i> (Relação sinal-ruído)
SER	<i>Symbol Error Rate</i> (Taxa de erro de símbolo)
SS	<i>Spread Spectrum</i> (Espalhamento espectral)

Lista de Símbolos

Resume-se aqui a simbologia empregada ao longo deste trabalho.

Os vetores colunas são representados por caracteres minúsculos em **negrito**. As matrizes são representadas por caracteres maiúsculos também em **negrito**. Variáveis escalares são apresentadas em *itálico*.

O símbolo $f'(\cdot)$ utilizado no Capítulo 2, Seção 2.3, representa a derivada da função f com relação ao seu argumento. Nos capítulos seguintes, a utilização de apóstrofo ($x'(n)$, $s'(n)$, etc.) refere-se sempre a sinais corrompidos por ruído.

Os símbolos utilizados com maior frequência são listados a seguir.

Símbolos gerais

\mathbb{N}	Conjunto dos números naturais (inteiros não-negativos)
\mathbb{R}	Conjunto dos números reais
\mathbb{R}^N	Conjunto dos vetores reais $1 \times N$
n	Índice de tempo discreto
t	Índice de tempo contínuo
$\#A$	Cardinalidade do conjunto A
$I_A(\cdot)$	Função indicadora para o conjunto A
$f^n(\cdot)$	n -ésima aplicação sucessiva de $f(\cdot)$
$f^{-1}(A)$	Contra-imagem do intervalo A sob a operação de $f(\cdot)$
$p(\cdot)$	Função densidade de probabilidade
$P(A)$	Probabilidade do evento A
$E[.]$	Esperança matemática

\bar{x}	Média temporal do sinal $x(n)$
z^{-1}	Atraso unitário
$[\cdot]^T$	Transposição de vetores ou matrizes

Sinais e sistemas caóticos

U	Domínio de definição dos mapas, $U = [-1, 1]$
$s(n)$	Órbita de um sistema dinâmico
$s(n, s_0)$	Órbita com condição inicial s_0
\mathbf{s}	$[s(0), s(1), \dots, s(N-1)]^T$
$f_T(\cdot)$	Mapa tenda definido na Eq. (2.9)
$f_I(\cdot)$	Mapa da família tenda inclinada definida na Eq. (2.11)
α	Parâmetro da família tenda inclinada
$f_Q(\cdot)$	Mapa quadrático definido na Eq. (1.1)
$\omega(s_0)$	Conjunto limite da órbita $s(n, s_0)$
S_{s_0}	Trajectoria de $s(n, s_0)$
$L(s_0)$	Número de Lyapunov de $s(n, s_0)$
$h(s_0)$	Expoente de Lyapunov de $s(n, s_0)$
\mathcal{P}	Operador de Frobenius-Perron
$p_*(\cdot)$	Função densidade invariante

Modulação digital utilizando sinais caóticos

M	Número de símbolos diferentes utilizados num sistema de comunicações digitais
N_b	Número de elementos do conjunto de funções de base
N	Comprimento das seqüências utilizadas para representar cada símbolo
$s_i(n)$	i -ésimo elemento do conjunto de funções de base ($1 \leq i \leq N_b$)
$\hat{s}_i(n)$	Estimativa de $s_i(n)$ obtida em receptores coerentes
$x_m(n)$	Seqüência enviada representando o m -ésimo símbolo ($1 \leq m \leq M$)
\mathbf{x}_m	$[x_{m1}, x_{m2}, \dots, x_{mN_b}]^T$ - coeficientes da combinação dos $s_i(n)$ que gera $x_m(n)$
\mathbf{z}_m	$[z_{m1}, z_{m2}, \dots, z_{mN_b}]^T$ - saídas dos correladores do receptor ao se transmitir $x_m(n)$

z_T	Limiar de decisão
N_{Sinc}	Número de pontos necessários à sincronização em receptores coerentes
k	Fator de diferenciação das energias dos símbolos do CSK unipolar (Eq. (4.5))
E_b	Energia média por símbolo
E_{bm}	Energia média utilizada na transmissão do m -ésimo símbolo
N_0	Dobro da DEP do ruído branco gaussiano presente no canal

Estimação de órbitas e condições iniciais

$r(n)$	Processo ruído branco gaussiano com média nula
\mathbf{r}	$[r(0), r(1), \dots, r(N-1)]^T$
$\hat{s}(n)$	Órbita estimada
$\hat{\mathbf{s}}$	$[\hat{s}(0), \hat{s}(1), \dots, \hat{s}(N-1)]^T$
$q(n)$	Seqüência de estados de uma órbita
\mathbf{q}	$[q(0), q(1), \dots, q(N-1)]^T$
\mathbf{q}_k	$[q(0), q(1), \dots, q(k)]^T$
$\hat{q}(n)$	Seqüência de estados estimada
$\hat{\mathbf{q}}$	$[\hat{q}(0), \hat{q}(1), \dots, \hat{q}(N-1)]^T$
$\text{mse}(\hat{\theta})$	Erro médio quadrático do estimador $\hat{\theta}$
N_S	Número de intervalos de partição para aplicação do algoritmo de Viterbi
U_j	j -ésimo subintervalo em que U foi dividido ($1 \leq j \leq N_S$)
Δ_S	Comprimento de cada subintervalo para partição uniforme
$\gamma(n, j)$	Probabilidade da seqüência de estados mais provável a terminar no estado j no instante n dado o vetor corrompido por ruído \mathbf{s}'
a_{ij}	Probabilidade de transição do estado i para o estado j
\mathbf{A}	$\{a_{ij}\}_{1 \leq i, j \leq N_S}$ - Matriz de transição de estados
SNR_{in}	Relação sinal-ruído antes da estimação
SNR_{out}	Relação sinal-ruído depois da estimação
G	Ganho de estimação
G_{dB}	Ganho de estimação em decibéis

Capítulo 1

Introdução

Este capítulo, situa esta tese no contexto atual da pesquisa de aplicações de sinais e sistemas caóticos em Telecomunicações. As Seções 1.1 e 1.2 traçam um histórico resumido do estudo dos sinais e sistemas caóticos e de suas aplicações em Engenharia de Telecomunicações. A Seção 1.3 expõe alguns desafios atuais para a aplicação prática de modulações digitais que utilizam sinais caóticas como portadoras. A partir deles, a Seção 1.4 estabelece os objetivos deste trabalho. Por fim, a organização da tese é explicada na Seção 1.5.

1.1 Sinais caóticos

Um *signal caótico* é determinístico, aperiódico e apresenta dependência sensível às condições iniciais, ou seja, se o sistema que o gerou for iniciado com uma condição ligeiramente diferente, o sinal obtido pode apresentar valores completamente distintos do sinal anterior após algumas iterações (ALLIGOOD; SAUER; YORK, 1996).

Um exemplo simples de sistema que gera sinais caóticos é o mapa quadrático

$$s(n+1) = f_Q(s(n)) = -2s^2(n) + 1. \quad (1.1)$$

O mapa $f_Q(\cdot)$ é semelhante ao clássico mapa logístico (ALLIGOOD; SAUER; YORK, 1996) com o domínio de definição alterado para $U = [-1, 1]$. Começando com $s_0 \in U$, obtém-se por meio de iterações sucessivas da Eq. (1.1) uma seqüência de pontos $[s(0), s(1), s(2), \dots]$. A

função $s(n)$, $n \in \mathbb{N}$, é um *senal* ou *órbita* de $f_Q(\cdot)$ com condição inicial $s(0)$.

Ainda que aparentemente simples, as órbitas desse sistema apresentam comportamento complicado. Sabe-se que para quase todas as condições iniciais, ele apresenta órbitas caóticas limitadas no intervalo $[-1, 1]$. Os gráficos (a) e (b) da Figura 1.1 mostram duas de suas órbitas, $s_1(n)$ e $s_2(n)$, com condições iniciais $s_1(0) = 0,7$ e $s_2(0) = 0,70000001$. Vê-se que apesar das condições iniciais estarem próximas, depois de um tempo curto de evolução, as órbitas originadas ficam completamente separadas, como mostra o módulo da diferença entre elas (Figura 1.1(c)).

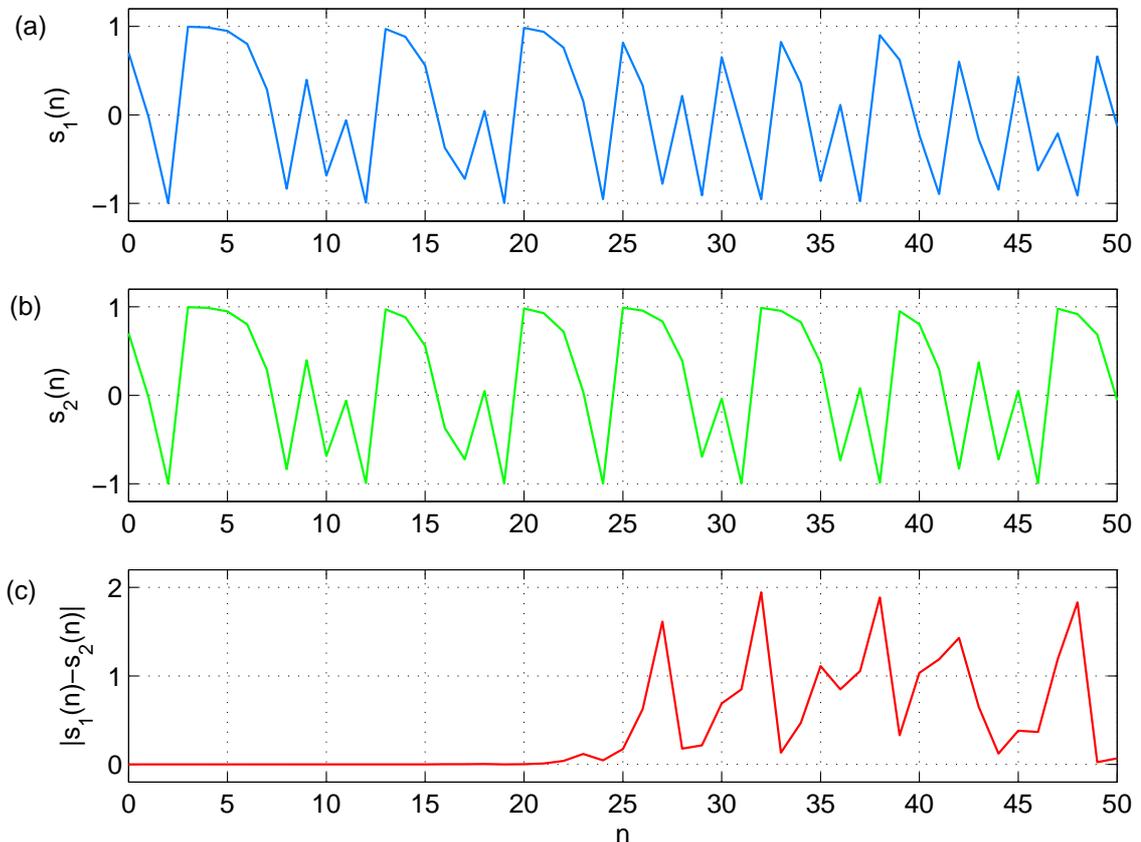


Figura 1.1: Órbitas do mapa quadrático da Eq. (1.1) com diferentes condições iniciais: (a) $s_1(n)$ com $s_1(0) = 0,7$; (b) $s_2(n)$ com $s_2(0) = 0,70000001$; (c) módulo da diferença entre as duas órbitas.

O primeiro trabalho em que aparece a palavra “caos” com o sentido matemático atual foi publicado por Li e Yorke (1975). Porém, o início do estudo dos sistemas e sinais que apresentam comportamento caótico pode ser localizado muito antes na história da Ciência.

Sistemas dinâmicos regidos por equações diferenciais são estudados desde o século XVII.

A partir do trabalho fundamental de Sir Isaac Newton escrito em 1687 (NEWTON, 1999) essas equações vem sendo utilizadas para descrever todo tipo de processo que se desenvolve continuamente no tempo.

Nos séculos XVIII e XIX numerosas técnicas para encontrar soluções para estas equações foram desenvolvidas. No entanto, é freqüentemente impossível escrever essas soluções na forma de expressões algébricas simples, usando um número finito de termos. As soluções em séries envolvendo um número infinito de somas geralmente não convergem num tempo finito. Com as técnicas de então, quando passível de ser encontrado, o comportamento assintótico de uma solução limitada podia ser de três tipos:

- convergência para um ponto fixo; ou
- convergência para uma oscilação periódica; ou
- convergência para uma oscilação quasi-periódica. Esse comportamento aparece quando no sistema em estudo existem várias freqüências naturais incomensuráveis (FERRARA; PRADO, 1994).

Dessa forma, até o século XIX, acreditava-se serem esses os únicos comportamentos assintóticos possíveis. Essa situação só foi alterada quando o matemático francês Henri Poincaré (1890) estudou o clássico problema dos três corpos. Em seu trabalho, utilizando o conceito de variedade e análises qualitativas, Poincaré mostrou que as soluções assintóticas desse problema podem ser muito mais complicadas do que as três possibilidades conhecidas até então. Estava descoberto o que hoje denomina-se *caos*.

Baseado no trabalho de Poincaré, principalmente no seu enfoque qualitativo e geométrico, algumas áreas da Matemática nasceram e desenvolveram-se ao longo do século XX, como a topologia algébrica e geométrica. Porém, uma compreensão mais profunda da natureza das órbitas complicadas observadas por Poincaré só foi obtida na década de 1960.

Naquele momento, com o aparecimento dos computadores digitais, o emprego de simulações computacionais de equações diferenciais passou a fazer parte das pesquisas em diversas áreas e, com ela, o interesse por equações de diferenças e sistemas discretos aumentou. Com essas

simulações, as órbitas apontadas por Poincaré voltaram a ganhar interesse em duas frentes diferentes.

Por um lado, Edward Lorenz (1963) observou que mesmo modelos meteorológicos muito simplificados podem exibir *dependência sensível às condições iniciais* que, como já foi dito, é característica de sistemas caóticos. O trabalho de Lorenz é discutido de modo bastante acessível em (FERRARA; PRADO, 1994).

Por outro lado, o matemático americano Stephen Smale (vencedor da medalha Fields da *International Mathematical Union* em 1966), juntamente com os soviéticos D. V. Anosov, V. I. Arnold e A. N. Kolmogorov e os brasileiros J. Palis e M. M. Peixoto entre muitos outros, analisaram e descreveram o comportamento caótico utilizando principalmente propriedades dos difeomorfismos, ou seja, mapas diferenciáveis cujos inversos também são diferenciáveis (SMALE, 1967).

A partir daí, na década de 1970, apareceram vários artigos sobre aplicações de dinâmica não-linear nas mais diversas áreas. Trabalhos como o do ecologista Robert May (1974) e do físico Mitchell Feigenbaum (1978) entre muitos outros mostraram que o comportamento caótico aparece muito mais freqüentemente do que se esperava tanto nas soluções das equações diferenciais e de diferenças, quanto nos sistemas naturais que elas modelam.

Na década de 1980, a computação de alta velocidade surgiu como uma poderosa aliada dos pesquisadores de dinâmica não-linear. Um dos trabalhos mais importantes daquela época é sem dúvida o de Benoit Mandelbrot (1982) e a sua *geometria fractal*.

A pesquisa na área passou a ter um “componente experimental” e a possibilidade de “visualizar” resultados matemáticos bastante abstratos tornando o assunto mais atraente e acessível e chamando a atenção de pesquisadores de inúmeras áreas. Desde então, o número de aplicações de sistemas não-lineares e caos cresceu muito. Em (STROGATZ, 1998) citam-se dezenas de exemplos dessas aplicações.

Na área de Engenharia de Telecomunicações não foi diferente. Principalmente após a publicação do artigo (PECORA; CARROLL, 1990) sobre sincronização de sistemas caóticos, observou-se um grande número de trabalhos propondo aplicações destes sistemas em comunicações que estão detalhadas na Seção 1.2.

1.2 Aplicações de sinais caóticos em comunicações

Os sinais caóticos possuem algumas propriedades interessantes do ponto de vista da Engenharia de Telecomunicações. Em termos espectrais, os sinais caóticos caracterizam-se por ocuparem uma larga faixa de frequências, terem função de autocovariância com formato impulsivo e função de covariância cruzada com valores muito baixos entre órbitas com condições iniciais diferentes.

Como exemplo, a Figura 1.2 mostra para o mapa $f_Q(\cdot)$ da Eq. (1.1), uma estimativa da densidade espectral de potência da órbita $s_1(n)$ com condição inicial $s_1(0) = 0,7$, sua função de autocovariância $C_{s_1s_1}$ e a função de covariância cruzada com outra órbita $s_2(n)$, $C_{s_1s_2}$ com condição inicial $s_2(0) = 0,4$. Admite-se ergodicidade assintótica para que a densidade espectral de potência esteja bem definida.

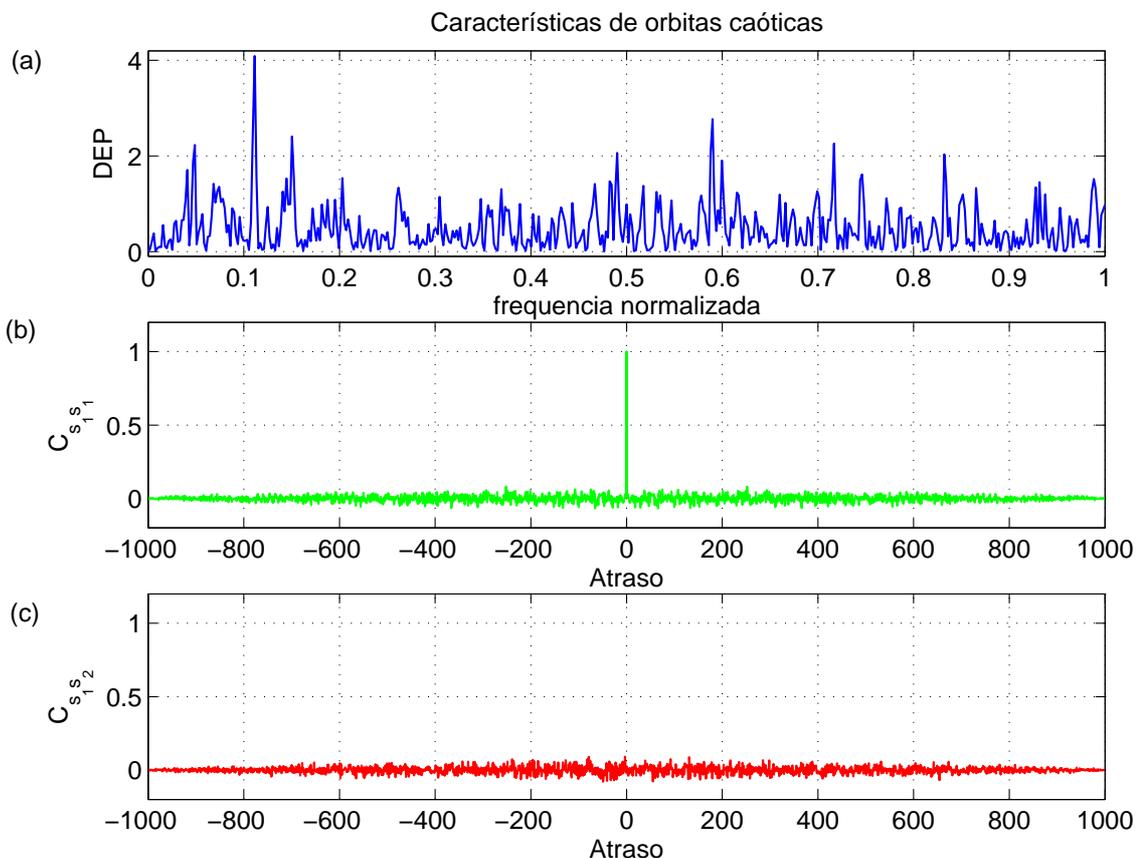


Figura 1.2: Características espectrais e de covariância de órbitas do mapa $f_Q(\cdot)$ da Eq. (1.1). (a) Estimativa da densidade espectral de potência (DEP) da órbita $s_1(n)$ com condição inicial $s_1(0) = 0,7$; (b) função de autocorrelação de $s_1(n)$; (c) função de correlação cruzada entre $s_1(n)$ e a órbita $s_2(n)$ com condição inicial $s_2(0) = 0,4$. $C_{s_1s_1}$ foi normalizada de forma que $C_{s_1s_1}(0) = 1$.

Por essas características, os sinais caóticos são candidatos naturais a atuarem no espalhamento de sinais de informação. Assim, ao se utilizar um sinal caótico para modular sinais de banda estreita independentes desse, os sinais resultantes adquirem largura de banda maior ao mesmo tempo em que se atenua o nível de sua densidade espectral, caracterizando assim um sistema que realiza espalhamento espectral (LATHI, 1998). Dessa forma, os sistemas que utilizam modulação com sinais caóticos possuem as mesmas qualidades dos sistemas de espalhamento espectral convencionais (HAYKIN, 2000).

Sinais caóticos permitem que um grande número de sinais para espalhamento possam ser facilmente produzidos como consequência da dependência sensível às condições iniciais. Também, como visto no exemplo do mapa $f_Q(\cdot)$ da Eq. (1.1), sinais caóticos podem ser gerados de forma extremamente simples. Assim, pode ser que eles venham a ser uma alternativa interessante e de baixo custo para comunicações por espalhamento espectral.

A grande maioria dos trabalhos publicados nos últimos anos envolvendo sinais e sistemas caóticos aplicados em comunicações pode ser dividida em três grandes temas: seqüências para espalhamento espectral por seqüência direta, modulação analógica e modulação digital.

1.2.1 Espalhamento espectral por seqüência direta

O fato dos sinais caóticos de tempo discreto serem seqüências aperiódicas fez com que surgissem vários trabalhos sobre sua aplicação em sistemas de espalhamento espectral por seqüência direta (DS-SS - *Direct Sequence - Spread Spectrum*). A idéia básica é substituir as seqüências de espalhamento binárias convencionais, como as seqüências de Gold (WADE, 2000), por seqüências caóticas geradas por mapas não-lineares. As principais vantagens delas são o número infinito de seqüências de espalhamento existentes e a menor vulnerabilidade às interferências do sinal espalhado (LAU; TSE, 2003).

Um trabalho interessante é (ROVATTI; MAZZINI; SETTI, 2004) em que os autores mostram que sistemas DS-SS utilizando seqüências caóticas podem ter desempenho superior aos sistemas DS-SS convencionais em termos de capacidade e probabilidade de erro de símbolos em situações com múltiplos usuários.

1.2.2 Modulação analógica

Dentre as várias abordagens para transmissão de informação analógica utilizando sinais caóticos, destacam-se os sistemas propostos por Cuomo e Oppenheim (1993a) e por Wu e Chua (1993).

O primeiro, chamado de *mascaramento caótico*, consiste em somar o sinal de informação analógico a ser transmitido à saída de um sistema caótico. No receptor, por meio de um processo chamado de *sincronização caótica*, o sinal caótico original é reconstruído e a informação analógica é então extraída pela subtração do sinal caótico reproduzido do sinal recebido.

No sistema proposto por Wu e Chua (1993) o sinal de informação é injetado num sistema caótico, alterando sua dinâmica. Desta forma, o sinal caótico gerado pelo sistema contém a informação analógica. Utilizando a sincronização caótica, o receptor é capaz de regenerar o sinal caótico sem as modificações causadas pelo sinal de informação que é obtido comparando-se o sinal recebido com o regenerado.

Esses sistemas foram estudados em detalhes na dissertação de mestrado do autor dessa tese (EISENCRAFT, 2001). Uma das conclusões desse trabalho foi que esses sistemas de modulação analógica dependem fundamentalmente da sincronização caótica. Esse tipo de sincronismo, proposto em (PECORA; CARROLL, 1990) depois do trabalho pioneiro (YAMADA; FUJISAKA, 1983), é muito sensível ao ruído limitando assim o emprego desses sistemas, como analisado em (EISENCRAFT; GERKEN, 2000). Não foi encontrado até o momento um sincronismo robusto o suficiente que permita a aplicação prática desses sistemas em canais em que o ruído esteja fortemente presente, apesar de muitos estudos terem sido feitos desde então (veja e.g. (MILLERIOUX; DAAFOUZ, 2001; BOUTAYEB; DAROUACH; RAFARALAH, 2002)).

1.2.3 Modulação digital

Num sistema de comunicação digital convencional cada símbolo entre os M possíveis é transmitido usando sempre a mesma forma de onda. O equivalente em banda base de tempo discreto desse sistema consiste em representar cada símbolo por uma seqüência de N pontos, gerada a partir de um conjunto de seqüências de base ortonormais. Quando funções de base senoidais

são utilizadas sem técnicas de espalhamento espectral, o sinal modulado resultante é de banda estreita.

Em um sistema de modulação digital caótica, as seqüências de base são trechos de comprimento N de sinais caóticos com condições iniciais que variam a cada símbolo. Como esses sinais são aperiódicos, um mesmo símbolo é quase sempre representado por seqüências diferentes a cada vez que for transmitido. Esse fato propicia vantagem quanto ao sigilo de informação à custa de dificultar a estimação das estatísticas de longo prazo a partir de seqüências finitas. Por exemplo, a energia utilizada para transmitir um dado símbolo não é constante, o que leva a erros de recepção mesmo na ausência de ruído, caso essa energia precise ser estimada no receptor para a demodulação.

O diagrama de blocos de um sistema de comunicação digital caótico é mostrado na Figura 1.3 para o caso em que a base é constituída por um único sinal. Os componentes deste diagrama serão analisados em detalhes nos Capítulos 3 e 4.

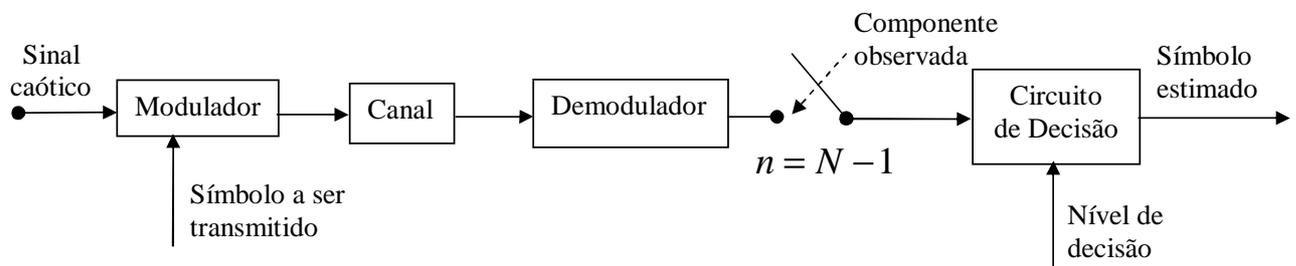


Figura 1.3: Diagrama de blocos de um sistema de comunicação digital caótico com uma única função de base.

Segundo (KENNEDY; ROVATTI; SETTI, 2000; LATHI, 1998) uma modulação digital caótica pode ser considerada como uma modulação por espalhamento espectral (SS - *Spread Spectrum*). Pela definição desses autores, um sistema SS deve satisfazer dois critérios:

1. a largura de banda do sinal transmitido é muito maior do que a mínima necessária para a quantidade de informação sendo transmitida;
2. esta largura de banda é determinada por um *signal de espalhamento* que independe da informação.

Seguindo essa definição, as modulações digitais caóticas são sistemas SS em que sinais caóticos

são responsáveis pelo espalhamento. Porém, referências como (HAYKIN, 2000) e (SKLAR, 2001) incluem uma terceira condição para classificar uma modulação como SS:

3. No receptor, a recuperação dos dados originais é realizada por correlação do sinal espalhado recebido com uma réplica sincronizada do sinal de espalhamento.

Sob essa restrição, apenas os sistemas de modulação caóticos com receptores coerentes podem ser considerados como de espalhamento espectral.

De qualquer forma, pela sua característica de banda larga, os sistemas que utilizam modulação caótica tendem a compartilhar as mesmas propriedades dos sistemas SS convencionais, ou seja, dificuldade de detecção não-autorizada, mitigação do desvanecimento devido ao multipercorso e robustez a interferências de banda estreita (HAYKIN, 2000; LAU; TSE, 2003).

Visando usufruir dessas vantagens, muitos sistemas de modulação de informação digital que utilizam sinais caóticos têm sido propostos. Um dos métodos mais estudados é o chamado *chaveamento caótico* (CSK - *Chaos Shift Keying*) em que símbolos diferentes são mapeados em trechos de sinal caótico com energias médias diferentes ou gerados por sistemas diferentes (DEDIEU; KENNEDY; HASLER, 1993).

No CSK, a detecção no receptor pode usar ou não sincronização caótica. No primeiro caso, conhecido como *detecção coerente*, considera-se que estão disponíveis no receptor cópias sincronizadas dos sinais da base utilizada no transmissor. A detecção pode ser feita avaliando-se o erro de sincronismo (PARLITZ et al., 1992; DEDIEU; KENNEDY; HASLER, 1993) ou por meio de correlação convencional (KOLUMBÁN; KENNEDY; CHUA, 1998).

Por analogia aos sistemas com portadoras senoidais que dispensam a detecção de fase, os sistemas com portadoras caóticas que desnecessitam sincronização caótica são chamados *não-coerentes*. Nesse caso, a demodulação é feita a partir de propriedades usadas para diferenciar os sinais caóticos, como o valor médio, a energia ou a dinâmica do sistema gerador de cada um dos sinais (KENNEDY; KOLUMBÁN, 2000).

Outra técnica de modulação bastante discutida na literatura é o *chaveamento caótico diferencial* (DCSK - *Differential Chaos Shift Keying*) (KOLUMBÁN; KENNEDY; CHUA, 1998) que emprega recepção diferencial. Essa técnica dá uma estrutura especial ao sinal transmitido correspondente a um dado símbolo; no caso binário cada símbolo DCSK é representado por

dois trechos de sinais caóticos. O primeiro trecho é o de *referência* e o segundo é o de *informação*. Dependendo do símbolo transmitido, o trecho de informação é uma cópia exata ou com sinal invertido do sinal de referência.

Existe uma série de outros métodos de comunicação digital sugeridos usando sinais caóticos baseados no CSK ou no DCSK. Alguns estão citados na Tabela 1.1.

Tabela 1.1: Algumas modulações digitais caóticas baseadas no CSK e no DCSK.

Sistema	Referência
Chaveamento liga-desliga caótico (COOK - <i>Chaotic On-Off-Keying</i>)	(KENNEDY; KOLUMBÁN, 2000)
DCSK com frequência modulada (FM-DCSK - <i>Frequency Modulated DCSK</i>)	(KOLUMBÁN et al., 1998)
Chaveamento com atraso e correlação (CDSK - <i>Correlation Delay Shift Keying</i>)	(SUSHCHIK; TSIMRING; VOLKOVSKII, 2000)
Chaveamento caótico simétrico (SCSK - <i>Symmetric CSK</i>)	(SUSHCHIK; TSIMRING; VOLKOVSKII, 2000)
CSK em quadratura (QCSK - <i>Quadrature CSK</i>)	(GALIAS; MAGGIO, 2001)
CSK com estimação de máxima verossimilhança (ML-CSK - <i>Maximum Likelihood CSK</i>)	(KISEL; DEDIEU; SCHIMMING, 2001)
DCSK modificado (M-DCSK - <i>Modified DCSK</i>)	(MANDAL; BANERJEE, 2004)

Além disso, alguns trabalhos propõem o controle da dinâmica simbólica de um sistema caótico de forma a transmitir informações digitais. Podem ser citados como exemplos (HAYES; GREBOGI; OTT, 1993; HAYES et al., 1994). Bollt (2003) faz uma extensa lista de referências a respeito.

1.3 Algumas linhas de pesquisa atuais

Dentre os três temas descritos na seção anterior, esta tese concentra-se em sistemas de modulação digital. A escolha foi feita por ser esse trabalho uma evolução da pesquisa apresentada na dissertação de mestrado do autor na qual aborda-se principalmente sistemas de comunicação analógica usando caos (EISENCRAFT, 2001).

Apesar das várias possíveis aplicações de sinais caóticos em Telecomunicações, ainda existem limitações que precisam ser melhor pesquisadas para possibilitar uma implementação prática. Algumas delas são listadas a seguir.

- I. **Determinação da largura de banda.** O projeto de sistemas de comunicação requer que se conheça a largura de banda dos sinais envolvidos. No entanto, ainda inexiste uma forma de se obter teoricamente a largura de banda de um sinal vinculado a um sistema caótico, tendo-se exclusivamente simulações numéricas como alternativa prática.
- II. **Sincronização caótica.** Sistemas coerentes requerem cópias sincronizadas dos sinais da base no receptor para fazer a demodulação. Assim, é necessário um nível de sincronismo bom entre transmissor e receptor. Como os sinais caóticos apresentam a dependência sensível às condições iniciais, esse problema é difícil e as soluções que existem funcionam de forma precária na presença de relações sinal-ruído baixas ou quando há distorções no canal, como mostram as simulações computacionais apresentadas em (EISENCRAFT; GERKEN, 2000; EISENCRAFT, 2001).
- III. **Avaliação em condições não-ideais.** Preocupação constante no projeto de sistemas de comunicação é a influência de condições não-ideais do canal no desempenho do sistema. Porém, no caso de modulação caótica, apenas problemas relacionados a propagação multipercursos foram estudados para alguns sistemas e por meio de simulações numéricas (KENNEDY et al., 2000). Certamente, nesse tópico ainda há muita pesquisa a ser feita.
- IV. **Integração com a tecnologia existente.** Um aspecto sempre relevante no desenvolvimento de uma nova tecnologia que objetiva tornar-se prática é a possibilidade de integração com os sistemas já existentes. Nesse sentido, têm surgido trabalhos sobre a

compatibilidade e a coexistência entre sistemas de comunicação baseada em caos e os convencionais. Por exemplo, os Capítulos 7 e 8 de (LAU; TSE, 2003) e (LAU; TSE; HAU, 2004).

- V. **Adequação e uniformidade de notação.** A notação empregada nos trabalhos que descrevem sistemas de modulação digital caótica ainda é confusa em relação à empregada usualmente em Engenharia de Telecomunicações. Há pouca preocupação em tratar os sistemas de modulação utilizando portadoras caóticas como caso particular de sistemas de modulação digital. São exceções (KOLUMBÁN; KENNEDY; CHUA, 1997, 1998; KENNEDY; ROVATTI; SETTI, 2000). Porém esses autores trabalham com notação de tempo contínuo, pouco adequada quando se estudam mapas de tempo discreto.
- VI. **Definição dos mapas a serem utilizados.** Um ponto importante é a escolha do mapa caótico a ser utilizado. Cada mapa gera sinais caóticos com características diferentes. Sendo assim, a questão de como escolher o mapa gerador a ser empregado num dado sistema pode ser importante. Apesar disso, esse assunto foi muito pouco explorado na literatura. Normalmente são utilizados sem maiores discussões, os mapas lineares por partes, seja por sua simplicidade ou por possuírem densidade invariante uniforme.
- VII. **Exploração mais intensa das propriedades da dinâmica caótica.** A detecção não-coerente permite uma gama muito grande de possibilidades, das quais boa parte ainda não foi explorada. Até o momento, o FM-DCSK é o sistema mais promissor em termos de taxas de erro de símbolo. No entanto, é importante notar que no FM-DCSK, como proposto originalmente (KOLUMBÁN et al., 1998), não se explora explicitamente as propriedades dos sinais caóticos, ou seja, seu desempenho seria basicamente o mesmo se fosse utilizado um sinal aleatório. É de se esperar que um sistema, cujo receptor aproveite melhor as propriedades estatísticas das órbitas do mapa utilizado, tenha melhor desempenho. Uma idéia nesse sentido é descrita em (KISEL; DEDIEU; SCHIMMING, 2001). Porém, neste artigo permanecem várias questões em aberto.

1.4 Objetivos

Nesta tese busca-se reduzir as limitações relacionadas à notação, definição dos mapas a serem utilizados e exploração das propriedades da dinâmica caótica em sistemas de modulação digital usando caos, correspondentes respectivamente às linhas de pesquisa V, VI e VII comentadas acima. A partir da descrição em tempo discreto de modulações caóticas e de técnicas de estimação de sinais caóticos, procura-se propor sistemas cujos receptores explorem as propriedades da dinâmica caótica permitindo obter melhores desempenhos.

Os objetivos da tese são:

- estudar sistemas de modulação digital por portadoras caóticas e comparar seus desempenhos propondo uma notação de tempo discreto aplicável tanto a sistemas convencionais quanto a sistemas com portadoras caóticas;
- estudar o problema de estimação de órbitas caóticas e de suas condições iniciais propondo limites de desempenho e comparações entre estimadores. Generalizar o estimador de Viterbi deduzido por Dedieu e Kisel (1999) para permitir aplicá-lo a uma classe maior de mapas.
- estudar sistemas de comunicação digital com portadoras caóticas que utilizam estimação da órbita na demodulação propondo e avaliando sistemas que possam empregar uma classe maior de mapas ou recepção diferencial.

Este trabalho envolve conhecimentos e resultados de três áreas diferentes: sistemas dinâmicos não-lineares, sistemas de comunicações e teoria de estimação. São incluídos os resultados necessários de cada uma dessas vertentes, procurando torná-lo acessível a um número maior de interessados.

1.5 Organização da tese

A tese está estruturada em sete capítulos.

No Capítulo 2 são apresentados definições e resultados da teoria de sistemas dinâmicos empregados no restante da tese. A intenção é estabelecer a notação utilizada e tornar o trabalho acessível a um público maior.

No Capítulo 3 são discutidos fundamentos de comunicações digitais estabelecendo uma base teórica comum para modulações digitais convencionais com portadoras senoidais e modulações caóticas.

No Capítulo 4 são descritos em detalhes, à luz da teoria de sistemas de comunicações, os sistemas CSK, DCSK e alguns de seus derivados. O desempenho desses sistemas é analisado teórica e numericamente por meio de simulações computacionais, quando eles são submetidos a canais nos quais é adicionado ruído branco gaussiano.

O Capítulo 5 trata do problema de estimação de condições iniciais e órbitas de sistemas dinâmicos discretos unidimensionais. São deduzidos limites de desempenho para estimadores não-enviesados e, em seguida, são analisados dois algoritmos de estimação, o MLE e o algoritmo de Viterbi. Ao final do capítulo são utilizadas simulações computacionais para compará-los.

No Capítulo 6, são estudados sistemas de modulação digital usando portadoras caóticas que utilizam os métodos de estimação descritos no capítulo anterior. Sistemas baseados no ML-CSK proposto em (KISEL; DEDIEU; SCHIMMING, 2001) e um sistema novo, proposto pelo autor, o ML-DCSK são analisados por meio de simulações computacionais.

Finalmente, no Capítulo 7, são relacionadas as contribuições, apresentadas as conclusões e sugeridos temas para trabalhos futuros.

Capítulo 2

Sinais caóticos em tempo discreto: natureza e caracterização

Este capítulo define e comenta os principais conceitos relacionados a sistemas dinâmicos não-lineares e caos empregados nos capítulos subsequentes. Como a literatura contém vários textos sobre o assunto (e.g. (ALLIGOOD; SAUER; YORK, 1996; DEVANEY, 2003; FERRARA; PRADO, 1994; GUCKENHEIM; HOLMES, 2002; MONTEIRO, 2002; STROGATZ, 1998)), tendo em vista limitar o texto ao necessário, são discutidos apenas conceitos referentes a sistemas unidimensionais de tempo discreto. As extensões para os casos multidimensionais e de tempo contínuo podem ser encontradas nas referências citadas.

Na Seção 2.1 define-se a notação utilizada na representação de mapas e sinais. Na Seção 2.2 são comentados os diversos tipos de estabilidade relacionados ao estudo de órbitas caóticas. A seguir, na Seção 2.3, são formalizados os conceitos de dependência sensível às condições iniciais e número de Lyapunov. A Seção 2.4 apresenta a definição de sinais caóticos e a Seção 2.5 exemplifica alguns mapas capazes de gerá-los. Por fim, a Seção 2.6 expõe os conceitos de densidade invariante e operador de Frobenius-Perron e a Seção 2.7 resume o capítulo.

2.1 Conceitos e definições básicas

Informalmente, um sistema dinâmico é um conjunto de possíveis estados com uma regra determinística que define o estado presente em função dos anteriores. Sistemas dinâmicos de tempo discreto são conhecidos por várias denominações diferentes como *relações recursivas*, *mapas iterados* ou simplesmente *mapas*. Apresenta-se a seguir diversas definições e notações usadas neste trabalho, baseadas na referência (ALLIGOOD; SAUER; YORK, 1996).

Definição 1. Seja $f(\cdot)$ uma função cujo espaço de partida $U \subset \mathbb{R}$ (domínio) é igual ao espaço de chegada (contra-domínio) e $s(0) \in U$. A equação de diferenças $s(n+1) = f(s(n))$, com $n \in \mathbb{N}$ representa um *sistema dinâmico de tempo discreto* ou *mapa*.

Definição 2. A *órbita* ou *senal* correspondente a $s(0)$ gerado a partir de $f(\cdot)$ e $s(0) \equiv s_0$ fixo é a função $s(n) = f^n(s_0)$, em que $f^n(\cdot)$ representa a n -ésima aplicação sucessiva de $f(\cdot)$ e $f^0(s_0) = s_0$. O ponto s_0 é chamado de *condição inicial* de $s(n)$. Quando for importante destacar a condição inicial s_0 de uma órbita, ela será indicada por $s(n, s_0)$.

Definição 3. A *trajetória* da órbita $s(n, s_0)$ de $f(\cdot)$ é o conjunto $S_{s_0} = \{s(n, s_0), n \in \mathbb{N}\}$.

Definição 4. Um ponto c é dito *ponto fixo* de $f(\cdot)$ ou do mapa $s(n+1) = f(s(n))$ se $f(c) = c$ e, portanto, $s(n, c) = c, \forall n \in \mathbb{N}$.

Como a periodicidade de uma órbita e conceitos relacionados são muito importantes no estudo de sinais caóticos, pois a ausência de período é uma das principais características desses sinais, definem-se a seguir esses termos.

Definição 5. Seja o mapa $s(n+1) = f(s(n))$ e p um ponto do domínio de $f(\cdot)$. O ponto p é dito *eventualmente periódico* de período $k > 0$ se para algum $n_* \in \mathbb{N}$, $f^{n+k}(p) = f^n(p)$ para todo $n \geq n_*$ e k é o menor natural que satisfaz essa condição. A órbita $s(n, p)$ é dita *eventualmente periódica* de período k . No caso em que $n_* = 0$, p é dito *periódico* e $s(n, p)$ é *periódica* de período k .

Definição 6. Seja $s(n+1) = f(s(n))$, com $f(\cdot)$ contínua e diferenciável. Uma órbita $s(n, a)$ é dita *assintoticamente periódica* se convergir para uma órbita periódica quando $n \rightarrow \infty$; ou

seja, existe uma órbita periódica $s(n, p)$ tal que $\lim_{n \rightarrow \infty} |s(n, a) - s(n, p)| = 0$. O ponto a é dito *assintoticamente periódico*. Uma órbita (respectivamente, ponto) não assintoticamente periódica (periódico) é chamada de *aperiódica* (*aperiódico*).

2.2 Estabilidade

Os conceitos relacionados à estabilidade são fundamentais para a análise do comportamento de sistemas dinâmicos. Seja na implementação física ou nas simulações computacionais, não é possível obter precisão absoluta na geração de órbitas, mesmo fixadas as condições iniciais e os parâmetros que as definem. A análise da estabilidade, em suas diversas definições, garante se um comportamento teórico esperado poderá ser observado ou não.

São tratados a seguir a estabilidade segundo Lagrange, a estabilidade para órbitas estacionárias e a estabilidade para órbitas genéricas. Encerrando a seção, os conceitos de conjunto ω -limite e estabilidade estrutural de um sistema são abordados.

2.2.1 Estabilidade no sentido de Lagrange

Uma órbita $s(n)$ apresenta *estabilidade limitada de acordo com Lagrange* se $s(n)$ é limitada (MONTEIRO, 2002), ou seja, existe uma constante C tal que $|s(n)| < C$ para todo $n \in \mathbb{N}$.

2.2.2 Estabilidade de órbitas estacionárias

Uma órbita é dita estacionária se é um ponto fixo ou uma órbita periódica. As definições a seguir são baseadas nas referências (GUCKENHEIM; HOLMES, 2002; STROGATZ, 1998).

Definição 7. Um ponto fixo c de um mapa $s(n+1) = f(s(n))$ é:

1. *atrator* se existe $\delta > 0$ tal que, para qualquer órbita $s(n, s_0)$ com $|s_0 - c| < \delta$, $\lim_{n \rightarrow \infty} s(n, s_0) = c$;
2. *estável segundo Lyapunov* se, dado $\varepsilon > 0$, existe um $\delta > 0$ tal que se $|s_0 - c| < \delta$ então $|s(n, s_0) - c| < \varepsilon$ para todo $n \geq 0$;

3. *assintoticamente estável* se é atrator e estável segundo Lyapunov;
4. *instável* se não é atrator e nem estável segundo Lyapunov.

As trajetórias que se iniciam próximas de um atrator podem sair de suas proximidades a curto prazo, porém, devem aproximar-se dele a longo prazo. As trajetórias que se iniciam próximas de um ponto fixo estável segundo Lyapunov devem permanecer nas proximidades desse ponto para todo $n \geq 0$, não convergindo, necessariamente, para ele. Essas situações são mostradas de forma esquemática nas Figuras 2.1(a) e (b). As trajetórias próximas de um ponto fixo assintoticamente estável além de convergirem para ele, permanecem o tempo todo numa vizinhança dele.

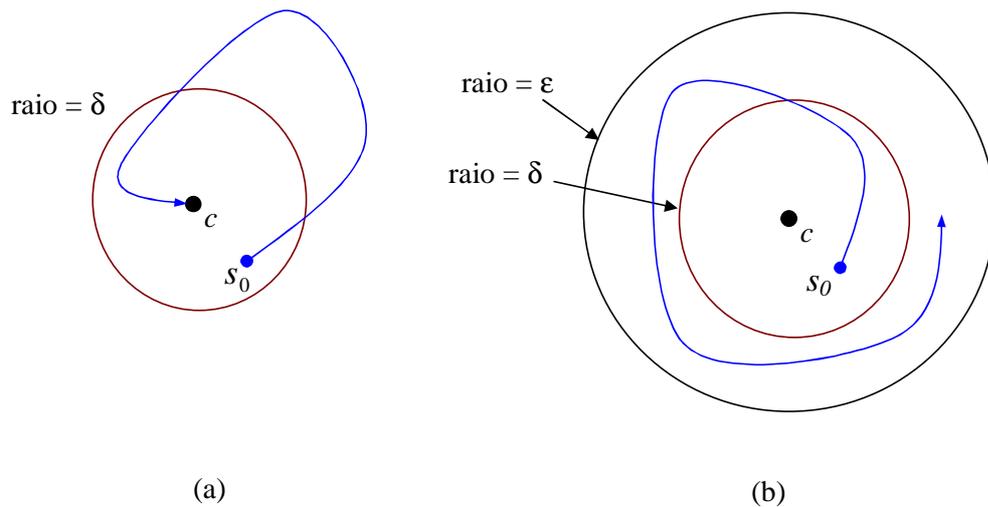


Figura 2.1: Diagrama esquemático do comportamento de uma órbita próxima (a) de um ponto fixo atrator (b) de um ponto fixo estável segundo Lyapunov.

Se uma órbita $s(n, p)$ é periódica de período k para o mapa $s(n + 1) = f(s(n))$, p é um ponto fixo do mapa $s(n + 1) = f^k(s(n))$. Assim, a análise da estabilidade da órbita periódica $s(n, p)$ é feita através da análise da estabilidade do ponto fixo p do sistema $s(n + 1) = f^k(s(n))$.

2.2.3 Estabilidade de órbitas genéricas

A estabilidade orbital generaliza as definições da subseção anterior para um caso qualquer. Evidentemente, a estabilidade de órbita estacionária é um caso particular. Dentre as várias definições possíveis, neste trabalho é utilizada a *estabilidade orbital segundo Lyapunov*, da forma como estabelecida por Cesari (1971), adaptada aqui para o caso discreto.

Definição 8. Seja a órbita $s(n, s_0)$ do mapa $s(n+1) = f(s(n))$ e A um intervalo real aberto tal que $S_{s_0} \subset A$. Esta órbita é dita *estável segundo Lyapunov* se:

1. existe um $\delta > 0$ tal que se $|r_0 - s_0| \leq \delta$ então $S_{r_0} \subset A$;
2. dado $\varepsilon > 0$, existe um $\delta_1, 0 < \delta_1 \leq \delta$, tal que se $|r_0 - s_0| \leq \delta_1$, então $|s(n, s_0) - s(n, r_0)| \leq \varepsilon$ para todo $n \geq 0$.

A órbita $s(n, s_0)$ é dita *assintoticamente estável segundo Lyapunov* se 1. e 2. são satisfeitas e

3. existe um $\delta_2, 0 < \delta_2 \leq \delta$ tal que $|r_0 - s_0| \leq \delta_2$ implica em $\lim_{n \rightarrow \infty} |s(n, s_0) - s(n, r_0)| = 0$.

Deve-se ressaltar que essas definições de órbitas estáveis (segundo Lyapunov) são bastante restritivas deixando de fora muitas órbitas que a intuição leva a chamar de estáveis. Por exemplo, órbitas que percorram pontos muito parecidos mas em instantes diferente de suas vizinhas são consideradas instáveis. Além disso, essa definição de estabilidade não é invariante a mudanças de coordenadas. Mais detalhes podem ser vistos em (EISENCRAFT, 2001).

Pode-se definir a estabilidade de uma órbita de uma maneira mais genérica de forma que ela se torne independente da escolha das coordenadas que a descrevem. Por exemplo, definindo-se a distância de um ponto s até S_{s_0} como

$$d(s, S_{s_0}) = \inf_{n \in \mathbb{N}} \{|s - s(n, s_0)|\} \quad (2.1)$$

e o tubo de raio ε da órbita $s(n, s_0)$ como

$$T(\varepsilon, S_{s_0}) = \{s \in U / d(s, S_{s_0}) < \varepsilon\} \quad (2.2)$$

pode-se tomar as seguintes definições.

Definição 9. A órbita $s(n, s_0)$ é *estável* se dado $\varepsilon > 0$, existe um δ tal que se $|r_0 - s_0| < \delta$ então $s(n, r_0) \in T(\varepsilon, S_{s_0})$, para todo $n \geq 0$. Se, além disso, existe δ_1 com $0 < \delta_1 < \delta$ tal que se $|r_0 - s_0| < \delta_1$ então $\lim_{n \rightarrow \infty} d(s(n, r_0), S_{s_0}) = 0$, a órbita é *assintoticamente estável*.

Com essas definições, a estabilidade deixa de depender da velocidade e da parametrização (EISENCRAFT, 2001).

Nas próximas seções e capítulos, a estabilidade orbital sempre se refere à estabilidade segundo Lyapunov.

2.2.4 Conjuntos limites

O conjunto limite de uma órbita é o conjunto de pontos cuja vizinhança é freqüentemente visitada por ela. Formalmente, tem-se a definição a seguir.

Definição 10. O *conjunto limite* da órbita $s(n, s_0)$ de $f(\cdot)$ é dado por

$$\omega(s_0) = \{s \in U / \forall \{n_*, \varepsilon\}, \exists n > n_*, |f^n(s_0) - s| < \varepsilon\}. \quad (2.3)$$

Se $\omega(s_0)$ é o conjunto limite de um órbita e r_0 é uma outra condição inicial, então diz-se que a órbita $s(n, r_0)$ (ou o ponto r_0) é *atraído* para $\omega(s_0)$ se $\omega(r_0) \subset \omega(s_0)$.

Pontos de uma órbita podem ou não estar contidos em seu conjunto limite. O conjunto limite pode não ter nenhum ponto em comum com a órbita, como no caso do conjunto limite de uma órbita convergindo assintoticamente para um ponto fixo atrator. Nesse caso, o conjunto limite é um único ponto, o ponto fixo atrator. A órbita é atraída para esse ponto fixo.

Definição 11. Um *atrator* é um conjunto limite que atrai um conjunto de condições iniciais de medida não nula. Esse conjunto de condições iniciais é chamado de *bacia de atração*.

Um conjunto $A \subset B$ é *denso* em B se arbitrariamente próximo de cada ponto de B existe um ponto de A . Em outras palavras, para cada s_* em B e cada $\varepsilon > 0$, a vizinhança $\{s \in U / |s_* - s| < \varepsilon\}$ contém um ponto de A . Nas próximas seções encontram-se órbitas que são densas no intervalo $[-1, 1]$ o que significa que seu conjunto limite é $[-1, 1]$.

2.2.5 Estabilidade estrutural

A *estabilidade estrutural* é conceitualmente diferente das definições de estabilidade vistas anteriormente. Enquanto nessas últimas a estabilidade é investigada perturbando-se as condições

iniciais, para a estabilidade estrutural interessa a robustez das órbitas sob uma perturbação do mapa.

Seja um mapa $f(\cdot)$ com derivada $f'(\cdot)$ definido num espaço de funções contínuas. Um mapa $g(\cdot)$ é uma perturbação de tamanho ε do mapa $f(\cdot)$ se ele pertence a uma vizinhança de raio ε de $f(\cdot)$ e se sua derivada $g'(\cdot)$ pertence a uma vizinhança de raio ε de $f'(\cdot)$ (FERRARA; PRADO, 1994).

Definição 12. Um sistema é denominado *estruturalmente estável* se para qualquer perturbação suficientemente pequena do mapa que o define, o conjunto das órbitas resultantes é topologicamente equivalente¹ àquele das equações sem a perturbação.

Uma definição mais detalhada desse conceito juntamente com as condições necessárias para o seu aparecimento pode ser encontrada, por exemplo, em (PALIS; MELO, 1982).

Em todas as aplicações de mapas dos próximos capítulos sempre se considera que os parâmetros escolhidos na definição do mapa $f(\cdot)$ são tais que garantem estabilidade estrutural dos sistemas.

2.3 Número e expoente de Lyapunov

Um conceito fundamental para o estudo de sinais caóticos é a *dependência sensível às condições iniciais* (DCI) definida a seguir segundo (ALLIGOOD; SAUER; YORK, 1996).

Definição 13. Seja $s(n+1) = f(s(n))$ um mapa em $U \subset \mathbb{R}$. Um ponto $s_0 \in U$ tem *dependência sensível às condições iniciais* (DCI) se existe uma distância não-nula ε tal que pelo menos algum ponto arbitrariamente próximo de s_0 é eventualmente mapeado a pelo menos ε unidades da imagem correspondente a s_0 . Mais precisamente, existe $\varepsilon > 0$ tal que qualquer vizinhança $\{s \in U / |s - s_0| < \delta\}$ de s_0 contém pelo menos um ponto s_* tal que $|f^{n_*}(s_*) - f^{n_*}(s_0)| \geq \varepsilon$ para algum $n_* \in \mathbb{N}$. Nessas condições a órbita $s(n, s_0)$ tem DCI.

Valendo-se apenas dessa definição é muito difícil verificar essa propriedade para uma dada órbita, exceto em casos excepcionais.

¹Dois conjuntos de órbitas são topologicamente equivalentes se existe um homeomorfismo (função contínua com inversa contínua) que leva um no outro preservando o sentido temporal das trajetórias. Mais detalhes podem ser obtidos, por exemplo, no Capítulo 1 de (PALIS; MELO, 1978).

A seguir define-se o expoente de Lyapunov que representa uma taxa de divergência exponencial média entre duas órbitas suficientemente próximas e que possibilita a verificação da DCI de forma operacional. Daí a importância do estudo de sua definição e de métodos numéricos para o seu cálculo. Novamente, este trabalho limita-se ao caso discreto unidimensional. As extensões dessas definições para os casos contínuos e multidimensionais são detalhadas, por exemplo, em (EISENCRAFT, 2001).

O comportamento das órbitas do sistema dinâmico de tempo discreto $s(n+1) = f(s(n))$ nas proximidades de um ponto fixo é fortemente influenciado pela derivada de $f(\cdot)$ nesse ponto, supondo que ela exista. Por exemplo, se c é um ponto fixo de $f(\cdot)$ e o módulo da derivada $|f'(c)| = l > 1$, então toda órbita $s(n, s_0)$ com s_0 próximo de c afasta-se de c com uma *taxa multiplicativa de aproximadamente l por iteração*, até mover-se para longe de c . Isto pode ser visto na expansão em série de Taylor de $f(s_0)$ em torno de c :

$$\begin{aligned} f(s_0) &\approx f(c) + f'(c)(s_0 - c) \Rightarrow f(s_0) - f(c) \approx f'(c)(s_0 - c) \Rightarrow \\ &\Rightarrow |f(s_0) - c| \approx l |s_0 - c|. \end{aligned} \quad (2.4)$$

Para um ponto periódico p de período k , essa informação pode ser obtida olhando-se a derivada da k -ésima iteração de $f(\cdot)$ já que p é ponto fixo de $f^k(\cdot)$. Pela regra da cadeia,

$$\begin{aligned} f^{k'}(p) &= f'(f^{k-1}(p)) \cdot f'(f^{k-2}(p)) \cdot f'(f^{k-3}(p)) \cdot \dots \cdot f'(p) \Rightarrow \\ f^{k'}(p) &= \prod_{n=0}^{k-1} f'(s(n, p)), \end{aligned} \quad (2.5)$$

ou seja, $f^{k'}(p)$ é o produto das derivadas de $f(\cdot)$ calculadas nos k pontos distintos de S_p . Suponha que o módulo desse produto seja $l > 1$. Pelo mesmo raciocínio utilizado para pontos fixos, uma órbita com condição inicial próxima de p separa-se de $s(n, p)$ a uma taxa de aproximadamente l depois de k iterações. Faz sentido então descrever a *taxa multiplicativa média de separação por iteração entre as duas órbitas como sendo $L = l^{\frac{1}{k}}$* .

O número de Lyapunov generaliza as taxas obtidas acima para o caso em que os pontos não são necessariamente periódicos.

Definição 14. Seja o mapa $s(n+1) = f(s(n))$. Se $f(\cdot)$ for diferenciável nos pontos da trajetória da órbita $s(n, s_0)$, o seu *número de Lyapunov* é

$$L(s_0) = \lim_{N \rightarrow \infty} \left(\prod_{n=0}^{N-1} |f'(s(n))| \right)^{\frac{1}{N}} \quad (2.6)$$

se o limite existir. O *expoente de Lyapunov* $h(s_0)$ é definido como

$$h(s_0) = \ln(L(s_0)), \quad (2.7)$$

se $L(s_0)$ existir.

Segue da definição que o número de Lyapunov de um ponto fixo c do mapa $s(n+1) = f(s(n))$ é $|f'(c)|$ e o expoente de Lyapunov de um ponto periódico $s_0 = p$ de período k desse mesmo mapa é

$$h(p) = \frac{\sum_{n=0}^{k-1} \ln |f'(s(n, p))|}{k}. \quad (2.8)$$

A Definição 14 permite algumas considerações sobre estabilidade:

- um ponto fixo ou uma órbita periódica que tenha expoente de Lyapunov negativo será assintoticamente estável pelo argumento apresentado em (2.4), já que nesse caso $l < 1$;
- também pela expressão (2.4), conclui-se que expoente de Lyapunov positivo indica que o ponto fixo ou a órbita periódica é instável, pois nesse caso $l > 1$.

Esses fatos são importantes por causa do seguinte teorema, cuja demonstração é apresentada em (ALLIGOOD; SAUER; YORK, 1996):

Teorema 1. Seja o mapa $s(n+1) = f(s(n))$. Se a órbita $s(n, s_0)$ satisfaz $f'(s(n)) \neq 0$ para todo $n \in \mathbb{N}$ e é assintoticamente periódica tendo como limite a órbita periódica $s(n, p)$, então as duas órbitas têm expoentes de Lyapunov idênticos, assumindo que ambos estejam definidos.

Conclui-se assim que se uma órbita limitada possui expoente de Lyapunov positivo ela não pode ser assintoticamente periódica tendo como limite uma órbita periódica assintoticamente estável ou convergir para um ponto fixo assintoticamente estável. Assim, restam duas possibilidades para esse tipo de órbita

- ou ela é assintoticamente periódica convergindo para uma órbita periódica instável, caso pouco provável em geral; ou
- ela é *aperiódica*.

Essa conclusão é fundamental para a classificação de uma órbita como “caótica”.

Finalmente, outro conceito que é importante definir neste ponto é o de *mapas conjugados*. O número de Lyapunov é simples de ser calculado em mapas lineares por partes, já que esses apresentam derivada constante em cada trecho. Porém para mapas mais genéricos, a determinação desse número e das propriedades associadas a ele tornam-se mais difíceis. A conjugação permite deduzir propriedades das órbitas de um mapa a partir da “parecença” dele com outro eventualmente mais simples.

Definição 15. Seja $C(\cdot)$ um mapa contínuo e injetor. Os mapas $f(\cdot)$ e $g(\cdot)$ são *conjugados* se eles estão relacionados por uma mudança de coordenadas $C(\cdot)$ tal que $C \circ f(\cdot) = g \circ C(\cdot)$. A notação $C \circ f(\cdot)$ significa $C(f(\cdot))$ e $C(\cdot)$ é chamado de *mapa de conjugação*.

O mapa $C(\cdot)$ mostra uma correspondência entre pontos dos mapas $f(\cdot)$ e $g(\cdot)$. O seguinte teorema, demonstrado em (ALLIGOOD; SAUER; YORK, 1996) resume a importância da conjugação no estudo de sistemas dinâmicos não-lineares.

Teorema 2. Considere os mapas $f(\cdot)$ e $g(\cdot)$ conjugados por $C(\cdot)$. Pode-se afirmar que:

1. Se p é um ponto periódico de período k de $f(\cdot)$, então $C(p)$ é um ponto periódico de período k de $g(\cdot)$.
2. Se a derivada de $C(\cdot)$ é não-nula em todos os pontos da órbita $s(n, s_0)$ de $f(\cdot)$ então existe uma órbita correspondente $C(s(n, s_0))$ de $g(\cdot)$ com mesmo número de Lyapunov.

Esse teorema expressa que dados dois mapas conjugados, a cada órbita de um corresponde uma órbita do outro com mesmo período e mesmo número de Lyapunov.

2.4 Sinais caóticos

Existem diversas definições de sinais *caóticos* aceitas por diferentes pesquisadores. Aqui é utilizada a definição dada por Li e Yorke (1975) na forma como foi expressa em (ALLIGOOD;

SAUER; YORK, 1996).

Definição 16. Uma órbita limitada é dita *caótica* se é aperiódica e apresenta DCI.

Uma órbita com DCI, pela própria Definição 13, não é estável já que é sempre possível encontrar um ponto arbitrariamente próximo de sua condição inicial cuja órbita relacionada estará distante depois de um certo tempo de simulação. É exatamente aí que reside a “impresvisibilidade” característica dos sistemas caóticos. Como, numa situação prática, nunca se sabe com toda a precisão a condição inicial de uma órbita, é impossível prever o comportamento de um tal sistema após um certo tempo (conhecido como *tempo de horizonte*) dependente da incerteza na condição inicial.

Como foi visto, os expoentes de Lyapunov funcionam como uma quantificação da DCI. O seguinte teorema garante esse fato (ALLIGOOD; SAUER; YORK, 1996).

Teorema 3. Suponha que uma órbita $s(n, s_0)$ tenha número de Lyapunov $L(s_0)$ e expoente de Lyapunov $h(s_0)$. Essa órbita apresentará DCI se, e somente se, $L(s_0)$ for maior do que a unidade, isto é, $h(s_0) > 0$.

Esse teorema é bastante intuitivo considerando-se que os expoentes de Lyapunov medem a divergência exponencial de órbitas próximas. Como existem algoritmos disponíveis para o cálculo numérico dos expoentes de Lyapunov (EISENCRAFT, 2001), tem-se uma maneira mais fácil para determinar se uma órbita é caótica ou não:

Teorema 4. Uma órbita limitada $s(n, s_0)$ é caótica se, e somente se é aperiódica e $L(s_0) > 1$.

A determinação da periodicidade ou não de uma órbita não é simples. Nos procedimentos numéricos usa-se a condição $L(s_0) > 1$ como “suficiente” para considerar uma órbita como caótica, como é adotado em geral.

Outra definição muito usual na literatura é a de *atrator caótico* (ALLIGOOD; SAUER; YORK, 1996).

Definição 17. Seja o sistema dinâmico $s(n+1) = f(s(n))$ e uma órbita caótica $s(n, s_0)$ desse sistema. O conjunto $\omega(s_0)$ dessa órbita é chamado de *conjunto caótico* se $s_0 \in \omega(s_0)$. Um *atrator caótico* é um conjunto caótico que também é um atrator.

Em outras palavras, um conjunto caótico é o conjunto limite de uma órbita caótica que está contida no seu conjunto limite. A exigência de que a órbita caótica esteja contida no seu próprio conjunto limite garante que o conjunto caótico tenha uma órbita densa e que portanto seja irredutível.

2.5 Exemplos de mapas caóticos

Nesta seção são apresentados três exemplos de mapas discretos unidimensionais que geram órbitas caóticas. Inicialmente, é discutido o mapa tenda, linear por partes e com derivada em módulo constante. A seguir, a família de mapas tenda inclinada que também são lineares por partes mas com derivada diferente em cada trecho e por fim o mapa quadrático.

Nos três casos, adota-se como domínio $U = [-1, 1]$.

2.5.1 Mapa tenda $f_T(\cdot)$

Como primeiro exemplo, considere o mapa tenda

$$f_T(s) = 1 - 2|s| \quad (2.9)$$

e o sistema dinâmico

$$s(n+1) = f_T(s(n)) \quad (2.10)$$

com condição inicial $s_0 \in U$. A Figura 2.2 mostra um gráfico de $f_T(s)$ e uma órbita gerada a partir da Eq. (2.10).

Para $s \neq 0$, $|f'_T(s)| = 2$ e portanto $L(s_0) = L_T = 2$ para todas as órbitas que não passam por $s = 0$. O fato de que $L_T > 1$ implica no provável aparecimento de órbitas caóticas nesse sistema. Na verdade, utilizando a teoria de dinâmica simbólica, pode-se mostrar que existe uma quantidade não-enumerável de condições iniciais que geram órbitas caóticas para esse mapa, sendo seu atrator caótico o intervalo U (DEVANEY, 2003).

É importante ressaltar que para qualquer condição inicial s_0 racional, a órbita $s(n, s_0)$ converge para o ponto fixo $c = -1$. Assim, o comportamento caótico das órbitas desse mapa

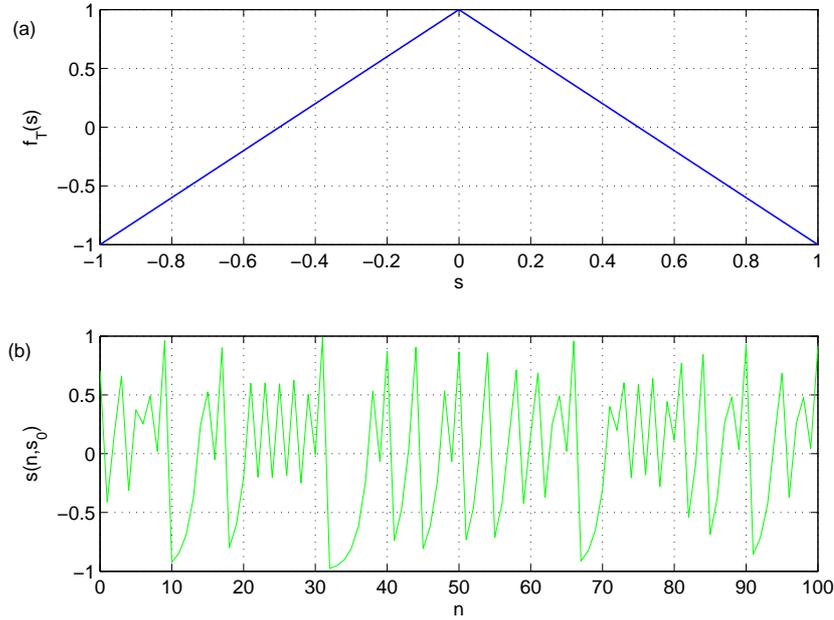


Figura 2.2: (a) Mapa tenda $f_T(s)$; (b) órbita com condição inicial $s_0 = \sqrt{2}/2$.

não pode ser observado na prática por meio de iterações da Eq. (2.10). Neste trabalho, para simular a utilização de condições iniciais irracionais utilizou-se a abordagem de sistemas lineares alimentados por entradas aleatórias descrita por Drake (1998).

2.5.2 Mapas tenda inclinada $f_I(\cdot)$

A família de mapas tenda inclinada é definida por

$$s(n+1) = f_I(s(n)) = \begin{cases} \frac{2}{\alpha+1}s(n) + \frac{1-\alpha}{\alpha+1}, & -1 \leq s(n) < \alpha \\ \frac{2}{\alpha-1}s(n) - \frac{\alpha+1}{\alpha-1}, & \alpha \leq s(n) \leq 1 \end{cases} \quad (2.11)$$

com parâmetro $-1 < \alpha < 1$ e $s(0) \in U$. O caso $\alpha = 0,8$ é mostrado na Figura 2.3 juntamente com a órbita $s(n, 0)$ desse mapa.

O número de Lyapunov $L(s_0)$ para uma órbita $s(n, s_0)$ que não passe por $s = \alpha$ pode ser obtido como uma média geométrica do módulo da derivada dos dois trechos lineares do mapa ponderada pelo comprimento de cada um dos segmentos (KISEL; DEDIEU; SCHIMMING,

2001). Como esse número não depende de s_0 , será indicado somente por L_I .

$$L_I = \left(\frac{2}{\alpha + 1} \right)^{\frac{\alpha+1}{2}} \left(\frac{2}{1 - \alpha} \right)^{\frac{1-\alpha}{2}}. \quad (2.12)$$

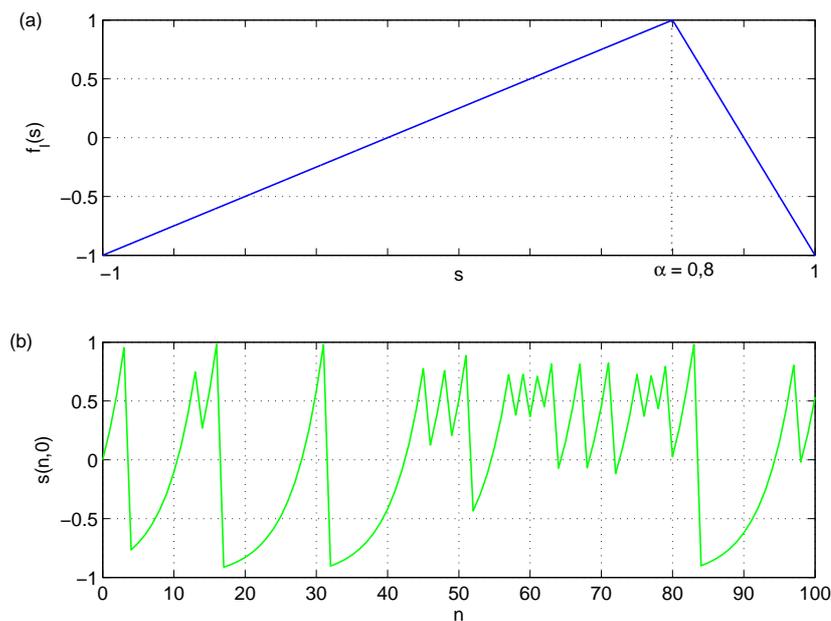


Figura 2.3: (a) Mapa tenda inclinada $f_I(s)$; (b) órbita com condição inicial $s_0 = 0$ para $\alpha = 0,8$.

A Figura 2.4 mostra a variação de L_I com α . Observa-se que, para os valores de α considerados as órbitas desse mapa apresentam $L_I > 1$ e portanto certamente órbitas com comportamento caótico. O valor máximo é $L_{I_{max}} = 2$ e ocorre para $\alpha = 0$ que corresponde ao mapa tenda $f_T(\cdot)$ do exemplo anterior.

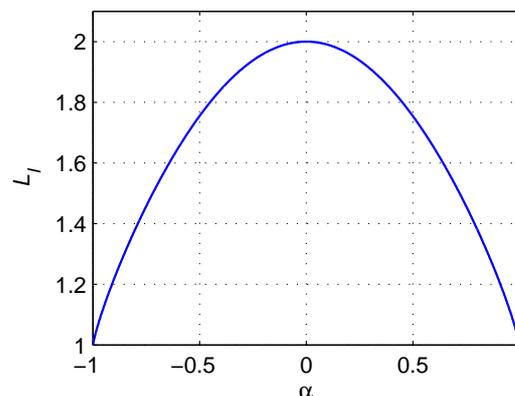


Figura 2.4: Número de Lyapunov de uma órbita do mapa $f_I(\cdot)$ em função do parâmetro α .

Um resultado interessante proveniente da teoria da dinâmica simbólica

(ALLIGOOD; SAUER; YORK, 1996) é que para qualquer $-1 < \alpha < 1$, $f_I(\cdot)$ apresenta órbitas caóticas densas em U sendo esse seu atrator caótico.

2.5.3 Mapa quadrático $f_Q(\cdot)$

Como último exemplo, considera-se o mapa quadrático $f_Q(\cdot)$ e o sistema dinâmico associado definido por

$$s(n+1) = f_Q(s(n)) = -2s^2(n) + 1. \quad (2.13)$$

Um gráfico de $f_Q(\cdot)$ é mostrado na Figura 2.5 juntamente com uma órbita. Este mapa tem dois pontos fixos instáveis, $c_1 = -1$ e $c_2 = 0,5$.

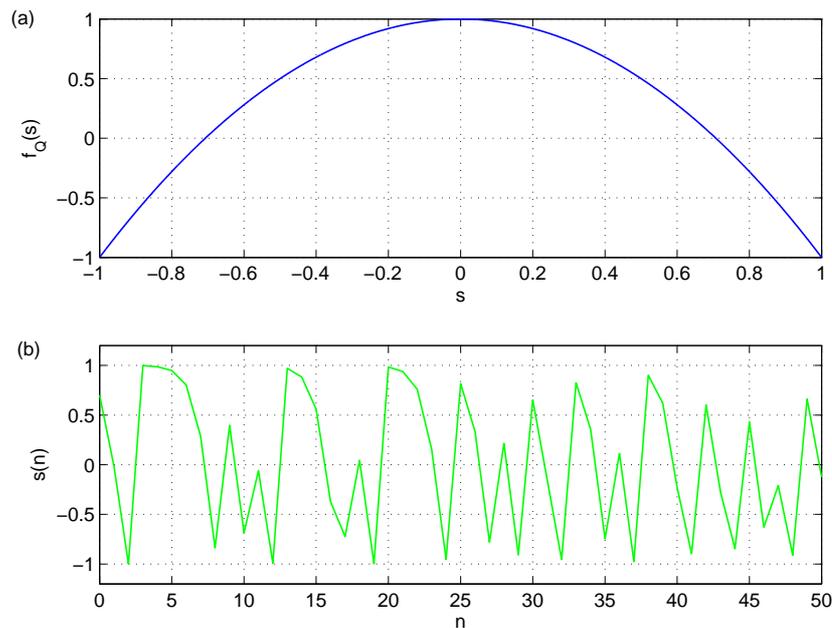


Figura 2.5: (a) Mapa quadrático $f_Q(\cdot)$; (b) órbita com condição inicial $s_0 = 0,7$.

Nesse caso, a derivada de $f_Q(\cdot)$ varia ponto a ponto sendo mais difícil o cálculo do número de Lyapunov e a conseqüente análise da dinâmica das órbitas desse mapa. Para resolver esse problema, exploram-se as similaridades entre o mapa $f_Q(\cdot)$ e o mapa $f_T(\cdot)$, através do conceito de conjugação. Pode-se mostrar que os mapas $f_T(\cdot)$ e $f_Q(\cdot)$ são conjugados com (ALLIGOOD; SAUER; YORK, 1996)

$$C_{TQ}(s) = -\cos \frac{\pi(s+1)}{2}. \quad (2.14)$$

Assim, o Teorema 2 permite afirmar que a cada uma das não-enumeráveis órbitas caóticas

de $f_T(\cdot)$ corresponde uma órbita caótica de $f_Q(\cdot)$ com número de Lyapunov $L_Q = L_T = 2$. É interessante ressaltar que algumas condições iniciais irracionais de $f_T(\cdot)$ são mapeadas em condições iniciais racionais de $f_Q(\cdot)$. Dessa forma, nesse caso, o comportamento caótico pode ser observado diretamente por meio de iterações da Eq. (2.13).

O conjunto atrator de todas as órbitas caóticas de $f_Q(\cdot)$ é novamente U . Como essas órbitas passam arbitrariamente próximas de todos os pontos de U é razoável que seu número de Lyapunov L_Q independa de s_0 da mesma forma como ocorre com as órbitas caóticas dos mapas lineares por partes analisados.

2.6 Densidade invariante e o operador de Frobenius-Perron

O estudo de órbitas caóticas em sistemas discretos unidimensionais pode ser bastante facilitado, estudando-se como seus mapas atuam sobre conjuntos de pontos (densidades) ao invés de sobre uma única condição inicial (LASOTA; MACKEY, 1985; SETTI et al., 2002).

Quando um dado mapa opera sobre uma densidade como condição inicial ao invés de sobre um único ponto, então as densidades sucessivas são dadas por um operador integral conhecido como *operador de Frobenius-Perron*. O objetivo aqui é fornecer uma interpretação intuitiva desse operador.

Órbitas caóticas são muito difíceis de serem caracterizadas individualmente. Considerando-se, por exemplo, o mapa $f_Q(\cdot)$, suas órbitas têm como atrator todo o intervalo U e, devido à DCI, seu comportamento de longo termo é imprevisível caso não se conheça sua condição inicial com precisão absoluta. Uma tentativa nesse sentido é construir um histograma para mostrar a frequência com que os pontos ao longo de uma trajetória caem em dadas regiões do domínio U . Esse histograma é obtido dividindo-se U em N_S intervalos disjuntos de forma que o j -ésimo intervalo, (negligenciando o ponto $+1$) é dado por

$$U_j = \left[\frac{2(j-1)}{N_S} - 1, \frac{2j}{N_S} - 1 \right), \quad j = 1, \dots, N_S. \quad (2.15)$$

A seguir, toma-se uma condição inicial s_0 e calcula-se uma longa trajetória de comprimento $N \gg N_S$. Assim, é imediato estimar a probabilidade $P(s(n) \in U_j)$ como a fração dos N pontos, chamada aqui de P_j , que estão no j -ésimo intervalo

$$P_j \approx \frac{\#\{f^n(s_0) \in U_j, n = 0, 1, \dots, N-1\}}{N}, \quad (2.16)$$

em que $\#A$ representa a cardinalidade do conjunto A .

Os histogramas² da Figura 2.6 mostram o resultado desse procedimento, tomando-se como condições iniciais $s_1(0) = 0,7$ e $s_2(0) = 0,4$, utilizando-se $N_S = 40$ subintervalos e $N = 10000$ pontos. Nesses histogramas, a área do retângulo sobre o intervalo U_j é a aproximação para P_j da Eq. (2.16). Existe uma simetria surpreendente no resultado já que os pontos concentram-se claramente próximos de -1 e 1 com mínimo próximo de $s = 0$, apesar das órbitas serem bastante diferentes. Repetindo o procedimento para outras condições iniciais leva, em geral, ao mesmo resultado. Assim, apesar da DCI das órbitas individualizadas, essa sensibilidade não se reflete na distribuição dos pontos ao longo da trajetória.

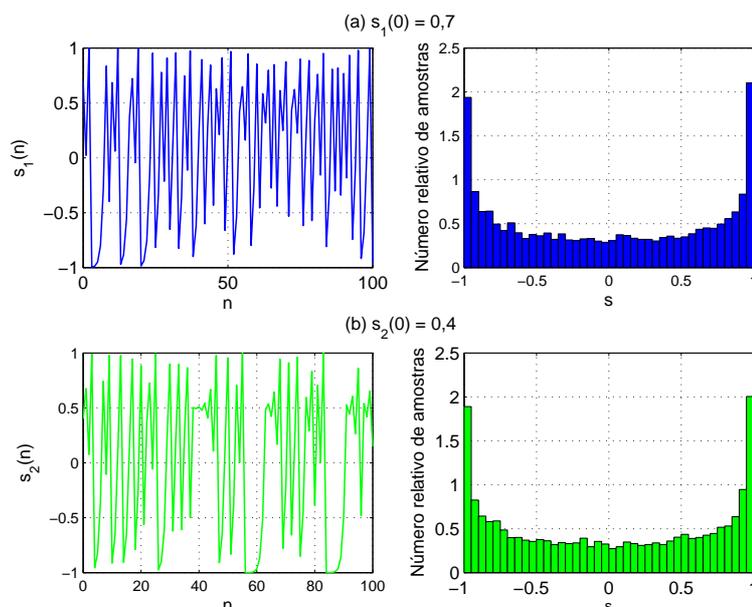


Figura 2.6: Órbitas e histogramas do mapa $f_Q(\cdot)$ obtidos com $N = 10000$ pontos, $N_S = 40$ intervalos e condições iniciais (a) $s_1(0) = 0,7$ e (b) $s_2(0) = 0,4$.

²Um histograma é uma aproximação da função densidade de probabilidades de uma variável aleatória em que o conjunto de valores que ela pode assumir é dividido em intervalos de mesmo comprimento e em que a área de cada barra representa a fração de valores observados naquele intervalo. Assim como ocorre com a função densidade, a soma da área dos retângulos do histograma é igual à unidade.

Note-se que para algumas condições iniciais selecionadas, comportamentos diferentes podem ocorrer. Por exemplo, algumas condições iniciais são atraídas para um dos pontos fixos de $f_Q(\cdot)$. Quando isso ocorre a trajetória assume um valor constante (c_1 ou c_2) a partir de certo n como ocorre na Figura 2.7(a). Alternativamente, para algumas outras condições iniciais irracionais a trajetória pode se tornar periódica e também não exibir o comportamento irregular das órbitas da Figura 2.6. Essa situação é ilustrada na Figura 2.7(b).

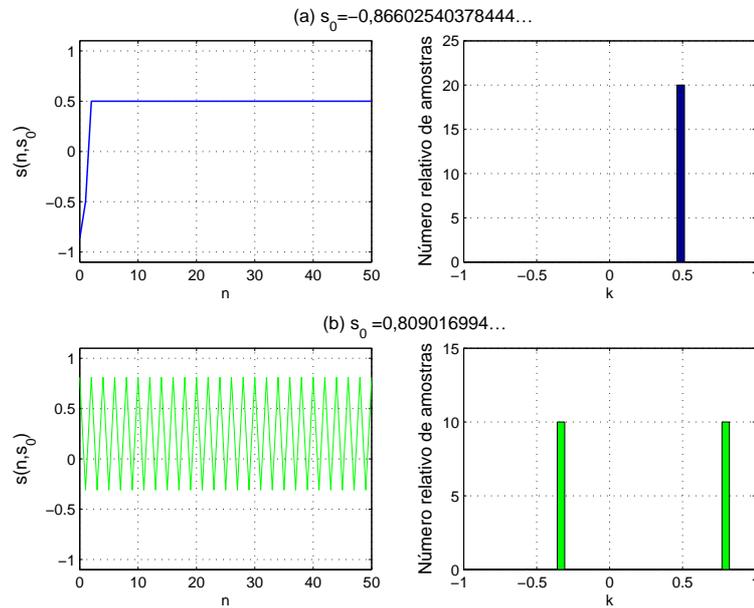


Figura 2.7: Órbitas e histogramas de $f_Q(\cdot)$ sem comportamento caótico. (a) $s_0 = -\sqrt{\frac{3}{4}}$; (b) $s_0 \approx 0,809016994$. Foram usadas $N = 10000$ amostras e $N_S = 40$ subintervalos.

O problema da descrição dos comportamentos esperados para órbitas de um mapa pode ser parcialmente resolvido abandonando-se o estudo de órbitas individuais em favor do exame do fluxo de densidades.

Considere um mapa $f(\cdot)$ definido em U e tome-se um grande número K de condições iniciais $s_{01}, s_{02}, \dots, s_{0K}$. A cada uma dessas condições iniciais aplica-se o mapa $f(\cdot)$, obtendo-se K novos pontos $s(1, s_{01}) \equiv s_{11}, s(1, s_{02}) \equiv s_{12}, \dots, s(1, s_{0K}) \equiv s_{1K}$. Para definir a densidade inicial e final desses pontos, é interessante introduzir o conceito de *função indicadora* de um conjunto A . Ela é simplesmente definida por

$$I_A(s) = \begin{cases} 1, & \text{se } s \in A \\ 0, & \text{se } s \notin A \end{cases} \quad (2.17)$$

Considere-se um intervalo $A_0 \subset U$ e um conjunto de K condições iniciais $s_{01}, s_{02}, \dots, s_{0K}$. Pode-se estimar a função densidade desses pontos para qualquer A_0 pela relação (LASOTA; MACKEY, 1985)

$$\int_{A_0} p_0(u) du \approx \frac{1}{K} \sum_{k=1}^K I_{A_0}(s_{0k}). \quad (2.18)$$

Assim se a função densidade $p_1(s)$ dos pontos $s_{11}, s_{12}, \dots, s_{1K}$ satisfaz

$$\int_A p_1(u) du \approx \frac{1}{K} \sum_{k=1}^K I_A(s_{1k}) \quad (2.19)$$

para $A \subset U$, pode-se obter uma relação entre p_1 e p_0 . Para fazê-lo é necessário introduzir o conceito de *contra-imagem* de um intervalo $A \subset U$ sob a operação do mapa $f(\cdot)$ ou

$$f^{-1}(A) = \{s \in U / f(s) \in A\}. \quad (2.20)$$

Como ilustrado na Figura 2.8, para o mapa $f_Q(\cdot)$, a contra-imagem de um intervalo é a união de dois intervalos.

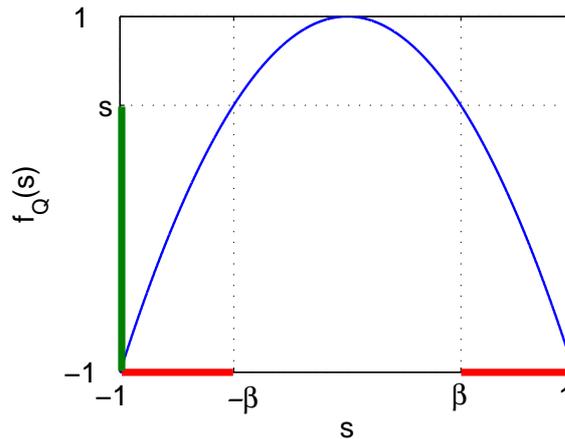


Figura 2.8: Contra-imagem do intervalo $[-1, s]$ por $f_Q(\cdot)$ representada por linhas grossas no eixo das abscissas com $\beta = \sqrt{\frac{1-s}{2}}$.

Note que para qualquer $A \subset U$,

$$s_{1k} \in A \Leftrightarrow s_{0k} \in f^{-1}(A). \quad (2.21)$$

Assim, tem-se a seguinte relação útil

$$I_A(f(s)) = I_{f^{-1}(A)}(s) \quad (2.22)$$

Com a Eq. (2.22) pode-se reescrever a Eq. (2.19) como

$$\int_A p_1(u) du \approx \frac{1}{K} \sum_{k=1}^K I_{f^{-1}(A)}(s_{0k}). \quad (2.23)$$

Dada a arbitrariedade de A_0 e A pode-se tomar $A_0 = f^{-1}(A)$. Com esta escolha os lados direitos das Eqs. (2.18) e (2.23) são iguais e, portanto

$$\int_A p_1(u) du = \int_{f^{-1}(A)} p_0(u) du. \quad (2.24)$$

Essa é a relação desejada entre $p_0(\cdot)$ e $p_1(\cdot)$. Dela percebe-se como a densidade inicial de estados $p_0(\cdot)$ é transformada por um dado mapa $f(\cdot)$ em uma nova densidade $p_1(\cdot)$.

Supondo $A = [a, s]$ obtém-se uma representação explícita para $p_1(\cdot)$ reescrevendo a Eq. (2.24)

$$\int_a^s p_1(u) du = \int_{f^{-1}([a,s])} p_0(u) du \quad (2.25)$$

e diferenciando em relação a s :

$$p_1(s) = \frac{d}{ds} \int_{f^{-1}([a,s])} p_0(u) du. \quad (2.26)$$

Fica claro que $p_1(\cdot)$ depende de $p_0(\cdot)$. Na literatura, esse fato é usualmente indicado por $p_1 = \mathcal{P}p_0$ (LASOTA; MACKEY, 1985), de forma que a Eq. (2.26) torna-se

$$\mathcal{P}p(s) = \frac{d}{ds} \int_{f^{-1}([a,s])} p(u) du. \quad (2.27)$$

em que se retirou o subscrito de p_0 . A Eq. (2.27) define explicitamente o *operador de Frobenius-Perron* \mathcal{P} correspondente à transformação $f(\cdot)$; esse operador é útil no estudo da evolução de densidades.

Por exemplo, para ilustrar a utilidade da Eq. (2.27) e do conceito do operador de Frobenius-Perron retorna-se ao mapa quadrático $f_Q(\cdot)$ aplicando a Eq. (2.27) a uma fórmula analítica da contra-imagem do intervalo $[-1, s]$.

A Figura 2.8 mostra que os pontos extremos dos dois intervalos constituintes de $f_Q^{-1}([-1, s])$ podem ser calculados resolvendo-se uma equação quadrática. Assim,

$$f_Q^{-1}([-1, s]) = \left[-1, -\sqrt{\frac{1-s}{2}}\right] \cup \left[\sqrt{\frac{1-s}{2}}, 1\right]. \quad (2.28)$$

Com isto, a Eq. (2.27) pode ser escrita como:

$$\mathcal{P}p(s) = \frac{d}{ds} \int_{-1}^{-\sqrt{\frac{1-s}{2}}} p(u) du + \frac{d}{ds} \int_{\sqrt{\frac{1-s}{2}}}^1 p(u) du, \quad (2.29)$$

e após realizar as diferenciações indicadas resulta

$$\mathcal{P}p(s) = \frac{1}{2\sqrt{2(1-s)}} \left[p\left(-\sqrt{\frac{1-s}{2}}\right) + p\left(\sqrt{\frac{1-s}{2}}\right) \right]. \quad (2.30)$$

Essa equação é uma fórmula explícita para o operador de Frobenius-Perron correspondente ao mapa $f_Q(\cdot)$ e mostra como $f_Q(\cdot)$ transforma uma dada densidade $p(\cdot)$ em outra $\mathcal{P}p(\cdot)$.

Por exemplo, tome-se uma densidade inicial $p(s) = \frac{1}{2}$ para $s \in [-1, 1]$. Então, como os termos dentro dos colchetes na Eq. (2.30) são constantes, um cálculo simples fornece:

$$\mathcal{P}p(s) = \frac{1}{2\sqrt{2(1-s)}}. \quad (2.31)$$

Pode-se então substituir essa expressão para $\mathcal{P}p(\cdot)$ no lugar de $p(\cdot)$ no segundo membro da Eq. (2.30) obtendo-se

$$\begin{aligned} \mathcal{P}(\mathcal{P}p(s)) &= \mathcal{P}^2 p(s) = \frac{1}{2\sqrt{2(1-s)}} \left[\frac{1}{2\sqrt{2\left(1 + \sqrt{\frac{1-s}{2}}\right)}} + \frac{1}{2\sqrt{2\left(1 - \sqrt{\frac{1-s}{2}}\right)}} \right] \\ &= \frac{1}{4\sqrt{2(1-s)}} \left[\frac{1}{\sqrt{2 + \sqrt{2(1-s)}}} + \frac{1}{\sqrt{2 - \sqrt{2(1-s)}}} \right]. \end{aligned} \quad (2.32)$$

Na Figura 2.9(a) são mostrados gráficos de $p(s) = \frac{1}{2}$, $\mathcal{P}p(s)$ dada pela Eq. (2.31) e $\mathcal{P}^2p(s)$ dada pela Eq. (2.32). Nela se vê que essas densidades aproximam-se de uma densidade limite. Essa é dada por:

$$p_*(s) = \begin{cases} \frac{1}{\pi\sqrt{1-s^2}}, & -1 \leq s \leq 1 \\ 0, & \text{caso contrário} \end{cases} \quad (2.33)$$

como pode ser verificado substituindo-se $p_*(s)$ na equação $\mathcal{P}p_*(s) = p_*(s)$, conduzindo portanto a densidade invariante. Esse resultado pode também ser obtido a partir da conjugação entre $f_T(\cdot)$ e $f_Q(\cdot)$ (KENNEDY; ROVATTI; SETTI, 2000).

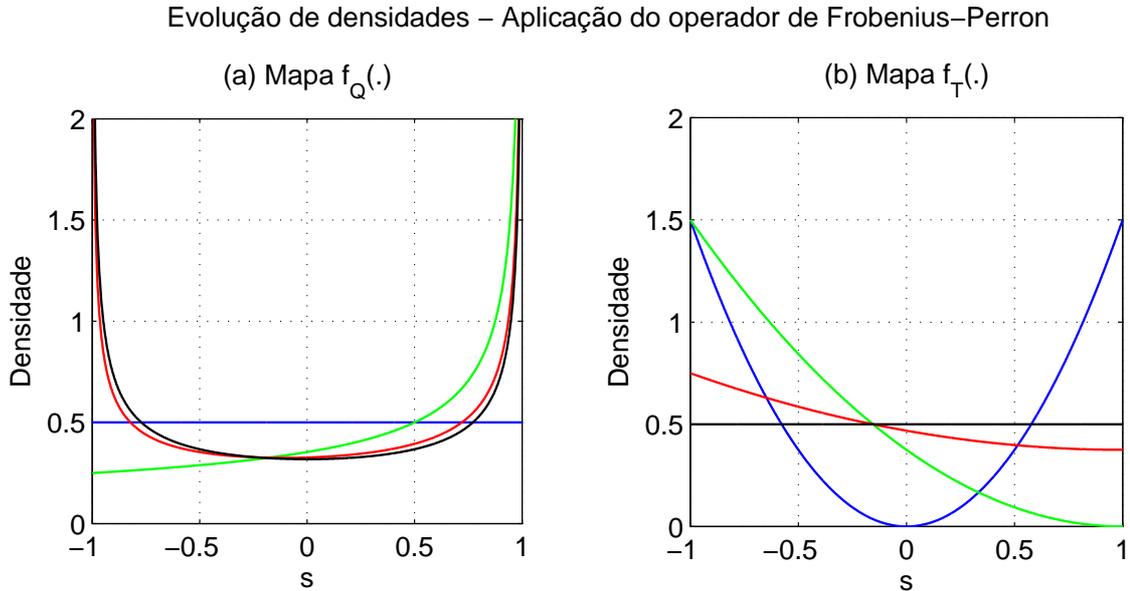


Figura 2.9: (a) Evolução da densidade constante $p(s) = \frac{1}{2}$ por $f_Q(\cdot)$; (b) Evolução da densidade $p(s) = \frac{3}{2}s^2$ por $f_T(\cdot)$. Em azul $p(s)$, em verde $\mathcal{P}p(s)$, em vermelho $\mathcal{P}^2p(s)$ e em preto a densidade invariante, $p_*(s)$.

Note-se também a semelhança entre o gráfico de $p_*(s)$ da Figura 2.9(a) com os histogramas da Figura 2.6. Na verdade, pode-se mostrar que, para o mapa quadrático, a distribuição dos pontos ao longo de uma órbita típica aproxima-se da densidade invariante única $p_*(s)$ da mesma forma que as iterações de densidades (LASOTA; MACKAY, 1985), propriedade que pode ser reconhecida como uma forma de *ergodicidade* desse mapa (KENNEDY; ROVATTI; SETTI, 2000).

Como outro exemplo, considera-se a família de mapas tenda inclinada $f_I(\cdot)$ definida na Eq. (2.11) e ilustrada na Figura 2.3. Tome-se um intervalo $[-1, s] \subset [-1, 1]$ cuja contra-

imagem sob $f_I(\cdot)$ é dada por

$$f_I^{-1}([-1, s]) = [-1, s_1] \cup [s_2, 1] \quad (2.34)$$

sendo $s_1 = -\frac{(\alpha+1)s+(\alpha-1)}{2}$ e $s_2 = \frac{(\alpha-1)s+\alpha+1}{2}$. Sendo assim, o operador de Perron-Frobenius é

$$\begin{aligned} \mathcal{P}p(s) &= \frac{d}{ds} \left(\int_{-1}^{s_1} p(u)du + \int_{s_2}^1 p(u)du \right) \\ &= \frac{1}{2} [(\alpha+1)p(s_1) + (1-\alpha)p(s_2)]. \end{aligned} \quad (2.35)$$

A Eq. (2.35) corresponde ao operador de Frobenius-Perron para os mapas $f_I(\cdot)$. Observa-se que também, neste caso, as densidades $p(\cdot)$ são levadas por \mathcal{P} a uma densidade invariante como se vê na Figura 2.9(b) para o caso $\alpha = 0$, ou seja, para o mapa tenda $f_T(\cdot)$. Nessa figura, considerou-se como densidade inicial $p(s) = \frac{3}{2}s^2$ que é levada a $\mathcal{P}p(s) = \frac{3}{8}(s-1)^2$ e $\mathcal{P}^2p(s) = \frac{3}{32}(s^2 - 2s + 5)$. Nesse caso, como $\mathcal{P}\frac{1}{2} = \frac{1}{2}$, é fácil ver que a distribuição invariante é uniforme,

$$p_*(s) = \begin{cases} \frac{1}{2}, & -1 \leq s \leq 1 \\ 0, & \text{caso contrário} \end{cases} \quad (2.36)$$

A ergodicidade observada para o mapa quadrático também se verifica nesse exemplo (KENNEDY; ROVATTI; SETTI, 2000).

2.7 Conclusões

Neste capítulo foram definidos e exemplificados conceitos fundamentais da teoria de sistemas dinâmicos não-lineares necessários para o desenvolvimento dos próximos capítulos. Entre os conceitos destacam-se os de órbita caótica, número de Lyapunov, densidade invariante e o operador de Frobenius-Perron. Como o foco dos próximos capítulos é o caso unidimensional de tempo discreto, limitou-se apenas ao caso de interesse.

Essa revisão pretendeu estabelecer uma notação coerente tornando os resultados da tese acessíveis mesmo a leitores que não sejam especialistas.

Além disso, com relação a caos, órbitas e sistemas caóticos, diversos autores usam definições levemente diferentes. Assim, o autor considerou importante explicitar as definições utilizadas nesta tese.

A Tabela 2.1 mostra os enunciados dos teoremas deste capítulo e a Tabela 2.2 lista as principais definições dadas.

Tabela 2.1: Teoremas do Capítulo 2

Teorema	Enunciado
1	Seja o mapa $s(n+1) = f(s(n))$. Se a órbita $s(n, s_0)$ satisfaz $f'(s(n)) \neq 0$ para todo $n \in \mathbb{N}$ e é assintoticamente periódica tendo como limite a órbita periódica $s(n, p)$, então as duas órbitas têm expoentes de Lyapunov idênticos assumindo que ambos estejam definidos.
2	Considere os mapas $f(\cdot)$ e $g(\cdot)$ conjugados por $C(\cdot)$. Pode-se afirmar que: 1. Se p é um ponto periódico de período k de $f(\cdot)$, então $C(p)$ é um ponto periódico de período k de $g(\cdot)$. 2. Se a derivada de $C(\cdot)$ é não-nula em todos os pontos da órbita $s(n, s_0)$ de $f(\cdot)$ então existe uma órbita correspondente $C(s(n, s_0))$ de $g(\cdot)$ com mesmo número de Lyapunov.
3	Suponha que uma órbita $s(n, s_0)$ tenha número de Lyapunov $L(s_0)$ e expoente de Lyapunov $h(s_0)$. Essa órbita apresentará DCI se, e somente se, $L(s_0)$ for maior do que a unidade, isto é, $h(s_0) > 0$.
4	Uma órbita limitada $s(n, s_0)$ é caótica se, e somente se é aperiódica e $L(s_0) > 1$.

Tabela 2.2: Principais definições do Capítulo 2

Conceito	Localização
Mapa	Definição 1
Órbita	Definição 2
Condição inicial	Definição 2
Trajectoria	Definição 3
Ponto fixo	Definição 4
Órbita eventualmente periódica	Definição 5
Órbita periódica	Definição 5
Órbita assintoticamente periódica	Definição 6
Órbita aperiódica	Definição 6
Ponto fixo atrator	Definição 7
Ponto fixo estável segundo Lyapunov	Definição 7
Ponto fixo assintoticamente estável	Definição 7
Ponto fixo instável	Definição 7
Órbita estável segundo Lyapunov	Definição 8
Órbita assintoticamente estável segundo Lyapunov	Definição 8
Conjunto limite	Definição 10
Atrator	Definição 11
Bacia de atração	Definição 11
Estabilidade estrutural	Definição 12
Dependência sensível às condições iniciais	Definição 13
Número de Lyapunov	Definição 14
Expoente de Lyapunov	Definição 14
Mapas conjugados	Definição 15
Órbita caótica	Definição 16
Conjunto caótico	Definição 17
Atrator caótico	Definição 17
Contra-imagem	Equação 2.20
Operador de Frobenius-Perron	Equação 2.27

Capítulo 3

Elementos básicos de comunicação digital

Este capítulo resume conceitos básicos ligados a modulações digitais convencionais. Pretende-se definir uma estrutura genérica de sistema de comunicação junto com medidas de desempenho que se apliquem a qualquer modulação digital, inclusive caótica, pois assim é mais fácil compreender suas vantagens e desvantagens e compará-las.

Em geral a transmissão de informação digital é feita por meio de um canal analógico limitado em frequência. Para que a informação possa ser transmitida através do meio analógico, ela deve primeiramente ser mapeada num conjunto de sinais de duração finita. O receptor, baseado no conjunto de sinais que chega até ele deve ser capaz de determinar as informações digitais transmitidas.

Essa estrutura básica permanece inalterada tanto para sistemas convencionais quanto para sistemas que usam sinais caóticos. Porém, para esses últimos, uma série de aspectos adicionais devem ser analisados, o que é feito no próximo capítulo.

Na Seção 3.1, descreve-se e justifica-se o modelo de canal que será usado nas simulações. Em seguida, (Seção 3.2), discute-se o modelo equivalente de banda base em tempo discreto para sinais passa-banda. Esse modelo simplifica bastante a simulação e a análise dos sistemas de comunicação estudados. A Seção 3.3 trata dos processos de modulação e demodulação apresentando o conceito de funções de base ou sinais de base (WOZENCRAFT; JACOBS,

1965). Na Seção 3.4 são descritas as principais configurações de recepção de sistemas de comunicações digitais. Ao final do capítulo, esses conceitos são ilustrados com o exemplo clássico do chaveamento de fase (PSK - *Phase Shift Keying*), no caso binário e com recepção coerente.

3.1 Estrutura utilizada

O objetivo de um sistema de comunicação digital, esquematizado na Figura 3.1, é transportar informação de uma fonte digital para um receptor da forma mais eficiente possível.

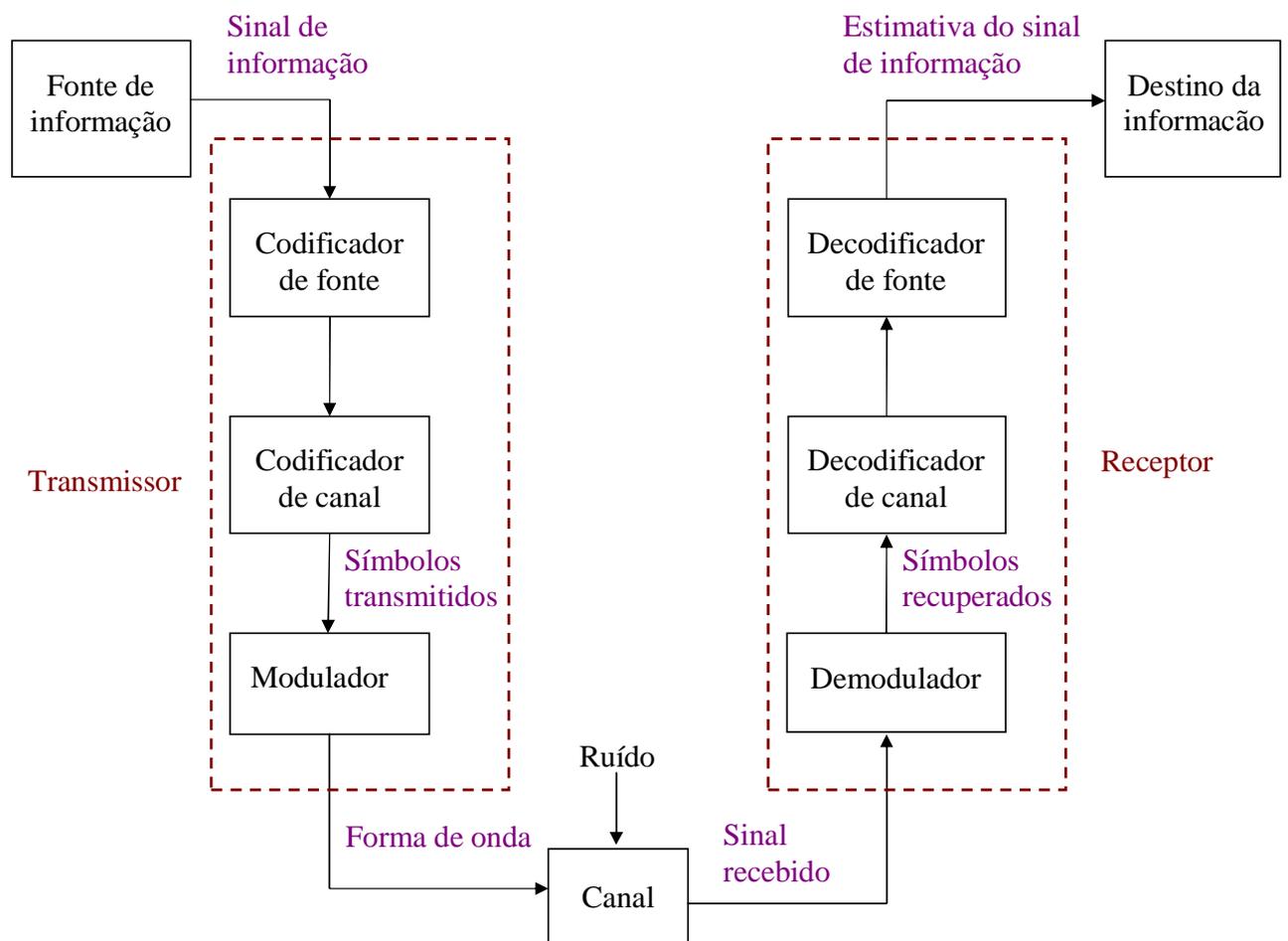


Figura 3.1: Diagrama de blocos de um sistema de comunicações digitais.

Ao converter uma seqüência de informação digital em formas de onda é comum intercalar processos de *codificação de fonte e de canal* para o aumento da segurança da transmissão, a compressão dos dados e a correção de alguns erros. Essas codificações geram a seqüência de

símbolos a ser transmitida.

A seguir, os símbolos são utilizados para determinar propriedades de uma portadora. Esse processo é conhecido como *modulação*. Alguns exemplos de modulações digitais comuns são o chaveamento de amplitude (ASK - *Amplitude Shift Keying*), chaveamento de frequência (FSK - *Frequency Shift Keying*) e o chaveamento de fase (PSK - *Phase Shift Keying*) em que o símbolo transmitido determina, respectivamente, a amplitude, a frequência ou a fase de uma portadora senoidal, respectivamente.

O *canal* é o meio físico pelo qual o sinal se propaga desde o transmissor até o receptor.

A tarefa do *demodulador* no receptor é produzir uma estimativa da seqüência de símbolos transmitida a partir da forma de onda que chega a ele. O papel do *decodificador de canal* e do *decodificador de fonte* é gerar uma estimativa da informação original a partir da seqüência de símbolos obtida no demodulador.

3.1.1 O modelo de canal

Para se avaliar o desempenho de uma dada modulação, o modelo de canal utilizado deveria representar de forma razoável os fenômenos que ocorrem entre a saída do transmissor e a entrada do receptor. Na prática, no entanto, adotam-se modelos simplificados. Mesmo assim, devem ser levados em conta pelo menos dois fenômenos quase sempre presentes:

- ruído térmico modelado como ruído branco gaussiano aditivo (AWGN - *Additive White Gaussian Noise*) com média nula e
- limitação em banda do canal gerada pela presença de filtros passa-banda na entrada do receptor para evitar interferências e sinais indesejados.

Dessa forma, o modelo mais simples de canal para avaliar um sistema de modulação é o mostrado na Figura 3.2 em que $r(t)$ é AWGN com média nula. Vale ressaltar que o filtro passa-banda indicado tem como finalidade selecionar a frequência de transmissão no receptor, pois muitas vezes os sistemas são multiplexados em frequência (LATHI, 1998).

Nesse modelo de canal considera-se que o sinal é corrompido por AWGN com média nula. Existem três motivos para que esse modelo seja considerado nesse trabalho (HAYKIN, 2000;

PROAKIS, 1995):

- o ruído AWGN permite muitas vezes a obtenção de resultados teóricos exatos;
- o ruído térmico, presente em todo sistema de comunicação, tem esse aspecto;
- experimentalmente, o desempenho relativo de numerosos sistemas de modulação quando se usa um canal AWGN não se altera significativamente com o uso de canal com características espectrais mais complicadas.

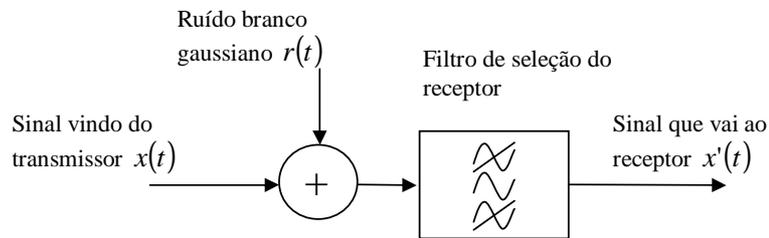


Figura 3.2: Modelo de canal de comunicações usado nas simulações.

Em um sistema de comunicação digital, o desempenho é usualmente medido em termos da taxa de erro de símbolo (SER - *Symbol Error Rate*). Tal taxa varia com a relação entre a energia média por símbolo e o dobro da densidade espectral de potência do ruído (E_b/N_0).

Como ilustração, a Figura 3.3 mostra a SER em função de E_b/N_0 para algumas modulações digitais binárias convencionais (LATHI, 1998) sem codificação em canal AWGN: ASK coerente, ASK não-coerente, PSK diferencial (DPSK - *Differential PSK*) e PSK coerente.

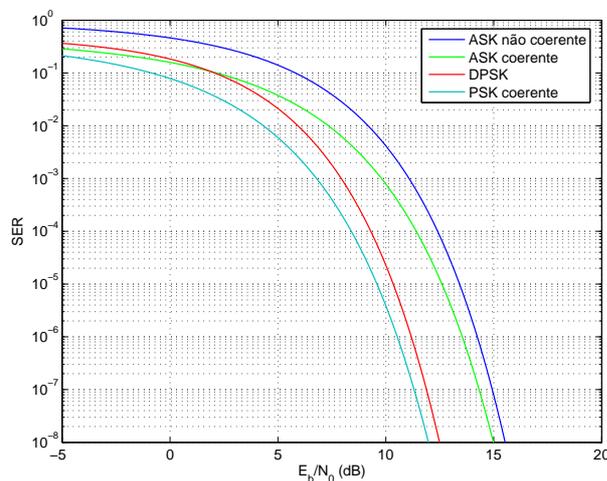


Figura 3.3: Exemplos de curvas SER para modulações digitais convencionais em AWGN.

3.2 Modelos equivalentes de tempo discreto

Para facilitar as simulações e análises de modulações digitais é usual trabalhar com modelos equivalentes em banda base de tempo discreto, cujos resultados são válidos para os modelos originais. Este modelamento permite representar os sinais transmitidos por meio de seqüências de comprimento finito. No caso das modulações caóticas, esses modelos também oferecem a vantagem de se usar diretamente a notação de sistemas dinâmicos discretos definida no Capítulo 2.

3.2.1 Sistemas passa-banda

Nem sempre os sinais produzidos por uma fonte de informação podem ser transmitidos diretamente por um dado canal. Por isso, esses sinais, chamados de sinais em banda base, são modificados para facilitar a transmissão. No caso convencional, o sinal em banda base, suposto ter uma largura de banda B , é usado para modificar algum parâmetro de uma portadora senoidal com frequência $f_0 \gg B$. A Figura 3.4 mostra o espectro de magnitude típico do sinal passa-banda transmitido.

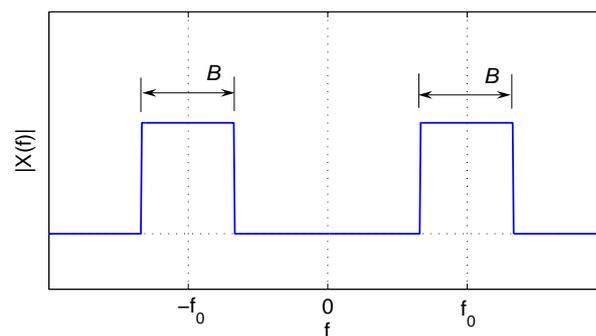


Figura 3.4: Espectro de magnitude típico de sinal gerado por um sistema passa-banda.

Esse sinal é transmitido através do canal AWGN descrito na Figura 3.2. Admitindo-se que o filtro de seleção do receptor seja ideal com faixa de passagem de largura B centrada em f_0 , todos os sinais interferentes que não caiam dentro da banda de frequências ocupada pelo sinal transmitido são eliminados. A saída do filtro é da forma:

$$x'(t) = x(t) + r'(t) \quad (3.1)$$

em que $r'(t)$ denota o ruído gaussiano filtrado.

3.2.2 Modelo equivalente em banda base de tempo discreto

Para a maioria das finalidades práticas, o estudo de sistemas de comunicação passa-bandas é realizado de forma mais conveniente utilizando-se sistemas equivalentes em banda base. O sinal passa-banda é descrito (PROAKIS, 1995) por

$$x(t) = x_c(t) \cos(2\pi f_0 t) - x_s(t) \sin(2\pi f_0 t). \quad (3.2)$$

A representação passa-baixas de $x(t)$, denotada por $x_l(t)$, é em geral complexa com parte real $x_c(t)$ e parte imaginária $x_s(t)$:

$$x_l(t) = x_c(t) + jx_s(t). \quad (3.3)$$

Tanto $x_c(t)$ quanto $x_s(t)$ são sinais passa-baixas, normalmente chamados respectivamente de componente em fase e quadratura do sinal passa-banda (HAYKIN, 2000).

Em termos de equivalentes passa-baixas, a Eq. (3.1) pode ser reescrita como

$$x'_l(t) = x_l(t) + r'_l(t) \quad (3.4)$$

A versão em tempo discreto do modelo em banda base é obtida, amostrando-se o sinal passa-baixas equivalente com um período de amostragem

$$T_a = \frac{1}{B}. \quad (3.5)$$

Pelo teorema da amostragem (HAYKIN, 2000), o sinal equivalente em banda base de tempo contínuo pode ser recuperado a partir do sinal amostrado por meio de um filtro passa-baixas ideal com largura de banda $B/2$. Para simplificar a notação, denota-se $x_l(nT_a)$, $x'_l(nT_a)$ e

$r'_l(nT_a)$ respectivamente por $x(n)$, $x'(n)$ e $r(n)$. Assim, a Eq. (3.4), no tempo discreto torna-se:

$$x'(n) = x(n) + r(n). \quad (3.6)$$

Pode-se mostrar que, se $r(t)$ é AWGN então as amostras $r(n)$ são estatisticamente independentes (PROAKIS, 1995).

3.3 Modulação e demodulação

O conjunto de seqüências equivalentes de tempo discreto das formas de onda empregadas num dado sistema com M símbolos distintos é representado por $x_m(n)$, $m = 1, 2, \dots, M$, sendo que $x_m(n)$ assume valores não-nulos apenas no intervalo $0 \leq n \leq N - 1$. Para se transmitir o m -ésimo símbolo, o sinal representado em tempo discreto por $x_m(n)$ é transmitido pelo canal analógico em que pode ser distorcido. Além disso, poderá ter adicionado um ruído. Na demodulação o sinal recebido, cujo equivalente $x'_m(n)$ é uma seqüência estocástica, será mapeado de volta em um símbolo.

Para simplificar o processo de demodulação, é interessante definir um conjunto com menos sinais, denominado *funções de base*, sendo que os sinais representados por $x_m(n)$ são combinações lineares dos elementos dessa base (WOZENCRAFT; JACOBS, 1965).

Seja $s_i(n)$, $i = 1, 2, \dots, N_b$, $n = 0, 1, \dots, N - 1$, uma base de seqüências ortonormais, ou seja,

$$\sum_{n=0}^{N-1} s_i(n)s_j(n) = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases}, \quad 1 \leq i, j \leq N_b. \quad (3.7)$$

Dessa forma, cada um dos M sinais $x_m(n)$ é representado por uma combinação das N_b seqüências $s_i(n)$, sendo $N_b \leq M$:

$$x_m(n) = \sum_{i=1}^{N_b} x_{mi}s_i(n), \quad m = 1, 2, \dots, M. \quad (3.8)$$

Os coeficientes x_{mi} na Eq. (3.8) podem ser interpretados como as componentes de um vetor coluna N_b -dimensional \mathbf{x}_m . Cada símbolo determina um vetor \mathbf{x}_m . A seqüência a ser transmi-

tida para cada \mathbf{x}_m pode então ser gerada de acordo com a Eq. (3.8). A Figura 3.5 exemplifica a geração da sequência $x_m(n)$ para o caso $N_b = 2$.

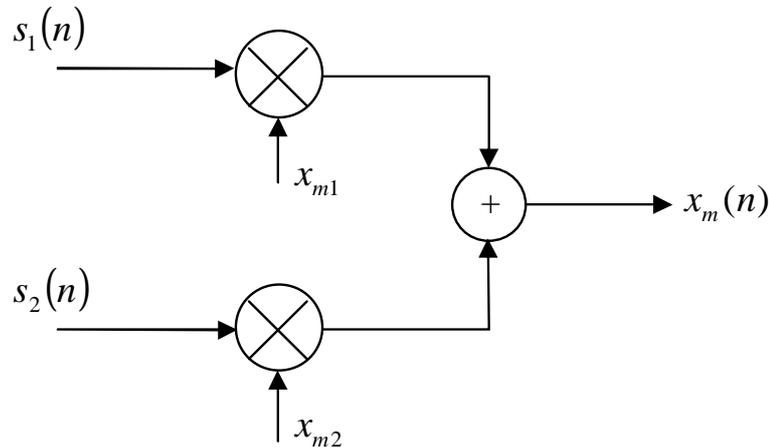


Figura 3.5: Geração de $x_m(n)$ a partir das funções de base, caso $N_b = 2$.

Como as seqüências de base são ortonormais, o vetor \mathbf{x}_m pode ser recuperado a partir do sinal transmitido desde que todos os sinais de base $s_i(n)$ sejam conhecidos. Para isso, basta calcular

$$x_{mi} = \sum_{n=0}^{N-1} x_m(n) s_i(n), \quad i = 1, 2, \dots, N_b. \quad (3.9)$$

Assim, uma forma de implementar um demodulador é por meio de um conjunto de N_b correlatores, cada um obtendo um dos pesos x_{mi} associado à função de base $s_i(n)$. Como existe uma relação biunívoca entre os vetores \mathbf{x}_m e os símbolos transmitidos, esses podem ser recuperados processando-se as saídas dos correlatores¹ e o sinal de informação original regenerado.

3.4 Configurações de receptores

O receptor deve ser capaz de reconhecer os símbolos enviados pelo canal, o que possibilita recuperar a informação transmitida. Por simplicidade, apenas a recuperação de um único símbolo isolado é considerada deixando-se de lado a interferência intersimbólica (LATHI, 1998).

¹Seguindo a literatura da área de comunicações na descrição de receptores, e.g. (LATHI, 1998; KENNEDY; ROVATTI; SETTI, 2000), usa-se o termo correlação para a operação constituída pelo produto de dois sinais num dado intervalo seguida pela somatória das amostras do sinal resultante. Quando os dois sinais envolvidos são idênticos, a operação também é chamada de *autocorrelação* quando são diferentes é chamada de *correlação cruzada*.

Considerando-se canal AWGN e que os possíveis símbolos transmitidos são equi-prováveis, deve ser utilizado o método de detecção por máxima verossimilhança (ML-*Maximum Likelihood*) para se obter um receptor ótimo, ou seja, aquele que minimiza a SER para uma dada E_b/N_0 (LATHI, 1998).

O método ML pode ser implementado através do receptor de correlação ou por filtro casado (HAYKIN, 2000). Nesse último caso, as seqüências de base são armazenadas localmente como a resposta impulsiva desses filtros. No caso de comunicação utilizando sinais caóticos, a seqüência transmitida muda a cada símbolo e a utilização de filtros casados não é possível. Assim, apenas a recepção por correlação é discutida aqui.

A Eq. (3.9) mostra como as componentes x_{mi} podem ser obtidas a partir do sinal transmitido por meio de correladores, se as N_b seqüências da base $s_i(n)$ são ortonormais e conhecidas no receptor. Note que, pensando num fluxo contínuo de símbolos, além das funções de base, o comprimento das seqüências N e a amostra em que $x_m(n)$ começou a ser transmitida também devem ser conhecidos. Esses últimos dados são conhecidos como *informações de temporização*. A idéia sugerida pela Eq. (3.9) é utilizada para projetar o receptor de correlação para $N_b = 2$ mostrado na Figura 3.6.

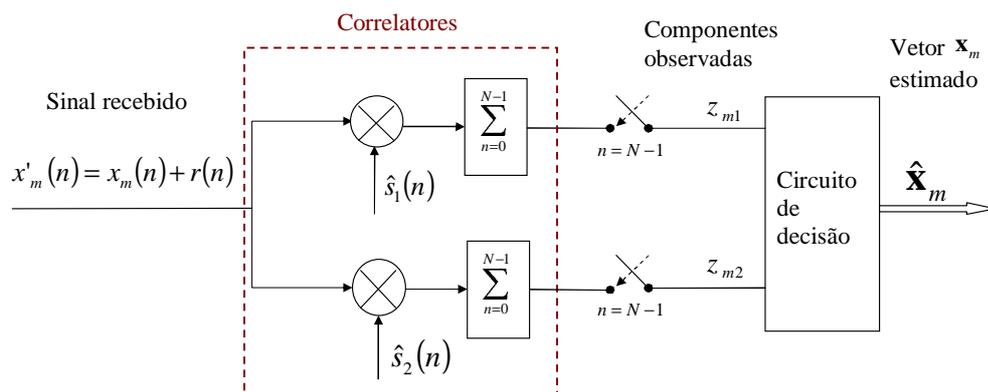


Figura 3.6: Diagrama de blocos de um receptor de correlação com $N_b = 2$.

Os componentes x_{mi} podem ser estimados por correlação, mas devido ao ruído, os valores obtidos possivelmente não coincidem com os valores nominais.

As saídas dos correladores, chamadas aqui de *componentes observadas* formam um *vetor de observação* \mathbf{z}_m . Elas são as entradas de um circuito decisor. Esse circuito aplica o método de detecção ML, ou seja, ele escolhe dentre os possíveis vetores \mathbf{x}_m qual está à menor distância eu-

clidiana do vetor de observação \mathbf{z}_m . O vetor estimado $\hat{\mathbf{x}}_m$, corresponde a um símbolo estimado que determina a informação recuperada.

É importante ressaltar que, num receptor de correlação, todas as seqüências de base e as informações de temporização são necessárias. Elas podem ser estimadas a partir do sinal ruidoso recebido. As seqüências de base estimadas são representadas por $\hat{s}_1(n)$ e $\hat{s}_2(n)$ na Figura 3.6.

Receptores nos quais são obtidas cópias das funções de base, ou seja, $\hat{s}_i(n) \approx s_i(n)$, são chamados de *receptores coerentes* (LATHI, 1998). Caso as seqüências de base $s_i(n)$ não possam ser recuperadas, são usados *receptores não-coerentes*. Nesse caso são usadas uma ou mais características robustas dos sinais $x_m(n)$, $m = 1, 2, \dots, M$ para diferenciá-los. A demodulação é obtida estimando-se uma ou mais características selecionadas do sinal recebido. Em aplicações em que a relação sinal-ruído no canal é muito baixa, pode ser muito difícil recuperar com precisão as funções de base a partir do sinal recebido. Nesses casos, a solução clássica é utilizar recepção não-coerente.

Uma vantagem atribuída aos receptores coerentes em relação aos não-coerentes para uma dada modulação é o seu melhor desempenho em canal AWGN (LATHI, 1998). Os desempenhos da modulação ASK com recepção coerente e não-coerente são mostrados na Figura 3.3 da página 68. Em contrapartida, as técnicas não-coerentes oferecem duas vantagens sobre a recepção coerente:

- quando a relação sinal-ruído no canal é baixa, os sinais de base $s_i(n)$ são dificilmente recuperados a partir do sinal que chega ao receptor porque os sinais $x'_m(n)$ são muito diferentes dos sinais $x_m(n)$. Nesse caso, a recepção não-coerente é a adotada;
- os receptores não-coerentes podem, em princípio, ser implementados com circuitos simples, já que não há necessidade de se recuperar os sinais da base.

3.5 Exemplo: o PSK binário

O diagrama de blocos do equivalente em banda base de tempo discreto de um sistema PSK binário é mostrado na Figura 3.7.

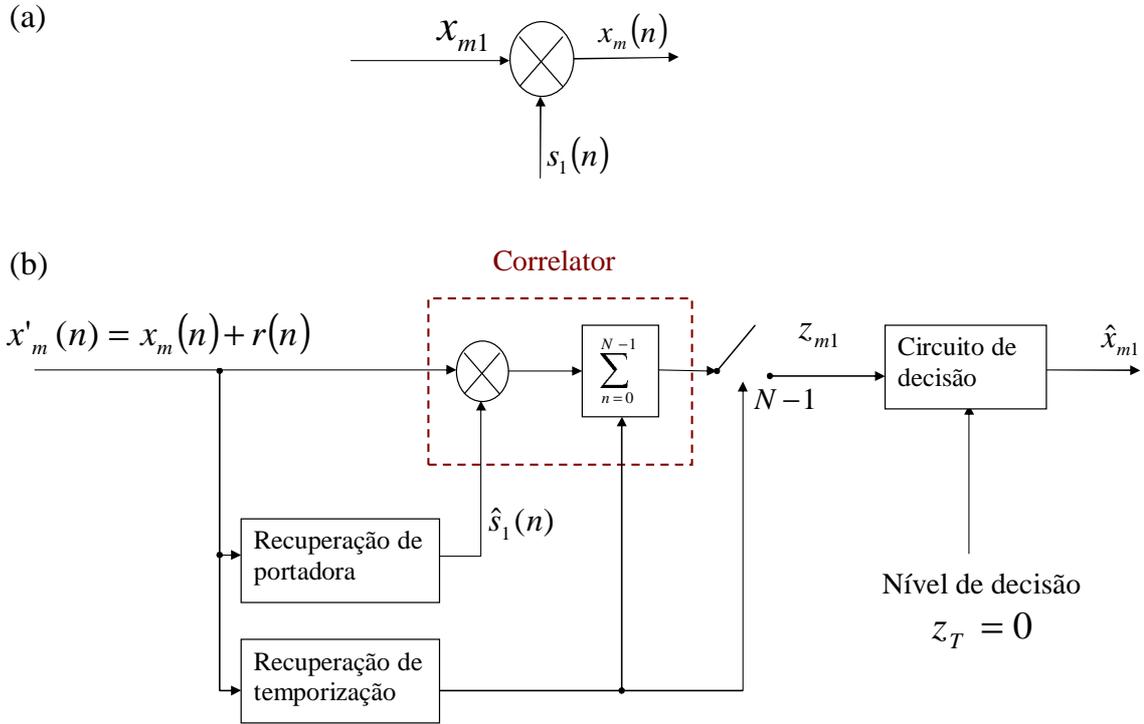


Figura 3.7: Diagrama de blocos de um sistema PSK binário com detecção coerente: (a) transmissor; (b) receptor.

No caso binário, os sinais transmitidos, $x_1(n)$ e $x_2(n)$, são

$$x_1(n) = \sqrt{\frac{E_b}{N}} \quad e \quad x_2(n) = -\sqrt{\frac{E_b}{N}} \quad (3.10)$$

com $0 \leq n \leq N - 1$.

É utilizada apenas uma função de base com energia unitária

$$s_1(n) = \frac{1}{\sqrt{N}}, \quad 0 \leq n \leq N - 1, \quad (3.11)$$

sendo $x_{11} = +\sqrt{E_b}$ e $x_{21} = -\sqrt{E_b}$.

O sinal recebido $x'_m(n)$ é o sinal transmitido $x_m(n)$ com $m = 1$ ou $m = 2$ adicionado a ruído branco gaussiano, ou seja, $x'_m(n) = x_m(n) + r(n)$. No receptor PSK binário ocorrem as seguintes operações:

- *recuperação de portadora* - a partir de $x'_m(n)$, estima-se a função de base $s_1(n)$;
- *recuperação de temporização* - extrai-se da seqüência de símbolos as informações de tem-

porização necessárias para que a correlação seja realizada sobre os pontos corretos, ou seja, $0 \leq n < N$. Além disso, o valor de N recuperado é necessário para definir o instante de obtenção da componente observada z_{m1} ;

- *obtenção de z_{m1}* - a componente observada é o resultado da correlação entre $x'_m(n)$ e a portadora recuperada $\hat{s}_1(n)$;
- *estimação de x_{m1}* - o coeficiente \hat{x}_{m1} é obtido pela análise do sinal de z_{m1} , como justificado a seguir.

Neste sistema, a componente observada é dada por

$$\begin{aligned} z_{m1} &= \sum_{n=0}^{N-1} x'_m(n) \hat{s}_1(n) = \sum_{n=0}^{N-1} (x_m(n) + r(n)) \hat{s}_1(n) = \sum_{n=0}^{N-1} (\pm \sqrt{E_b} s_1(n) + r(n)) \hat{s}_1(n) \\ &= \pm \sqrt{E_b} \sum_{n=0}^{N-1} s_1(n) \hat{s}_1(n) + \sum_{n=0}^{N-1} r(n) \hat{s}_1(n). \end{aligned} \quad (3.12)$$

Se a função de base é recuperada ou conhecida exatamente no receptor, então $\hat{s}_1(n) = s_1(n)$ e

$$z_{m1} = \pm \sqrt{E_b} \sum_{n=0}^{N-1} s_1^2(n) + \sum_{n=0}^{N-1} r(n) s_1(n) = \pm \sqrt{E_b} + \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} r(n). \quad (3.13)$$

Na ausência de ruído,

$$z_{m1} = \pm \sqrt{E_b}. \quad (3.14)$$

O comportamento da componente observada z_{m1} pode ser analisado por meio de histogramas, como os mostrados na Figura 3.8. Foi simulada a transmissão de 10000 símbolos com $E_b = 1$ e os valores admitidos por z_{m1} foram divididos em intervalos de comprimento 0,05. No histograma (a) estão apresentados os valores da componente observada no caso de canal ideal e o histograma (b) no caso de canal AWGN com $E_b/N_0 = 15\text{dB}$. Pode se ver que as amostras ficam concentradas em torno de $-\sqrt{E_b}$ e $+\sqrt{E_b}$.

Quando a segunda parcela na Eq. (3.13) é não-nula devido ao AWGN, z_{m1} torna-se uma variável aleatória com média $\pm \sqrt{E_b}$ e variância igual à potência média do ruído $r(n)$. Para minimizar a probabilidade de erro, isto é, para se obter um receptor ótimo, utiliza-se a detecção

por máxima verossimilhança. Nesse caso, a regra de decisão consiste em escolher $\hat{x}_{m1} = x_{11}$ se a saída do correlator for positiva no instante de decisão e $\hat{x}_{m1} = x_{21}$ caso contrário. O circuito de decisão é simplesmente um comparador com nível $z_T = 0$.

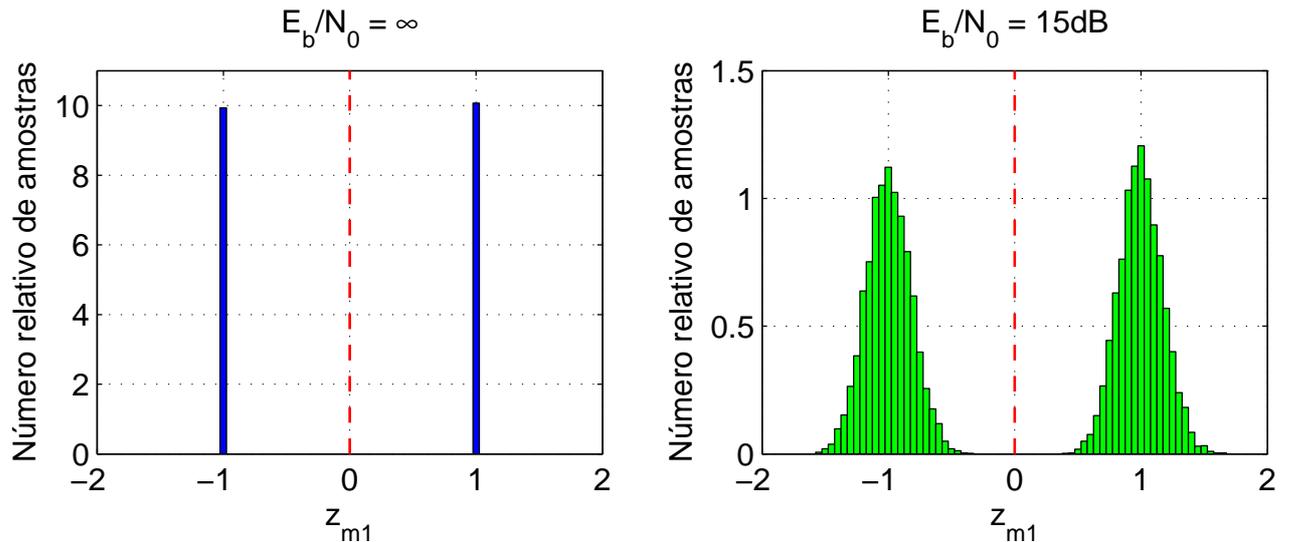


Figura 3.8: Histograma da componente observada z_{m1} com $E_b = 1$ (a) na ausência de ruído no canal e (b) com ruído no canal com $E_b/N_0 = 15\text{dB}$. Os histogramas foram calculados em $N_S = 40$ subintervalos tendo sido transmitidos 10000 símbolos. Em tracejado é indicado o nível de decisão.

Note que a taxa de erro é zero no caso sem ruído. Essa é uma propriedade importante de um esquema de modulação convencional baseado em portadoras senoidais e que não é sempre verdade para modulações caóticas.

3.6 Conclusões

Para facilitar a exposição sistemática dos sistemas de modulações digitais caóticos e a sua comparação com os sistemas convencionais, este capítulo expôs a base teórica comum utilizada nesses sistemas. A modulação digital consiste em mapear a informação a ser transmitida em variações de formas de onda, que no caso convencional são geralmente senoidais.

Introduziu-se o conceito de funções ou seqüências de base e o sistema equivalente passa-baixas de tempo discreto, o qual facilita as análises e simulações computacionais. Essas ferramentas possibilitam adotar uma notação unificada para todas as modulações digitais, baseadas ou não em portadoras senoidais, o que permite analisá-las com a mesma abordagem.

Capítulo 4

Modulação digital usando portadoras caóticas

Este capítulo descreve e analisa alguns sistemas de modulação digital propostos recentemente que utilizam funções de base caóticas e não as periódicas. Além disso, busca-se comparar seus desempenhos em termos de taxas de erro de símbolo (SER - *Symbol Error Rate*) quando submetidos a um canal AWGN.

O estudo concentra-se nas modulações chaveamento caótico (CSK - *Chaos Shift Keying*), chaveamento caótico diferencial (DCSK - *Differential Chaos Shift Keying*) e algumas de suas variantes, em particular o chaveamento caótico diferencial com modulação em frequência (FM-DCSK - *Frequency Modulated DCSK*).

No CSK, a informação é transmitida pelos coeficientes da combinação das seqüências de base usadas para gerar cada símbolo. No DCSK a informação também pode ser recuperada pela correlação entre partes das seqüências transmitidas.

Mesmo existindo alternativas, como a codificação da informação através do controle da dinâmica simbólica de um atrator (HAYES; GREBOGI; OTT, 1993; HAYES et al., 1994; Bollt, 2003), o CSK e o DCSK são os que têm chamado mais atenção dos pesquisadores na área de telecomunicações quando se tem em vista aplicações próximas da prática. Apesar de muita pesquisa ainda ser necessária, acredita-se que esses sistemas, principalmente o FM-DCSK, podem concorrer vantajosamente com os sistemas convencionais em algumas situações

especiais (KOLUMBÁN et al., 1998; KENNEDY; KOLUMBÁN, 2000; LAU; TSE, 2003).

Para facilitar a comparação entre os sistemas de modulação convencional e caótica utiliza-se a mesma terminologia do capítulo anterior.

Considera-se a transmissão de símbolos isolados, ou seja, desconsidera-se a interferência intersimbólica.

Nas Seções 4.1 e 4.2, descrevem-se detalhes da operação dos esquemas de modulação CSK e DCSK, supondo canal ideal. São utilizados seus equivalentes passa-baixas em tempo discreto para facilitar a notação e a comparação com os sistemas convencionais. Essa notação é mais adequada ao se tratar de sinais caóticos gerados por mapas, que são sistemas de tempo discreto por definição. Tal notação é uma distinção desta tese em relação à literatura do assunto.

Em seguida, são analisados nas Seções 4.3 e 4.4 o problema da variabilidade da energia por símbolo em sistemas de modulação caótica e o FM-DCSK, que em princípio resolve este problema.

Na Seção 4.5 é analisado qualitativamente o desempenho das modulações caóticas em canal AWGN. Em seguida, na Seção 4.6 são apresentados resultados de simulações computacionais dos sistemas não-coerentes e diferenciais tratados. Por fim, a Seção 4.7 apresenta as conclusões do capítulo.

4.1 O chaveamento caótico (CSK)

O CSK é uma modulação digital em que o símbolo a ser transmitido é codificado com os coeficientes de uma combinação linear de sinais caóticos gerados por diferentes atratores. Os atratores resultam de um mesmo sistema dinâmico para diferentes valores de um parâmetro de bifurcação (DEVANEY, 2003) ou de sistemas dinâmicos diferentes. Nessa seção, descreve-se o CSK supondo ausência de ruído no canal e recuperação perfeita das funções de base no receptor, quando necessária.

Usando a notação definida no Capítulo 3, o equivalente em banda base do sinal transmitido

associado ao m -ésimo símbolo é

$$x_m(n) = \sum_{i=1}^{N_b} x_{mi} s_i(n), \quad n = 0, 1, \dots, N-1, \quad m = 1, 2, \dots, M, \quad (4.1)$$

em que as seqüências de base $s_i(n)$ são trechos de sinais caóticos gerados por N_b atratores distintos. Como consequência da não-periodicidade, as seqüências $s_i(n)$ e, portanto, $x_m(n)$ são diferentes a cada símbolo transmitido. Os sinais $x_m(n)$ podem ser produzidos como mostrado na Figura 4.1 para o caso $N_b = 2$.

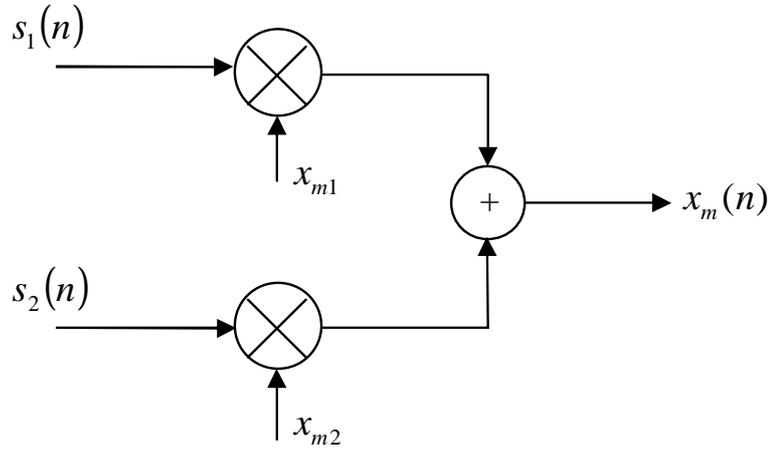


Figura 4.1: Geração de $x_m(n)$ a partir das seqüências de base caóticas, caso $N_b = 2$.

Impõe-se que as seqüências de base caóticas sejam ortonormais *em média*, ou seja,

$$E \left[\sum_{n=0}^{N-1} s_i(n) s_j(n) \right] = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases}, \quad 1 \leq i, j \leq N_b, \quad (4.2)$$

em que $E[\cdot]$ denota esperança matemática. A Eq. (4.2) identifica a importante característica dos esquemas de modulação digital caóticos: a ortonormalidade das funções de base pode ser definida apenas em termos de valores esperados. Como consequência, a correlação cruzada e a autocorrelação das seqüências de base são variáveis aleatórias.

Os coeficientes x_{mi} podem ser recuperados a partir do sinal transmitido por correlação com cópias das seqüências de base $\hat{s}_i(n)$ geradas localmente como mostrado na Figura 4.2 para o caso $N_b = 2$.

Supondo $\hat{s}_i(n) \approx s_i(n)$, o valor esperado da componente observada z_{mi} na saída do i -ésimo

correlator quando a seqüência $x_m(n)$ é transmitida é dado por

$$\begin{aligned} E[z_{mi}] &= E \left[\sum_{n=0}^{N-1} x_m(n) s_i(n) \right] = E \left[\sum_{n=0}^{N-1} \left(\sum_{j=1}^{N_b} x_{mi} s_j(n) \right) s_i(n) \right] \\ &= x_{mi} E \left[\sum_{n=0}^{N-1} s_i(n) s_i(n) \right] = x_{mi}. \end{aligned} \quad (4.3)$$

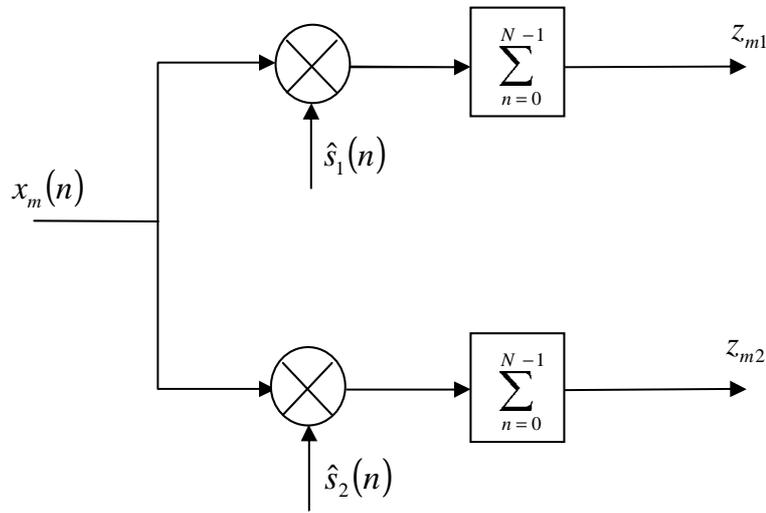


Figura 4.2: Obtenção do vetor de observação \mathbf{z}_m por correlação a partir do sinal transmitido, caso $N_b = 2$.

Assim, no caso de um canal ideal, o valor esperado do vetor de observações é \mathbf{x}_m .

Caso os coeficientes x_{mi} gerem sinais com energias diferentes para cada m , a demodulação também pode ser feita estimando-se a energia do sinal recebido no receptor. Nesta demodulação, não-coerente, são desnecessárias cópias das funções de base.

4.1.1 Modulação CSK com um mapa

No caso mais simples, apenas um atrator caótico é utilizado:

$$x_m(n) = x_{m1} s_1(n), \quad m = 1, 2, \dots, M; \quad n = 0, 1, \dots, N-1, \quad (4.4)$$

sendo $s_1(n)$ formado por N pontos de uma órbita desse atrator.

Um diagrama de blocos de um modulador CSK com $N_b = 1$ é mostrado na Figura 4.3.

No caso binário ($M = 2$), três possíveis escolhas para os valores de x_{m1} são usuais:

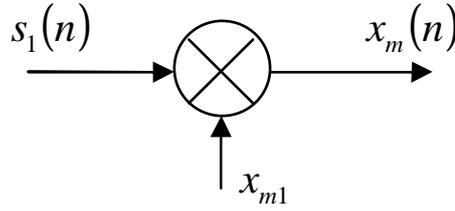


Figura 4.3: Modulador CSK com uma função de base.

- **CSK unipolar.** No CSK unipolar (KENNEDY; ROVATTI; SETTI, 2000), os símbolos representados por $x_1(n)$ e $x_2(n)$ são distinguíveis pela energia média transmitida para cada símbolo, E_{b1} e $E_{b2} = kE_{b1}$, respectivamente, em que $0 < k < 1$. Admitindo símbolos equiprováveis e sendo E_b a energia média por símbolo, então $x_1(n) = x_{11}s_1(n)$ e $x_2(n) = x_{21}s_1(n)$ com

$$x_{11} = \sqrt{\frac{2E_b}{1+k}} \text{ e } x_{21} = \sqrt{\frac{2kE_b}{1+k}}. \quad (4.5)$$

- **Chaveamento liga-desliga caótico (COOK - Chaotic On-Off Keying).** Caso particular do CSK unipolar para $k = 0$. No COOK $x_1(n) = \sqrt{2E_b}s_1(n)$ e $x_2(n) = 0$ (KOLUMBÁN; KENNEDY; CHUA, 1998). Equivalentemente,

$$x_{11} = \sqrt{2E_b} \text{ e } x_{21} = 0. \quad (4.6)$$

- **CSK bipolar.** No CSK bipolar (KENNEDY; ROVATTI; SETTI, 2000) $x_1(n) = x_{11}s_1(n)$ e $x_2(n) = x_{21}s_1(n)$ em que

$$x_{11} = \sqrt{E_b} \text{ e } x_{21} = -\sqrt{E_b}. \quad (4.7)$$

Para exemplificar a modulação CSK utilizando uma seqüência de base, considere o mapa tenda definido pela Eq. (2.9), repetida aqui por conveniência

$$s(n+1) = f_T(s(n)) = 1 - 2|s(n)|, \quad n \geq 0, \quad s[0] \in [-1, 1]. \quad (4.8)$$

Mostrou-se no final da Seção 2.6, página 62, que, para o mapa tenda, que pertence à família

dos mapas tenda inclinada, a função densidade invariante $p_*(s)$ que mede a probabilidade de um dado ponto de sua órbita estar numa vizinhança de s é uniforme:

$$p_*(s) = \begin{cases} \frac{1}{2}, & -1 \leq s \leq 1 \\ 0, & \text{caso contrário} \end{cases}. \quad (4.9)$$

Assim, o valor esperado da potência instantânea de um ponto dessa órbita é

$$E[s^2(n)] = \int_{-1}^1 s^2 \frac{1}{2} ds = \frac{1}{3}. \quad (4.10)$$

Portanto, para que um sinal constituído por N pontos dessa órbita tenha energia média unitária, deve-se multiplicar o sinal $s(n)$ em cada ponto por $\sqrt{3/N}$. Assim, para o estudo dos sistemas CSK com uma seqüência de base $s_1(n)$, esta pode ser obtida utilizando-se N pontos gerados através do mapa tenda $f_T(\cdot)$ com condição inicial distribuída uniformemente em $[-1, 1]$ e posteriormente multiplicados por $\sqrt{3/N}$.

A Figura 4.4 mostra os sinais transmitidos $x(n)$ para a seqüência de símbolos $\{1, 1, 0, 1, 0, 0, 1, 0\}$ utilizando cada um dos tipos de CSK discutidos com $N = 50$ amostras por símbolo e $E_b = 1$. Os símbolos “1” e “0” são transmitidos utilizando-se $x_1(n)$ e $x_2(n)$ respectivamente.

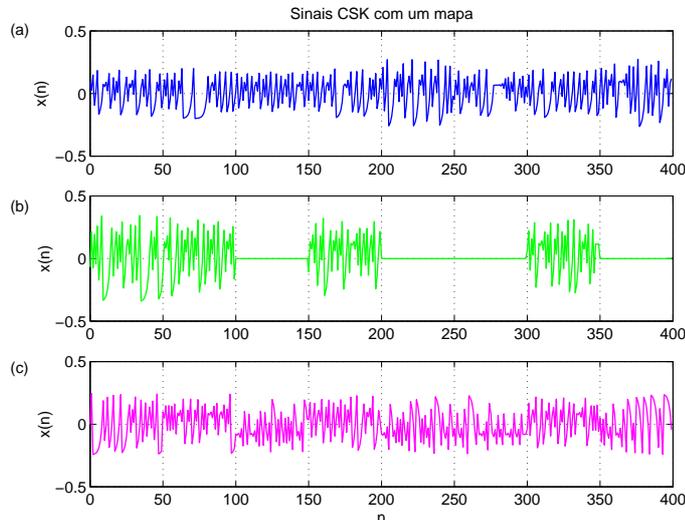


Figura 4.4: Sinais CSK transmitidos para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$: (a) CSK unipolar com $k = 0,5$; (b) COOK; (c) CSK bipolar. Em cada caso, $N = 50$ amostras e $E_b = 1$.

4.1.2 Demodulação do CSK com um mapa

No receptor CSK que emprega só um mapa, pode-se determinar o símbolo transmitido usando técnicas coerentes ou não-coerentes.

Receptor coerente

No receptor coerente por correlação para o CSK com uma função de base, mostrado esquematicamente na Figura 4.5, o valor esperado da componente observada z_{m1} é x_{m1} (Equação (4.3)), supondo que $\hat{s}_1(n) \approx s_1(n)$. Designa-se por z_T um nível de decisão adequado para cada esquema de modulação: se $z_{m1} > z_T$, o circuito de decisão estabelece $\hat{x}_{m1} = x_{11}$; se $z_{m1} < z_T$ então $\hat{x}_{m1} = x_{21}$.

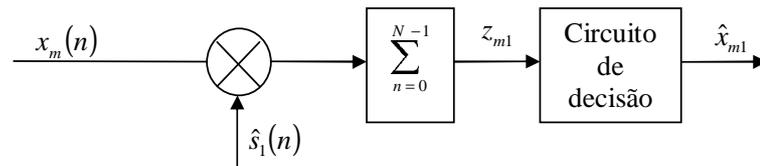


Figura 4.5: Receptor coerente para o CSK usando uma seqüência de base.

As Figuras 4.6(a)-(c) mostram os vetores de observação esperados na saída do correlator para os três esquemas de modulação CSK implementados com uma única função de base utilizando $E_b = 1$. Para um dado E_b fixo, supondo $\hat{s}_1(n) \approx s_1(n)$, a melhor discriminação é atingida pelo CSK bipolar seguido pelo COOK e pelo CSK unipolar, como se percebe na referida figura observando-se o afastamento entre os pontos \mathbf{z}_1 e \mathbf{z}_2

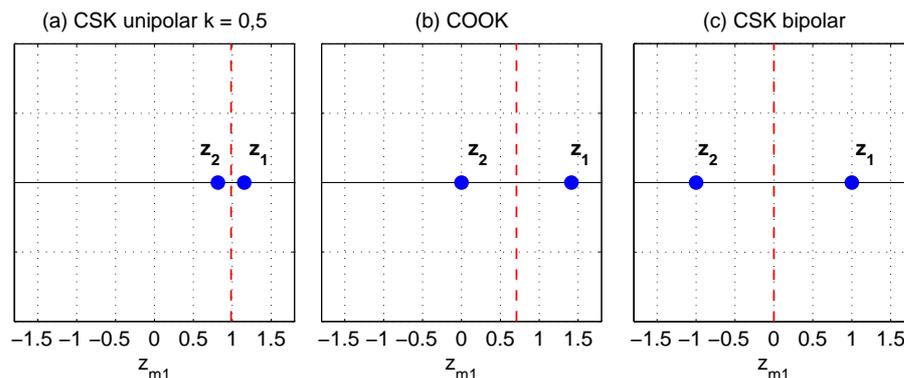


Figura 4.6: Valor esperado dos vetores de observação dos sistemas CSK com uma função de base e $E_b = 1$. (a) CSK unipolar com $k = 0,5$; (b) COOK; (c) CSK bipolar. A linha tracejada indica, em cada caso, o nível de decisão.

Deve-se ressaltar que no CSK bipolar os dois possíveis sinais transmitidos têm mesma energia. Assim, essa grandeza não pode ser usada para diferenciá-los no receptor. Como nos outros dois sistemas as energias médias de $x_1(n)$ e $x_2(n)$ são diferentes, pode-se utilizar demodulação não-coerente baseada na detecção da energia que é mais simples e apresenta melhores resultados em condições de baixa relação sinal-ruído, situação em que as seqüências de base dificilmente podem ser recuperadas.

Um requisito mínimo necessário do circuito do receptor coerente que gera a cópia local de $\hat{s}_1(n)$ é que consiga fazê-lo quer $x_{11}s_1(n)$ quer $x_{21}s_1(n)$ seja transmitido. Apesar de várias estratégias propostas sob o nome de *sincronização caótica* para recuperar a seqüência de base $s_1(n)$ nessas condições (PECORA; CARROLL, 1990; PARLITZ et al., 1992; MILLERIOUX; DAAFOUZ, 2001), essa abordagem da sincronização caótica não tem sido implementada adequadamente quando há ruído no canal ou descasamento de parâmetros entre transmissor e receptor o que compromete a aplicabilidade desses sistemas em situações reais (EISENCRAFT, 2001; EISENCRAFT; GERKEN, 2000).

Receptor não-coerente

As energias de $x_1(n)$ e $x_2(n)$ são diferentes nos sistemas de modulação COOK e CSK unipolar, o que permite a demodulação não-coerente de forma simples. O demodulador mostrado esquematicamente na Figura 4.7 estima a energia do sinal utilizado para transmitir cada símbolo. Note que o demodulador realiza uma autocorrelação e não a correlação cruzada com a função de base recuperada $\hat{s}_1(n)$.

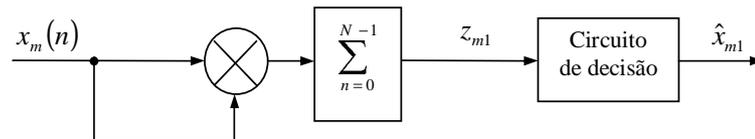


Figura 4.7: Receptor não-coerente para sinal CSK com uma função de base.

O valor esperado da componente de observação é

$$E[z_{m1}] = E\left[\sum_{n=0}^{N-1} x_m^2(n)\right] = E\left[x_{m1}^2 \sum_{n=0}^{N-1} s_1^2(n)\right] = x_{m1}^2 E\left[\sum_{n=0}^{N-1} s_1^2(n)\right] = x_{m1}^2. \quad (4.11)$$

A Figura 4.8 mostra histogramas de z_{m1} para o COOK e para o CSK unipolar com $k = 0,5$. Em cada caso, foram transmitidos 10000 símbolos com energia média por símbolo $E_b = 1$. Para cada símbolo foram utilizadas $N = 200$ amostras geradas a partir de $f_T(\cdot)$. Nestas condições, pelas Eqs. (4.5 - 4.6) e (4.11) para o COOK, $E[z_{11}] = 2$ e $E[z_{21}] = 0$ e para o CSK unipolar $E[z_{11}] = 1,333$ e $E[z_{21}] = 0,667$.

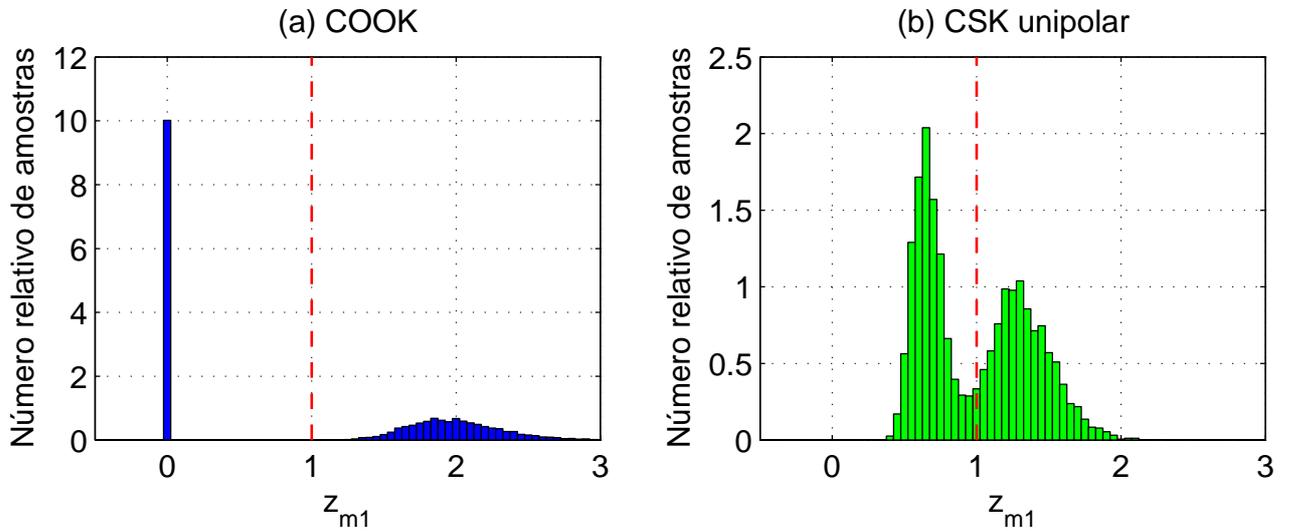


Figura 4.8: Histogramas de z_{m1} para (a) COOK e (b) CSK unipolar com $E_b = 1$ e $k = 0,5$: caso sem ruído.

É importante ressaltar que mesmo em condições ideais de canal, as amostras do sinal de observação não coincidem mas se concentram em torno de x_{m1}^2 , podendo haver erros na demodulação, pois o que é especificado é apenas a energia média para cada símbolo.

Esse receptor, apesar de não ser ótimo (LATHI, 1998), dispensa a recuperação de $s_1(n)$, evitando assim o problema da sincronização caótica independente da modulação.

4.1.3 Modulação CSK com dois mapas

Uma forma alternativa de sistema de comunicação binário CSK pode ser implementado utilizando dois mapas ou funções de base eliminando a necessidade de recuperar as seqüências de base caóticas independentemente da modulação. Neste caso,

$$x_m(n) = x_{m1}s_1(n) + x_{m2}s_2(n), \quad m = 1, 2, \dots, M; \quad n = 0, 1, \dots, N - 1. \quad (4.12)$$

Um diagrama de blocos de um modulador CSK desse tipo é mostrado na Figura 4.1 da página 80.

Um caso binário ($M = 2$) de especial importância é aquele em que as seqüências transmitidas são

$$x_1(n) = x_{11}s_1(n) \text{ e } x_2(n) = x_{22}s_2(n). \quad (4.13)$$

Neste caso, $\mathbf{x}_1 = [\sqrt{E_b}, 0]^T$ e $\mathbf{x}_2 = [0, \sqrt{E_b}]^T$.

Esse foi um dos primeiros sistemas de modulação caótica digital proposto na literatura, no início da década de 1990 (CUOMO; OPPENHEIM, 1993b; CUOMO; OPPENHEIM; STROGATZ, 1993; DEDIEU; KENNEDY; HASLER, 1993) e foi utilizado na experiência de campo relatada recentemente por Argyris et al. (2005). Porém, estes trabalhos não utilizam demodulação por correlação, que é ótima (LATHI, 1998).

4.1.4 Demodulação por correlação coerente do CSK com dois mapas

A demodulação coerente é a que melhor explora a utilização de duas funções de base. Um receptor de correlação como o mostrado na Figura 4.9 pode ser usado para estimar as componentes x_{mi} .

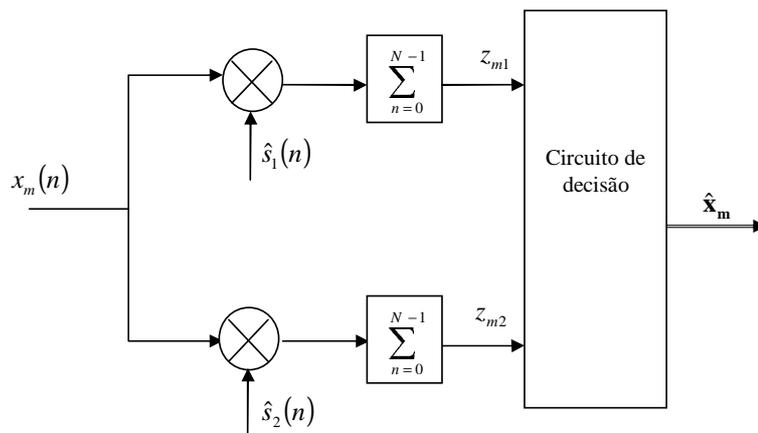


Figura 4.9: Receptor coerente para modulação CSK com duas funções de base.

Neste caso, supondo $\hat{s}_1(n) \approx s_1(n)$ e $\hat{s}_2(n) \approx s_2(n)$,

$$\begin{aligned} z_{m1} &= \sum_{n=0}^{N-1} x_m(n) s_1(n) = \sum_{n=0}^{N-1} (x_{m1} s_1(n) + x_{m2} s_2(n)) s_1(n) \\ &= x_{m1} \sum_{n=0}^{N-1} s_1^2(n) + x_{m2} \sum_{n=0}^{N-1} s_1(n) s_2(n) \end{aligned} \quad (4.14)$$

e

$$\begin{aligned} z_{m2} &= \sum_{n=0}^{N-1} x_m(n) s_2(n) = \sum_{n=0}^{N-1} (x_{m1} s_1(n) + x_{m2} s_2(n)) s_2(n) \\ &= x_{m1} \sum_{n=0}^{N-1} s_1(n) s_2(n) + x_{m2} \sum_{n=0}^{N-1} s_2^2(n). \end{aligned} \quad (4.15)$$

Levando-se em conta a ortonormalidade em média das seqüências de base $s_1(n)$ e $s_2(n)$, as componentes x_{mi} do vetor de sinal podem ser estimadas diretamente, pois

$$E[z_{m1}] = x_{m1} \text{ e } E[z_{m2}] = x_{m2}. \quad (4.16)$$

A decisão de qual o símbolo transmitido é simplificada consideravelmente no caso em que x_{12} e x_{21} são nulos. Neste caso, o valor esperado na saída dos correlatores é

$$E[z_{11}] = x_{11} E \left[\sum_{n=0}^{N-1} s_1^2(n) \right] = \sqrt{E_b} \quad (4.17)$$

$$E[z_{12}] = x_{12} E \left[\sum_{n=0}^{N-1} s_1(n) s_2(n) \right] = 0, \quad (4.18)$$

quando $x_1(n)$ é transmitido e

$$E[z_{21}] = x_{21} E \left[\sum_{n=0}^{N-1} s_1(n) s_2(n) \right] = 0 \quad (4.19)$$

$$E[z_{22}] = x_{22} E \left[\sum_{n=0}^{N-1} s_2^2(n) \right] = \sqrt{E_b}, \quad (4.20)$$

quando $x_2(n)$ é transmitido. Se $z_{m1} > z_{m2}$ então o circuito de decisão faz $\hat{\mathbf{x}}_m = \mathbf{x}_1$ e no caso contrário adota $\hat{\mathbf{x}}_m = \mathbf{x}_2$.

Os valores esperados para os vetores de observação \mathbf{z}_m para este sistema são mostrados na Figura 4.10 para $E_b = 1$. Note que a distância euclidiana entre \mathbf{z}_1 e \mathbf{z}_2 é $\sqrt{2E_b}$ que é menor do que a do CSK bipolar com uma seqüência de base, $2\sqrt{E_b}$. Isto significa que para obter o mesmo desempenho perante ruído, o CSK bipolar precisa, em princípio, de uma energia por símbolo menor. Porém, pela utilização de dois mapas, a sincronização caótica neste caso é simplificada.

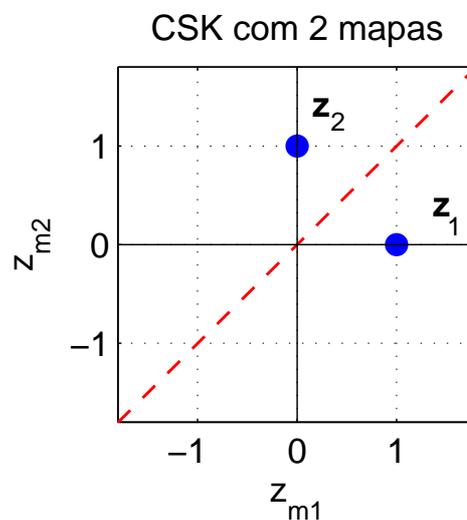


Figura 4.10: Valores esperados dos vetores de observação para demodulação coerente do CSK com duas seqüências de base e $E_b = 1$. A linha tracejada indica o limite teórico de decisão.

Tanto o CSK com duas funções de base quanto o FSK coerente convencional (PROAKIS, 1995) têm como distância entre os vetores de observação $\sqrt{2E_b}$. Porém, o FSK não sofre influência da sincronização caótica e da variabilidade da energia por símbolo presente no CSK. Esses dois fatores podem diminuir o desempenho do CSK comparado ao FSK.

4.2 O chaveamento caótico diferencial (DCSK)

O chaveamento caótico diferencial (DCSK) é uma variante do CSK com dois mapas em que as seqüências de base consistem de segmentos de formas de onda caótica repetidos. Como o DCSK é uma realização do CSK com dois mapas, ele pode utilizar a demodulação coerente, além de uma técnica diferencial simples.

4.2.1 Modulação DCSK

Assim como na forma mais simples do CSK binário com dois mapas (Seção 4.1.3), no DCSK, os sinais transmitidos são:

$$x_m(n) = x_{m1}s_1(n) + x_{m2}s_2(n), \quad m = 1, 2; \quad n = 0, 1, \dots, N - 1, \quad (4.21)$$

em que

$$\mathbf{x}_1 = [\sqrt{E_b}, 0]^T \quad e \quad \mathbf{x}_2 = [0, \sqrt{E_b}]^T. \quad (4.22)$$

No caso do DCSK, as seqüências de base ortogonais têm a forma

$$s_1(n) = \begin{cases} s(n), & 0 \leq n < \frac{N}{2} \\ s(n - \frac{N}{2}), & \frac{N}{2} \leq n < N \end{cases} \quad e \quad s_2(n) = \begin{cases} s(n), & 0 \leq n < \frac{N}{2} \\ -s(n - \frac{N}{2}), & \frac{N}{2} \leq n < N \end{cases} \quad (4.23)$$

em que $E[s_1(n)s_1(n)] = 1$, $s(n)$ é um trecho de sinal caótico e N é par. A primeira metade da função de base é chamada de *trecho de referência* e a segunda de *trecho de informação*, como mostrado de forma esquemática na Figura 4.11.

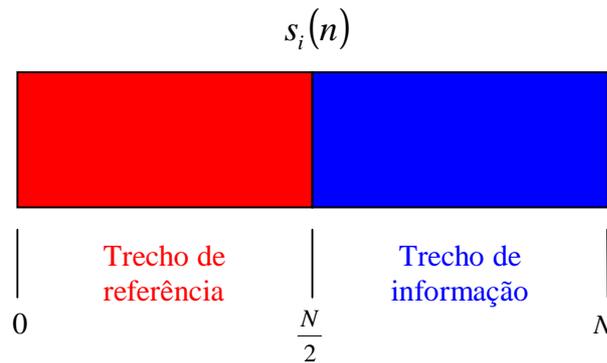


Figura 4.11: Estrutura dos sinais $s_i(n)$ usados no DCSK.

Substituindo as Eqs.(4.22 - 4.23) na Eq. (4.21), pode-se expressar o sinal DCSK como

$$x_m(n) = \begin{cases} \sqrt{E_b}s(n), & 0 \leq n < \frac{N}{2} \\ \sqrt{E_b}(-1)^{m+1}s(n - \frac{N}{2}), & \frac{N}{2} \leq n < N \end{cases} \quad (4.24)$$

A Figura 4.12 mostra o diagrama de blocos de um modulador DCSK.

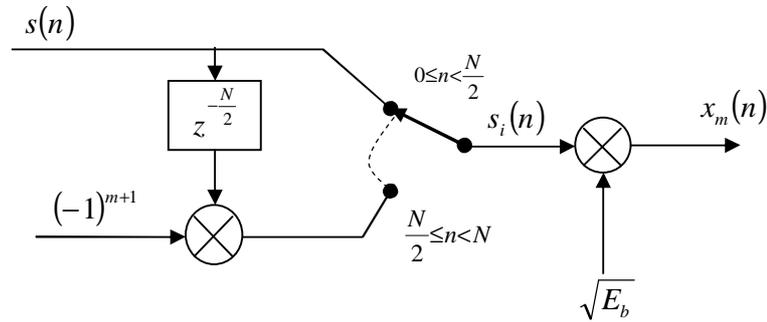


Figura 4.12: Diagrama de blocos de um modulador DCSK.

Um típico sinal DCSK $x(n)$ correspondente à seqüência de símbolos $\{1, 1, 0, 1, 0, 0, 1, 0\}$ utilizando o mapa tenda $f_T(\cdot)$ como gerador caótico é mostrada na Figura 4.13. Os símbolos “1” e “0” são transmitidos utilizando-se $x_1(n)$ e $x_2(n)$ respectivamente. O número de amostras por símbolo é $N = 50$ e $E_b = 1$.

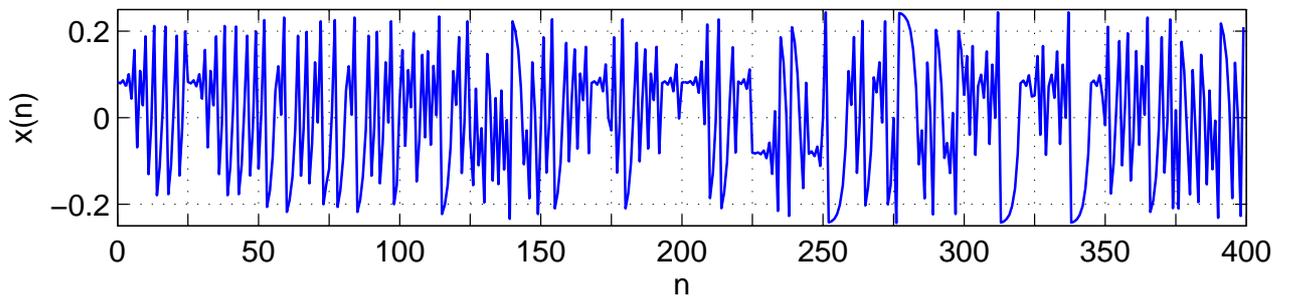


Figura 4.13: Sinal DCSK transmitido para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$ com $N = 50$ amostras e $E_b = 1$.

4.2.2 Demodulação DCSK

Como no sinal DCSK a informação é mapeada na correlação entre os trechos de referência e de informação, é possível demodular o sinal com um receptor diferencial (LATHI, 1998), além do receptor por correlação coerente.

Receptor coerente

Como já foi visto, o DCSK é um caso particular do CSK com duas funções de base e assim pode ser demodulado pelo receptor coerente mostrado na Figura 4.9 da página 87. Para os

sinais $x_1(n)$ e $x_2(n)$, os valores esperados dos vetores de observação são

$$E[\mathbf{z}_1] = [\sqrt{E_b}, 0]^T \quad e \quad E[\mathbf{z}_2] = [0, \sqrt{E_b}]^T \quad (4.25)$$

respectivamente (Eqs. (4.17 - 4.20)).

O diagrama mostrado na Figura 4.10 para chaveamento caótico coerente vale para o DCSK coerente também. Novamente, o desempenho do FSK coerente convencional é um limitante para este sistema.

Receptor diferencial

A estrutura das funções de base do DCSK torna possível a demodulação avaliando-se a correlação entre os trechos de referência e de informação.

O diagrama de blocos de um receptor DCSK diferencial é mostrado na Figura 4.14. O sinal recebido é atrasado de metade de sua duração $N/2$ e a correlação entre o sinal recebido e sua versão atrasada é determinada. A variável de observação z_{m1} é obtida amostrando-se a saída do correlator no instante $(N - 1)$.

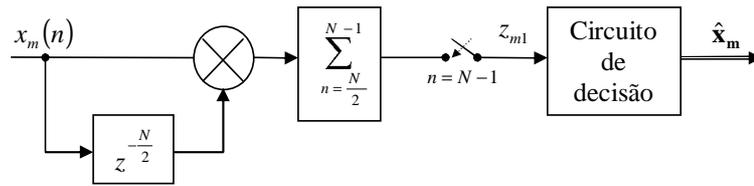


Figura 4.14: Diagrama de blocos de um receptor DCSK diferencial.

Neste caso,

$$\begin{aligned} z_{m1} &= \sum_{n=N/2}^{N-1} x_m(n)x_m(n - N/2) \\ &= \sum_{n=N/2}^{N-1} E_b s_m(n)s_m(n - N/2). \end{aligned} \quad (4.26)$$

Considerando que $E\left[\sum_{n=N/2}^{N-1} s_m^2(n)\right] = 1/2$, obtém-se

$$E[z_{11}] = +E_b/2 \quad e \quad E[z_{21}] = -E_b/2. \quad (4.27)$$

e assim conclui-se que a decisão de qual símbolo foi transmitido pode ser feita por um simples comparador de nível com referência no zero. Note que a distância euclidiana entre \mathbf{z}_1 e \mathbf{z}_2 é E_b .

Os valores esperados do vetor de observação do receptor diferencial para o DCSK são mostrados na Figura 4.15 ($E_b = 1$).

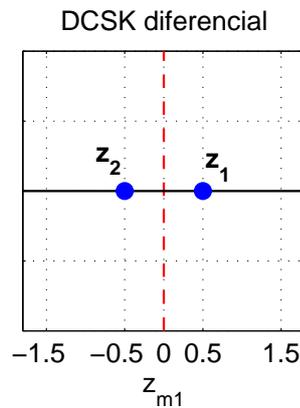


Figura 4.15: Valores esperados para os vetores de observação para um receptor DCSK diferencial.

A Figura 4.16 mostra um gráfico da saída do correlator da Figura 4.14 para o sinal transmitido da Figura 4.13 correspondente à seqüência de símbolos $\{1, 1, 0, 1, 0, 0, 1, 0\}$ para $E_b = 1$ e $N = 50$. A componente observada é o valor que a saída do correlator assume nos instantes $N - 1$, $2N - 1$, e assim sucessivamente. Esses valores são próximos de $+0,5$ nos casos em que o símbolo transmitido é “1” e $-0,5$ quando o símbolo transmitido é “0”, como esperado pela Eq. (4.27). O histograma das amostras do sinal de observação z_{m1} na ausência de ruído no canal é mostrado na Figura 4.17. Note que as amostras estão agrupadas em torno de $+E_b/2$ e $-E_b/2$.

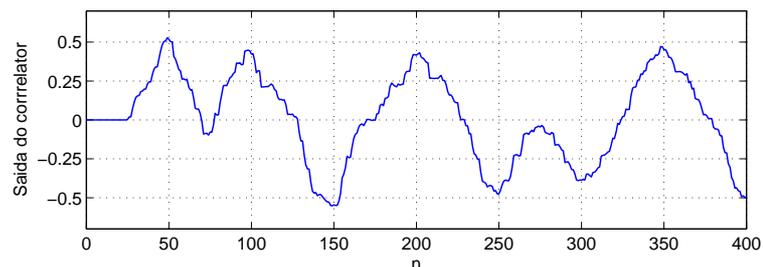


Figura 4.16: Saída do correlator DCSK diferencial para os símbolos $\{1, 1, 0, 1, 0, 0, 1, 0\}$. Novamente $E_b = 1$, $N = 50$ e o mapa tenda $f_T(\cdot)$ foi utilizado como gerador do sinal caótico.

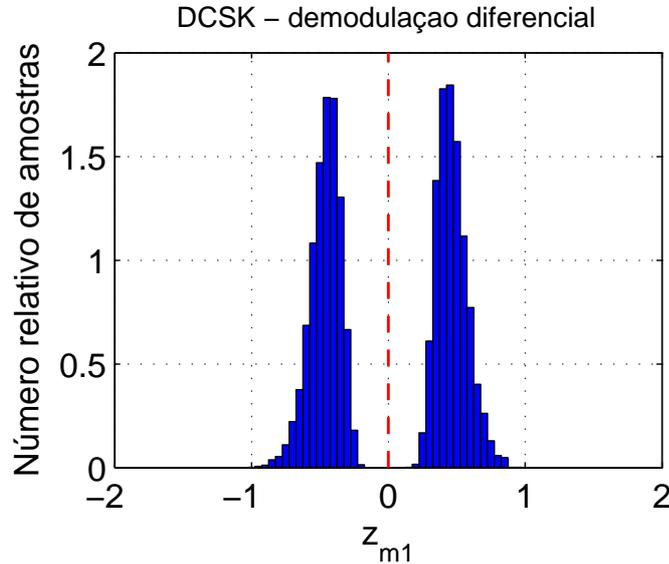


Figura 4.17: Histograma de z_{m1} para um receptor DCSK diferencial. Caso sem ruído e $E_b = 1$.

Comparando o CSK e o DCSK coerente com o DCSK com demodulação diferencial, o último apresenta as vantagens de independer dos problemas da sincronização caótica e ser pouco sensível a distorções do canal já que esse afeta tanto o trecho de referência quanto o de informação. Além disso, tem como nível de decisão $z_T = 0$ independente da potência de ruído no canal, o que não ocorre no CSK não-coerente.

Porém, o DCSK deixa muito a desejar quando comparado aos sistemas convencionais de modulação digital. Mesmo em situação ideal, esse sistema pode apresentar erros de recepção devido à variabilidade da energia por símbolo que faz com que a componente observada apresente variância não-nula. Esse problema é analisado em mais detalhes a seguir.

4.3 O problema da variabilidade da energia por símbolo

Considere um esquema de modulação convencional utilizando apenas uma função de base periódica, $s_1(n)$. Adota-se um número de amostras por símbolo N múltiplo inteiro do período de $s_1(n)$. Desta forma, a energia por símbolo

$$E_{bm} = x_{m1}^2 \sum_{n=0}^{N-1} s_1^2(n) \quad (4.28)$$

é constante para cada símbolo distinto. Em contrapartida, sinais caóticos são por definição aperiódicos. Assim, ao se utilizar uma base caótica, $s_1(n)$ é diferente a cada intervalo e

$$E_{s_1} = \sum_{n=0}^{N-1} s_1^2(n) \quad (4.29)$$

é diferente a cada símbolo transmitido.

A Figura 4.18 mostra histogramas de E_{s_1} para sinais $s_1(n)$ periódicos e caóticos. No histograma (a) considerou-se trechos do sinal periódico $s_1(n) = \frac{1}{2} \cos\left(\frac{\pi n}{4}\right)$ tendo-se $N = 8$ pontos que é o seu período; no (b) trechos de órbitas do mapa tenda normalizado, utilizando-se $N = 10$ e por fim, no histograma (c) considerou-se trechos de órbita do mapa tenda normalizado, utilizando-se $N = 100$ pontos.

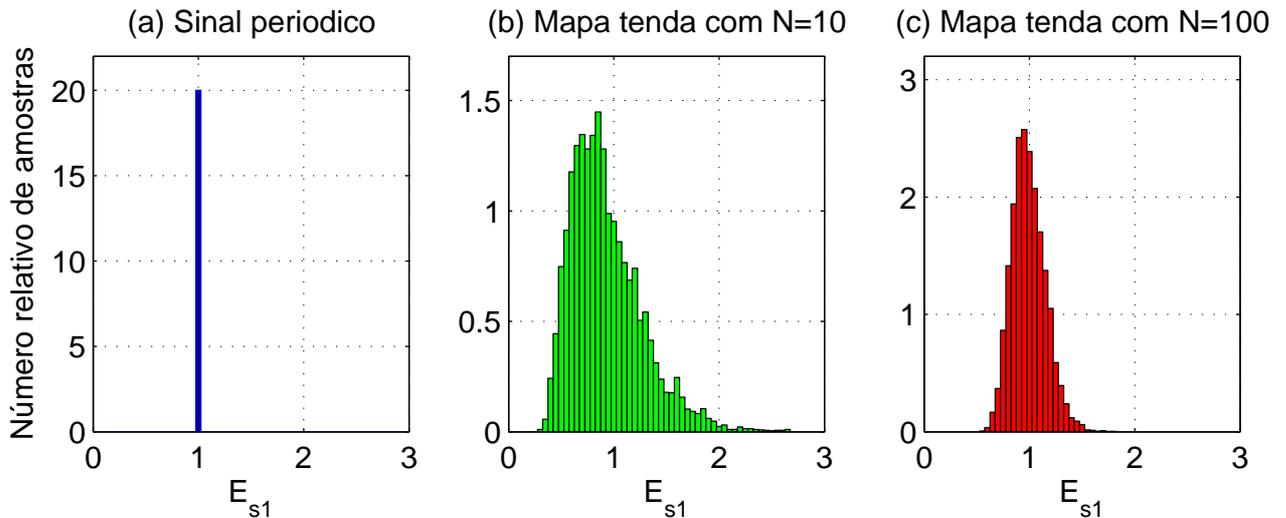


Figura 4.18: Histograma da energia de $s_1(n)$:(a) periódico; (b) órbita do mapa tenda normalizado com $N = 10$ e (c) órbita do mapa tenda normalizado com $N = 100$. Em cada caso foram considerados 10000 realizações.

No caso periódico, todos os valores de E_{s_1} são iguais a $E \left[\sum_{n=0}^{N-1} s_1^2(n) \right]$ com variância nula. No caso caótico, os valores de E_{s_1} estão concentrados em torno de $E \left[\sum_{n=0}^{N-1} s_1^2(n) \right]$ com variância não-nula.

Na Figura 4.19 mostra-se a evolução da média e do desvio-padrão de E_{s_1} em termos do número de pontos da seqüência para o mapa tenda normalizado. Para o cálculo tomou-se 10000 seqüências para cada N . Pode-se ver nesta figura que mesmo quando N é grande o desvio-padrão ainda se apresenta de forma importante e não decai abaixo de certo nível. Ou seja, a

partir de aproximadamente $N = 50$, aumentar o número de pontos utilizados na transmissão de um símbolo não reduz o problema de imprecisão da estimação da energia.

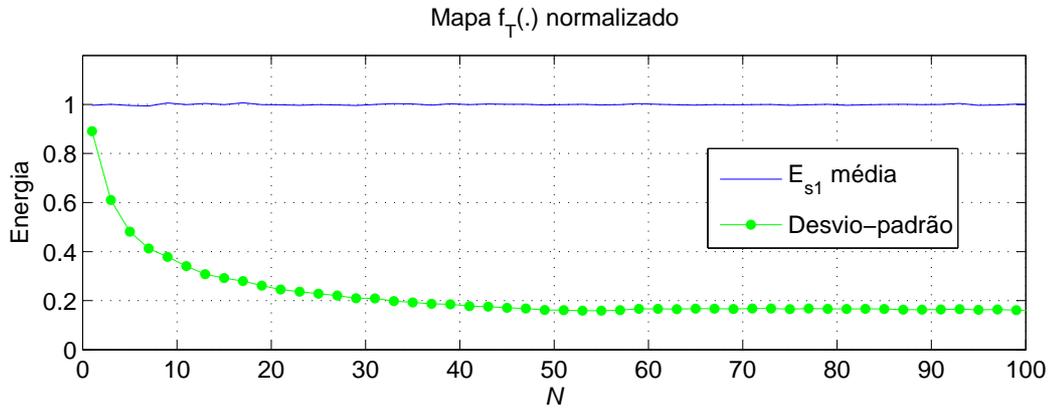


Figura 4.19: Média e desvio-padrão da estimação de E_{s1} em função de N para o mapa tenda normalizado.

O fato da energia por símbolo não ser constante é uma das principais desvantagens dos sistemas de comunicação utilizando sinais caóticos discutidos até aqui quando comparados aos sistemas convencionais. Pelas Figuras 4.8 e 4.17 vê-se que para estes sistemas podem ocorrer erros na recepção mesmo em condições ideais de canal, o que é indesejável na prática. Aumentar o número de pontos utilizados N não resolve a desvantagem já que o desvio-padrão na estimação de E_{s1} tem queda lenta. Além disso o aumento do número de pontos transmitidos por símbolo limita a taxa de transmissão máxima.

Uma solução alternativa é modificar o esquema de modulação de forma que a energia transmitida a cada símbolo seja mantida constante. O FM-DCSK (KOLUMBÁN et al., 1998), discutido em detalhes a seguir é um exemplo dessa última abordagem.

4.4 O FM-DCSK

O transmissor FM-DCSK gera um sinal DCSK com energia por símbolo constante. A idéia é aproveitar o fato de que a potência de um sinal modulado em frequência independe do sinal, desde que este varie lentamente em relação à portadora (LATHI, 1998). Desta forma, insere-se o sinal caótico na entrada de um modulador em frequência. Se a saída desse modulador for utilizada na implementação do DCSK, então a saída do correlador no receptor terá variância

nula no caso sem ruído e o problema de imprecisão da estimação da energia está solucionado.

A operação do modulador FM-DCSK é análoga à do DCSK, sendo que a única diferença é que as funções de base são construídas com um sinal caótico modulado em frequência. A Figura 4.20 ilustra o princípio do FM-DCSK em tempo contínuo, considerando T a duração de um símbolo.

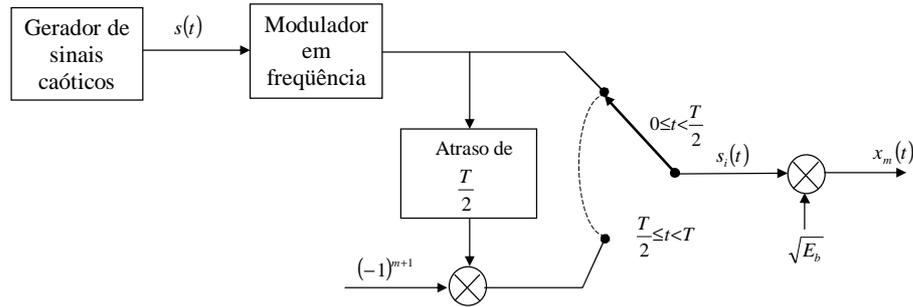


Figura 4.20: Diagrama de blocos de um modulador FM-DCSK em tempo contínuo.

Para efeitos de simulação e análise, é importante obter o modelo equivalente em tempo discreto do sistema FM-DCSK. Para isso, considera-se o trecho de referência do sinal FM-DCSK que é descrito por

$$x_m(t) = A \cos [2\pi (f_0 - K_f s(t)) t], \quad 0 \leq t < \frac{T}{2}, \quad (4.30)$$

sendo T a duração de um símbolo, f_0 a frequência da portadora, A e K_f constantes (LATHI, 1998). A constante K_f que define a intensidade da modulação é considerada unitária no desenvolvimento seguinte. Desta forma, pode-se reescrever a Eq. (4.30) como

$$x_m(t) = A [\cos(2\pi s(t)t) \cos(2\pi f_0 t) + \sin(2\pi s(t)t) \sin(2\pi f_0 t)], \quad 0 \leq t \leq \frac{T}{2}. \quad (4.31)$$

Portanto, de acordo com a Eq. (3.3) da página 70, o equivalente passa-baixas desse trecho de $x_m(t)$ é

$$x_{m_i}(t) = A [\cos(2\pi s(t)t) + j \sin(2\pi s(t)t)]. \quad (4.32)$$

Amostrando esse sinal com um período de amostragem conveniente T_a (Eq. (3.5) página 70) e

denotando $x_m(n) = x_{m_i}(nT_a)$, obtém-se o equivalente passa-baixas de tempo discreto

$$x_m(n) = A [\cos(2\pi s(n)n) + j \sin(2\pi s(n)n)] = Ae^{j2\pi ns(n)}, \quad 0 \leq n < N/2. \quad (4.33)$$

Repare que $|x_m(n)| = A$ para qualquer n . Assim, para que um símbolo seja representado com um sinal com energia E_b adota-se $A = \sqrt{\frac{E_b}{N}}$ e

$$x_m(n) = \sqrt{\frac{E_b}{N}} e^{j2\pi ns(n)}, \quad 0 \leq n < N/2. \quad (4.34)$$

O trecho de informação estende-se de $N/2 \leq n < N$ e repete o trecho de referência para $m = 1$ ou é igual ao oposto do trecho de referência para $m = 2$.

Usando a notação utilizada até aqui, para representar o FM-DCSK, tem-se

$$x_m(n) = x_{m1}s_1(n) + x_{m2}s_2(n) \quad (4.35)$$

em que

$$\mathbf{x}_1 = [\sqrt{E_b}, 0]^T \quad \text{e} \quad \mathbf{x}_2 = [0, \sqrt{E_b}]^T, \quad (4.36)$$

sendo as funções de base

$$s_1(n) = \begin{cases} \frac{1}{\sqrt{N}} e^{j2\pi ns(n)}, & 0 \leq n < \frac{N}{2} \\ \frac{1}{\sqrt{N}} e^{j2\pi(n-\frac{N}{2})s(n-\frac{N}{2})}, & \frac{N}{2} \leq n < N \end{cases} \quad \text{e} \quad (4.37)$$

$$s_2(n) = \begin{cases} \frac{1}{\sqrt{N}} e^{j2\pi ns(n)}, & 0 \leq n < \frac{N}{2} \\ -\frac{1}{\sqrt{N}} e^{j2\pi(n-\frac{N}{2})s(n-\frac{N}{2})}, & \frac{N}{2} \leq n < N \end{cases} \quad (4.38)$$

em que $s(n)$ é um sinal caótico e N é considerado par.

Substituindo as Eqs. (4.36 - 4.38) na Eq. (4.35), pode-se expressar o sinal FM-DCSK como

$$x_m(n) = \begin{cases} \sqrt{\frac{E_b}{N}} e^{j2\pi ns(n)}, & 0 \leq n < \frac{N}{2} \\ \sqrt{\frac{E_b}{N}} (-1)^{m+1} e^{j2\pi(n-\frac{N}{2})s(n-\frac{N}{2})}, & \frac{N}{2} \leq n < N \end{cases}. \quad (4.39)$$

Um diagrama de blocos de um modulador equivalente passa-baixas do FM-DCSK é mostrado na Figura 4.21.

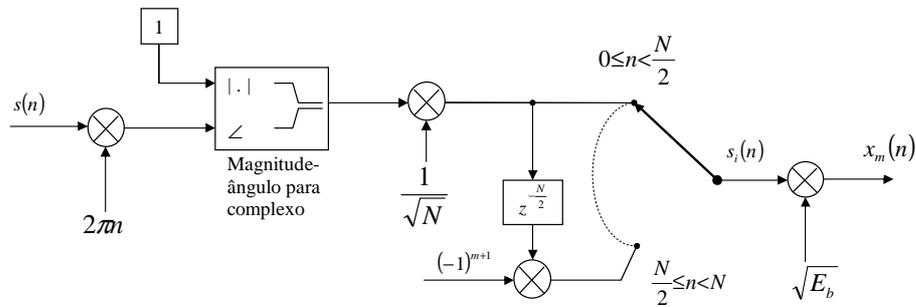


Figura 4.21: Diagrama de blocos de um modulador FM-DCSK em tempo discreto.

A Figura 4.22 mostra a parte real e a parte imaginária do equivalente passa-baixas $x(n)$ do sinal transmitido para a seqüência de símbolos $\{1, 1, 0, 1, 0, 0, 1, 0\}$ utilizando $N = 50$ amostras por símbolo e $E_b = 1$. Novamente os símbolos “1” e “0” são transmitidos utilizando $x_1(n)$ e $x_2(n)$ respectivamente.

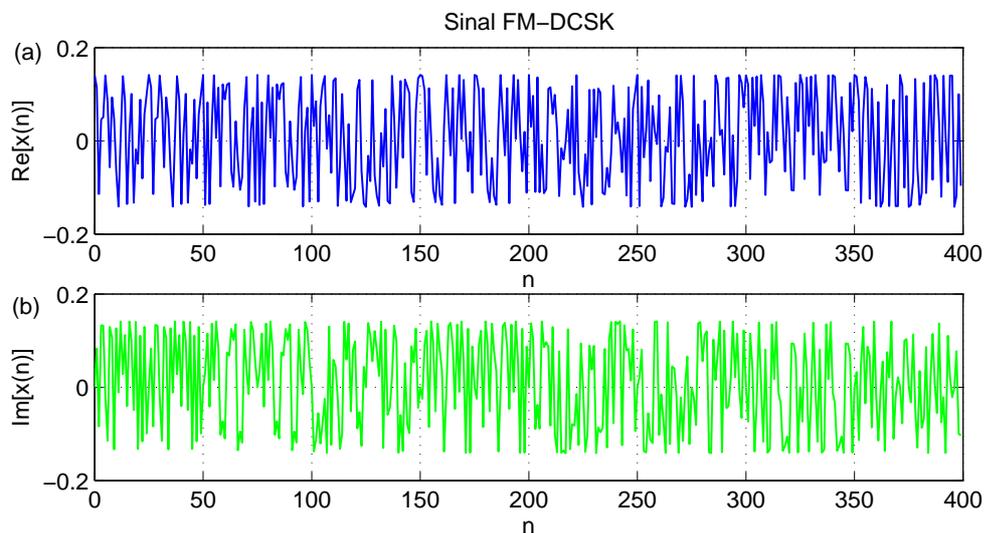


Figura 4.22: Parte real e parte imaginária do sinal FM-DCSK transmitido para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$ com $N = 50$ amostras e $E_b = 1$.

Note que a única diferença entre os esquemas de modulação FM-DCSK e DCSK é que os sinais caóticos modulados em FM são usados como funções de base ao invés dos próprios sinais caóticos. Com isto, a energia das funções de base torna-se constante, como ocorre nos esquemas de modulação convencionais. Conseqüentemente, o problema de imprecisão na estimação da energia não aparece.

A mesma idéia pode ser aplicada nos sistemas de modulação CSK utilizando apenas uma função de base. Porém o problema do nível de decisão dependente do ruído permanece.

Um receptor FM-DCSK simples pode ser construído usando o demodulador diferencial mostrado na Figura 4.14 da página 92. Note que no FM-DCSK a correlação entre os trechos de informação e de referência do sinal modulado em frequência carrega a informação; sendo assim, não há necessidade de demodulação em frequência no receptor. A entrada do demodulador DCSK diferencial é o sinal FM-DCSK neste caso.

O uso de um sinal caótico modulado em frequência como função de base tem um efeito interessante na implementação do sistema. Sinais gerados por mapas caóticos são quase exclusivamente sinais passa-baixas. No entanto, para a comunicação via rádio só podem ser usados sinais passa-banda na faixa de radiofrequência (RF). Por isso é necessário um modulador de RF convencional extra para gerar o sinal passa-bandas em RF para transmissão.

No entanto, se um sinal caótico modulado em frequência for usado como portadora, como no caso do FM-DCSK, a saída do modulador caótico já é um sinal passa-banda em RF e modulador e demodulador de RF extras tornam-se desnecessários (KENNEDY; ROVATTI; SETTI, 2000).

4.5 Comparação qualitativa do desempenho em canal AWGN

Nas seções anteriores mostrou-se que transmissões usando sinais caóticos podem ser demoduladas usando correlação de uma das seguintes formas:

- *recepção por correlação coerente*, em que as funções de base são recuperadas por sincronização,
- *recepção por correlação não-coerente*, em que a informação é extraída do sinal recebido a partir de autocorrelações,
- *recepção diferencial* em que a informação é extraída através da correlação entre os trechos de informação e de referência do sinal recebido.

A diferença entre estes esquemas reside no sinal utilizado para correlacionar com o sinal que chega ao receptor. Desta forma, as três configurações de receptores podem ser analisadas por meio de um diagrama de blocos comum conforme Figura 4.23.

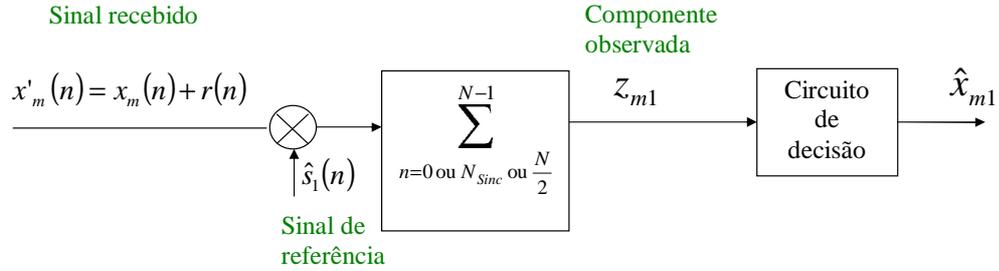


Figura 4.23: Diagrama de blocos genérico de um demodulador por correlação utilizando sinais caóticos com $N_b = 1$.

Comparando os receptores coerentes desenvolvidos para uma ou duas funções de base (Figuras 4.5 e 4.9 respectivamente), conclui-se que o diagrama de blocos geral da Figura 4.23 é válido para os dois casos. A única diferença é que, no caso de duas funções de base, dois correlatores devem ser usados.

Por simplicidade, no que segue, consideram-se apenas sistemas de transmissão binária utilizando uma única seqüência de base $s_1(n)$. Portanto os sinais transmitidos são $x_m(n) = x_{m1}s_1(n)$, $m = 1, 2$. Denota-se por $\hat{s}_1(n)$ o sinal de referência e por $x'_m(n)$ o sinal ruidoso que chega ao receptor:

$$x'_m(n) = x_m(n) + r(n) = x_{m1}s_1(n) + r(n). \quad (4.40)$$

sendo $r(n)$ um processo ruído branco gaussiano com média nula e potência σ_r^2 .

A decisão no receptor baseia-se na componente observada z_{m1} .

4.5.1 Receptor por correlação não-coerente

Em um receptor de correlação não-coerente, o sinal de referência $\hat{s}_1(n)$ é igual ao sinal que chega ao receptor $x'_m(n) = x_m(n) + r(n)$ e a componente observada pode ser expressa como

$$z_{m1} = \sum_{n=0}^{N-1} (x_m(n) + r(n))^2 = \sum_{n=0}^{N-1} x_m^2(n) + 2 \sum_{n=0}^{N-1} x_m(n)r(n) + \sum_{n=0}^{N-1} r^2(n). \quad (4.41)$$

No caso sem ruído, $r(n) = 0$ e o valor esperado da componente observada é igual à energia

média do símbolo transmitido, isto é,

$$E[z_{m1}] = x_{m1}^2 \left[\sum_{n=0}^{N-1} s_1(n)^2 \right] = x_{m1}^2. \quad (4.42)$$

A presença de ruído no canal faz com que a média da variável z_{m1} seja diferente de x_{m1}^2 . O valor esperado de z_{m1} depende do sinal caótico e da potência do ruído como expresso pela primeira e terceira parcela do segundo membro da Eq. (4.41), respectivamente. Neste caso z_{m1} é um estimador com viés da energia de $x_m(n)$ e o limiar do comparador usado como circuito de decisão depende explicitamente do nível de ruído.

Como exemplo, considerem-se os sistemas CSK unipolar com $k = 0,5$ e COOK com uma relação sinal-ruído (SNR - *Signal-to-Noise Ratio*) igual a 10dB no canal, sendo $E_b = 1$. Neste caso, a energia de $r(n)$ é $N\sigma_r^2 = 0,1$ e os valores esperados para z_{m1} podem ser obtidos usando as Eqs. (4.5), (4.6) e (4.41). Assim,

$$\text{CSKunipolar : } E[z_{11}] = \frac{2E_b}{1+k} + N\sigma_r^2 = 1,433; \quad E[z_{21}] = \frac{2kE_b}{1+k} + N\sigma_r^2 = 0,767. \quad (4.43)$$

$$\text{COOK : } E[z_{11}] = 2E_b + N\sigma_r^2 = 2,1; \quad E[z_{21}] = 0 + N\sigma_r^2 = 0,1. \quad (4.44)$$

As Figuras 4.24(c) e (d) mostram histogramas dos valores observados de z_m neste caso para 10000 símbolos transmitidos utilizando-se como gerador caótico o mapa tenda. Esses podem ser comparados aos histogramas das Figuras 4.24(a) e (b) que representam os casos sem ruído.

Entre as modulações usando recepção não-coerente, a maior distância entre os valores esperados de z_{m1} com um valor de E_b fixo é atingida pelo COOK. Neste caso, a distância entre z_{11} e z_{21} é $2E_b$.

O maior problema da recepção não-coerente é que o nível de decisão depende da potência do ruído, o que dificulta a implementação do circuito de decisão. A forma de produzir um estimador não-enviesado com maior distância entre as componentes observadas é o uso do CSK coerente e o DCSK.

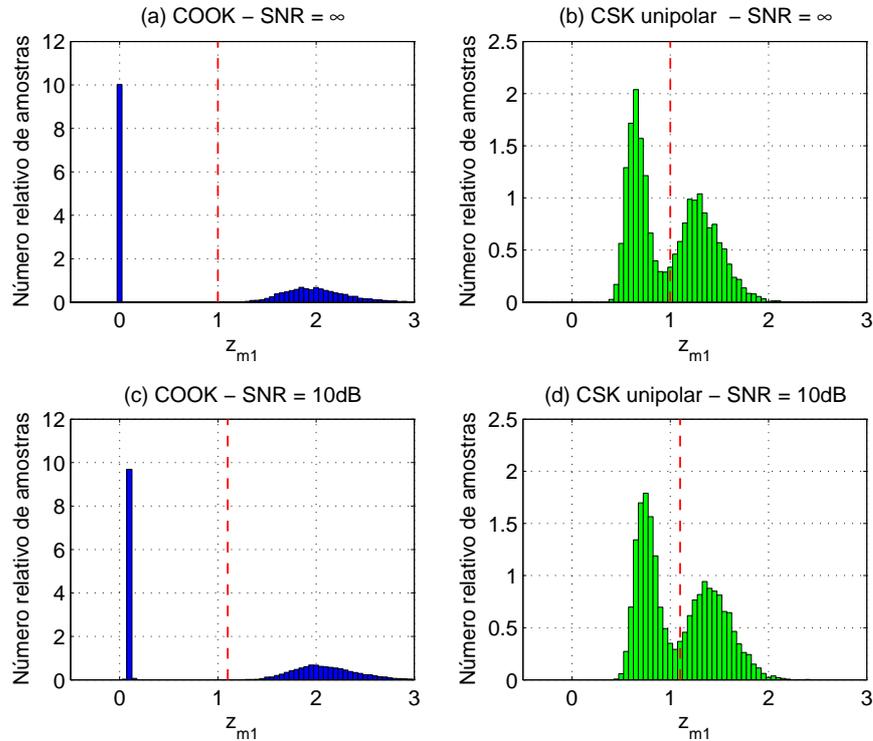


Figura 4.24: Histogramas da componente observada: (a) COOK e (b) CSK unipolar ($k = 0,5$) na ausência de ruído; (c) COOK e (d) CSK unipolar ($k = 0,5$) com $\text{SNR} = 10\text{dB}$. Em cada caso, $E_b = 1$ e 10000 símbolos foram transmitidos.

4.5.2 Receptor por correlação coerente e a sincronização caótica

Seja $\hat{s}_1(n)$ a função de base caótica recuperada em que $\hat{s}_1(n) \approx s_1(n)$ para $n \geq N_{\text{Sinc}}$. Neste caso, na Figura 4.23, o sinal de referência é o sinal recuperado.

Uma desvantagem do receptor CSK coerente é que a sincronização é perdida e recuperada toda vez que o símbolo transmitido muda (DEDIEU; KENNEDY; HASLER, 1993). Portanto, N_{Sinc} das amostras que compõem cada símbolo são necessárias à sincronização. Assim, o tempo de sincronização limita superiormente a taxa de símbolos e, portanto, a taxa de dados.

Como o transitório de sincronização não pode ser utilizado para transmitir informação, a componente observada é obtida por correlação somente no intervalo $N_{\text{Sinc}} \leq n < N - 1$. Sejam $x_m(n)$, $m = 1, 2$, os sinais transmitidos para a modulação CSK binária com uma única

seqüência de base $s_1(n)$. Então, a componente observada é dada por:

$$\begin{aligned} z_{m1} &= \sum_{n=N_{Sinc}}^{N-1} [x_m(n) + r(n)] \hat{s}_1(n) = \sum_{n=N_{Sinc}}^{N-1} [x_{m1}s_1(n) + r(n)] \hat{s}_1(n) = \\ &= x_{m1} \sum_{n=N_{Sinc}}^{N-1} s_1(n)\hat{s}_1(n) + \sum_{n=N_{Sinc}}^{N-1} r(n)\hat{s}_1(n). \end{aligned} \quad (4.45)$$

Cabe observar que z_{m1} é uma *variável aleatória* cuja esperança depende da energia por símbolo do sinal transmitido e da qualidade da sincronização.

Supondo que $r(n)$ e $\hat{s}_1(n)$ sejam não-correlacionados, o valor médio de z_{m1} independe do ruído; assim, o receptor é um *estimador não-enviesado* de x_{m1} ; em particular, o nível de decisão no comparador não depende do ruído no canal.

Como no caso de receptores coerentes convencionais que utilizam funções de base periódicas, o desempenho perante AWGN de um receptor coerente usando funções de base caóticas é teoricamente ótimo (LATHI, 1998). No entanto, a taxa de erro de símbolo (SER) também depende da qualidade da sincronização, ou seja, da proximidade entre o sinal de referência $\hat{s}_1(n)$ e a função de base caótica original $s_1(n)$. Qualquer erro de sincronização, especialmente a perda de sincronização, resulta em uma grande degradação do desempenho (EISENCRAFT, 2001).

As técnicas de sincronização caótica até hoje publicadas são muito sensíveis a ruído. Em particular, as funções de base $s_i(n)$ não podem ser recuperadas exatamente quando $x'_m(n) \neq x_m(n)$ (EISENCRAFT; GERKEN, 2000; EISENCRAFT, 2001; WILLIAMS, 2001). Desta forma, os receptores baseados em sincronização caótica não são adequados para ambientes de propagação com baixa SNR.

O número de amostras N_{Sinc} necessário à sincronização é outro fator que degrada o desempenho desses sistemas perante ruído. Como a informação não pode ser transmitida durante o transitório de sincronização, a energia correspondente a esse trecho de sinal é perdida o que implica degradação na SER.

Assim, embora um receptor de correlação CSK coerente seja superior a um receptor não-coerente no sentido de que o primeiro provê um estimador não-enviesado, seu desempenho

depende criticamente da habilidade de regenerar as funções de base no receptor. As técnicas de sincronização caótica existentes são insuficientemente robustas para um sistema de comunicação prático (EISENCRAFT; GERKEN, 2000). Em contrapartida, a recepção diferencial do DCSK e do FM-DCSK provê um estimador não-enviesado que permite o uso de um simples comparador com nível de decisão em zero e dispensa sincronização caótica.

Diante dessas dificuldades, o restante deste trabalho concentra-se em modulações não-coerentes e diferenciais, conforme atual tendência da literatura (KENNEDY; ROVATTI; SETTI, 2000; LAU; TSE, 2003).

4.5.3 Receptor diferencial

Em um receptor DCSK ou FM-DCSK diferencial, o sinal de referência $\hat{s}(n)$ é uma versão atrasada do sinal ruidoso que chega ao receptor. Note que amostras diferentes de ruído corrompem as entradas do correlator. A componente observada é dada por

$$z_{m1} = \sum_{n=\frac{N}{2}}^{N-1} [x_m(n) + r(n)] \left[x_m \left(n - \frac{N}{2} \right) + r \left(n - \frac{N}{2} \right) \right] \quad (4.46)$$

Substituindo-se a Eq. (4.24) na Eq. (4.46) obtém-se a componente observada no caso DCSK como

$$\begin{aligned} z_{m1} = & (-1)^{m+1} E_b \sum_{n=\frac{N}{2}}^{N-1} s^2(n) + \sqrt{E_b} \sum_{n=\frac{N}{2}}^{N-1} r(n) s \left(n - \frac{N}{2} \right) + (-1)^{m+1} \sqrt{E_b} \sum_{n=\frac{N}{2}}^{N-1} s(n) r \left(n - \frac{N}{2} \right) \\ & + \sum_{n=\frac{N}{2}}^{N-1} r(n) r \left(n - \frac{N}{2} \right) \end{aligned} \quad (4.47)$$

pois foi usado o fato de que $s(n) = s \left(n - \frac{N}{2} \right)$ para $N/2 \leq n < N$.

Para o DCSK o valor médio da primeira parcela é $E_b/2$ ou $-E_b/2$. No caso equivalente do FM-DCSK, o valor é constante e igual a $E_b/2$ ou $-E_b/2$. Os outros três termos contendo a seqüência AWGN tem média nula. Isso mostra que z_{m1} é um estimador *não-enviesado* de $\pm E_b/2$ neste caso. O nível de decisão é o zero e independe do nível de ruído no canal.

No caso da modulação DCSK, a variância de z_{m1} é determinada pela variação da energia

por símbolo do sinal caótico e pelo ruído no canal. Por isso, a incerteza de estimação da energia também influencia o desempenho do DCSK.

Para o FM-DCSK, a primeira parcela da Eq. (4.47) é igual a $\pm E_b/2$ e a incerteza de estimação da energia não aparece, além do que o nível de decisão é fixo e não há necessidade da sincronização caótica. Nesse sentido, o FM-DCSK é superior aos demais esquemas de modulação caótica analisados em termos de desempenho em canal AWGN, como é comprovado pelas simulações computacionais da próxima seção.

4.6 Curvas de desempenho em canal AWGN

Até aqui, foram descritos algumas modulações caóticas digitais e estas foram comparadas qualitativamente em termos de desempenho em canal AWGN. Discutiu-se também o problema de precisão na determinação da energia por símbolo devido à natureza aperiódica dos sinais caóticos. Nesta seção, é feita análise através da simulação numérica desses sistemas com recepção não-coerente e diferencial. Os resultados obtidos são expressos na forma usual na literatura de comunicações digitais: por meio de curvas de taxa de erro de símbolo (SER) em função de E_b/N_0 (relação entre a energia média por símbolo pelo dobro da densidade espectral do AWGN, dada por $S_r(\omega) = N_0/2$) (LATHI, 1998; HAYKIN, 2000; PROAKIS, 1995).

As curvas mostradas aqui foram obtidas por meio de simulação computacional dos equivalentes em banda base em tempo discreto dos sistemas contínuos, conforme discutido no Capítulo 3. Os trechos de sinal caótico de comprimento N utilizados para representar cada símbolo foram obtidos utilizando o mapa tenda $f_T(\cdot)$ normalizado e condições iniciais aleatórias uniformemente distribuídas no intervalo $[-1, 1]$. Após a demodulação, a informação recebida é comparada com a emitida. Se forem diferentes, conta-se um erro. No cálculo da taxas de erro considerou-se como critério de parada a ocorrência de 50 erros no receptor. Por exemplo, uma SER de $2,5 \cdot 10^{-4}$ significa que 50 entre $(50/2,5 \cdot 10^{-4}) = 2 \cdot 10^5$ símbolos foram interpretados de forma errônea.

4.6.1 Curvas de SER para o CSK com recepção não-coerente

O pior desempenho em canal AWGN foi obtido usando o esquema de modulação CSK unipolar, cujo diagrama de blocos do demodulador é mostrado na Figura 4.7 da página 85. Este mau desempenho é esperado devido à variabilidade da energia por símbolo e a dependência do nível de decisão com a potência do ruído no canal. Além disso, os vetores de observação estão menos distanciados do que no caso COOK (vide Figura 4.6 da página 84). Suas curvas de SER por E_b/N_0 para diversos valores de número de pontos N utilizados na transmissão e $k = 0,5$ são mostradas na Figura 4.25.

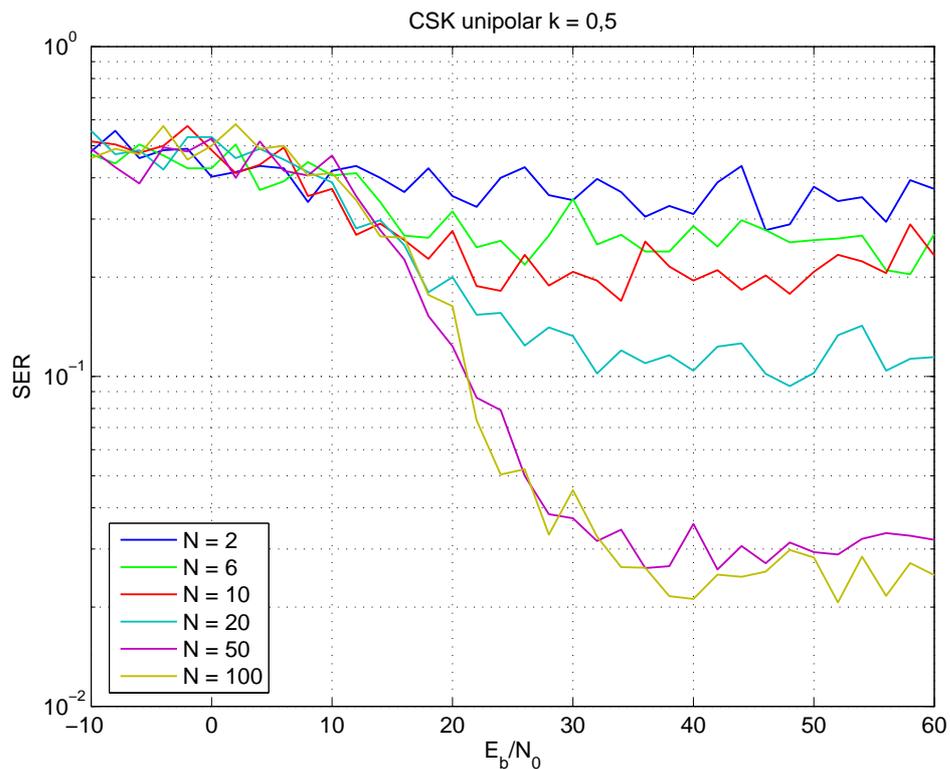


Figura 4.25: Taxas de erro de símbolo do CSK unipolar para $k = 0,5$.

Duas características chamam a atenção nessas curvas: a dependência da SER com o número de pontos N utilizados na transmissão e o mau desempenho do sistema mesmo para valores de E_b/N_0 próximos de 50dB.

A dependência com o comprimento da seqüência ou, no caso contínuo, com o tempo de símbolo, não ocorre nos sistemas de comunicação convencionais utilizando portadoras senoidais. Um resultado que é clássico para eles é que apenas a energia média utilizada por símbolo influi

no desempenho perante um canal AWGN (LATHI, 1998). Porém, no caso de portadoras caóticas, a energia utilizada na transmissão não é constante e deve ser estimada utilizando o sinal que chega ao receptor. A Figura 4.19 da página 96 mostra que, no caso do mapa tenda, até cerca de $N = 50$, a variância dessa estimativa diminui, o que justifica a melhora de desempenho do CSK conforme aumenta o valor de N até próximo desse valor.

Porém, a partir de aproximadamente $N = 50$ a variância da estimação da energia não decai mais tornando-se constante. Essa variância é que provoca os erros na determinação do símbolo transmitido mesmo em situações de E_b/N_0 e N altos. Por exemplo, há pouca diferença entre as SER obtidas para $N = 50$ e $N = 100$. De forma geral, o desempenho do CSK unipolar para este valor de k está longe do aceitável para aplicações práticas de telecomunicações.

O COOK é um caso especial do CSK unipolar em que a separação entre as componentes observadas é maximizada. A Figura 4.26 mostra as curvas de SER obtidas para este sistema. Novamente observa-se a dependência do desempenho com N . Devido à maior separação entre a energia dos sinais utilizados para representar cada símbolo, os resultados são melhores do que os do CSK com $k = 0,5$.

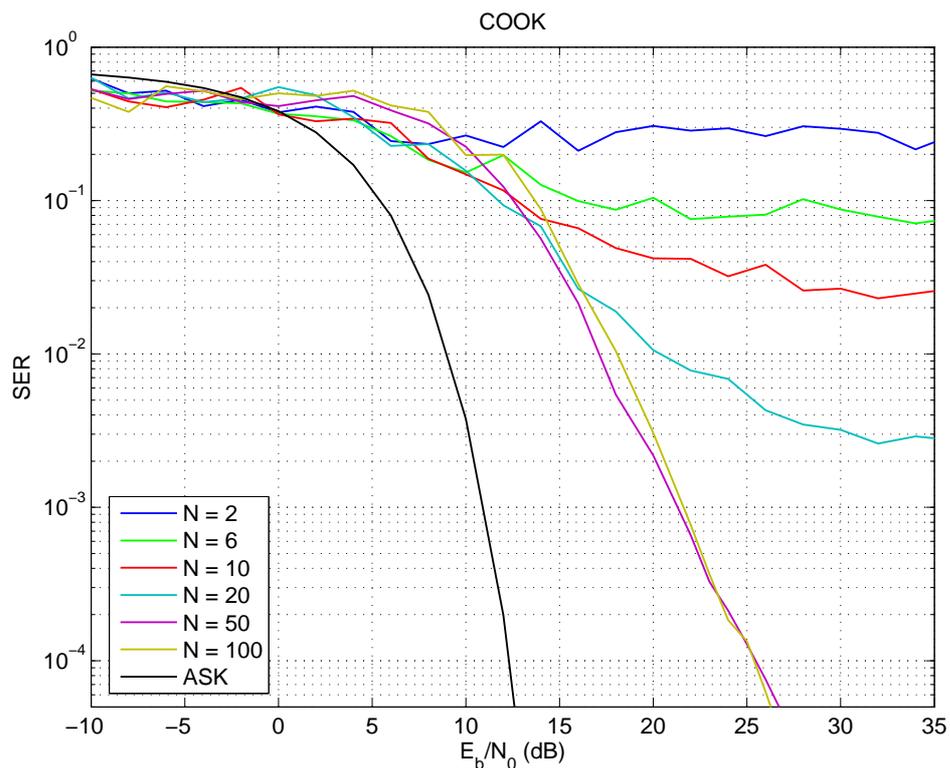


Figura 4.26: Taxas de erro de símbolo do sistema COOK.

Na Figura 4.26 também é mostrado, para efeito de comparação, o desempenho que seria obtido caso a seqüência de base caótica fosse substituída por um sinal com energia constante, representando um limite de desempenho. O sistema obtido dessa forma é o equivalente passa-baixas de tempo discreto do ASK não-coerente, cuja SER é dada por (LATHI, 1998)

$$\text{SER}_{\text{ASK}} = \frac{1}{2} \left[e^{-\frac{1}{2}E_b/N_0} + \frac{1}{2} \text{erfc} \left(\sqrt{\frac{1}{2} \frac{E_b}{N_0}} \right) \right]. \quad (4.48)$$

Por não apresentar o problema de imprecisão na estimativa da energia, o ASK tem um desempenho superior, e independente do número de pontos utilizados N .

É importante ressaltar que os resultados de SER foram obtidos com o nível de decisão sendo variado de acordo com o valor de E_b/N_0 . O controle do nível de decisão requer contínua determinação da potência de ruído no receptor o que aumenta a complexidade do circuito demodulador.

Supondo símbolos equi-prováveis, para garantir mínima probabilidade de erro, o nível de decisão deve ser alocado no ponto de cruzamento das funções densidades de probabilidade da variável observada condicionadas a transmissão de cada um dos símbolos (LATHI, 1998).

Numa primeira aproximação, o limiar z_T foi estimado como a média entre os valores esperados para a variável de observação z_{m1} quando são transmitidos $x_1(n)$ e $x_2(n)$. O cálculo exato desse limiar é deixado como trabalho futuro. O valor de z_T , apesar de poder melhorar ou piorar as taxas de erro obtidas não muda de forma qualitativa o comportamento das curvas obtidas, ou seja, a SER não tende a zero com o aumento de E_b/N_0 independentemente do limiar utilizado. Esse fato é observado, por exemplo, no histograma do CSK unipolar da Figura 4.24 em que as distribuições se sobrepõe mesmo na ausência de ruído.

Essa dependência do nível de decisão com a potência do ruído no canal certamente é um dos pontos fracos do CSK. Como ilustração, os histogramas da variável de observação para os casos $E_b/N_0 = 25\text{dB}$ e $E_b/N_0 = 40\text{dB}$ e $E_b = 1$ é mostrado na Figura 4.27 juntamente com o limiar utilizado em cada caso. Na mesma figura, é apresentado o nível de decisão adotado em função da E_b/N_0 no canal quando $E_b = 1$ e $N = 100$.

A melhoria no desempenho do sistema COOK (Figura 4.26) em relação ao CSK (Figura

4.25) decorre do maior distanciamento entre as energias médias dos sinais utilizados. A evolução da SER em função da constante k para uma E_b/N_0 fixa em 20dB é mostrada na Figura 4.28.

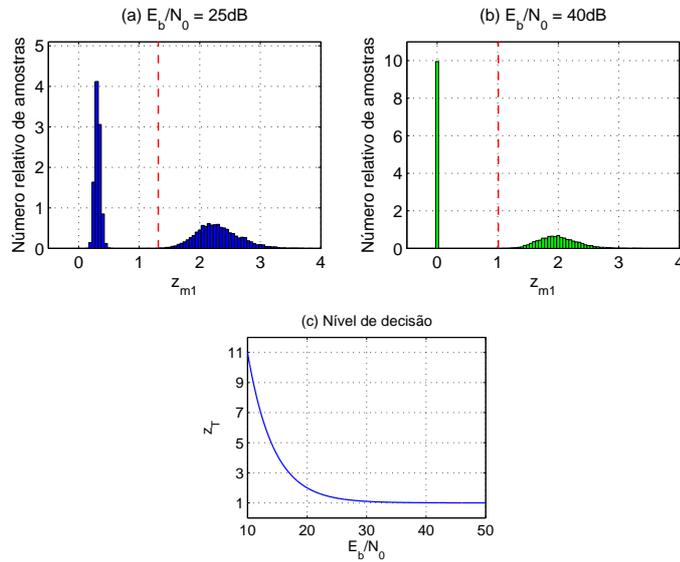


Figura 4.27: (a) Histograma de z_{m1} para o COOK com $E_b/N_0 = 25\text{dB}$; (b) idem para $E_b/N_0 = 40\text{dB}$; (c) nível de decisão em função da SNR no canal para $E_b = 1$ e $N = 100$.

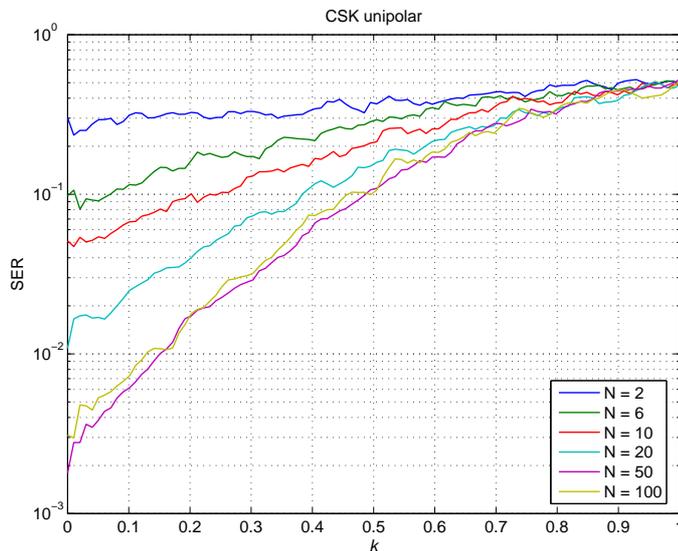


Figura 4.28: Taxas de erro de símbolo para o sistema CSK unipolar em função da constante k . A E_b/N_0 no canal foi mantida em 20dB.

Os resultados das simulações permitem concluir que os sistemas baseados em chaveamento caótico não-coerente apresentam desempenho muito aquém do necessário para ter alguma chance de concorrer com sistemas de comunicações convencionais. A dificuldade na determinação do limiar de detecção e o problema da incerteza da estimação da energia por símbolo

são os principais motivos para tanto.

4.6.2 Curvas de SER para o DCSK e FM-DCSK com recepção diferencial

No caso do sistema DCSK, não há necessidade de controle adaptativo do limiar o que torna o circuito de decisão bem mais simples. As curvas de SER em função da E_b/N_0 obtidas por meio de simulação desse sistema estão mostradas na Figura 4.29. Note que o problema de imprecisão da estimação ainda está presente. Porém, pelo gráfico obtido, para os valores de E_b/N_0 utilizados, as curvas não se estabilizam e decrescem monotonicamente. A variância de z_{m1} é função de N devido ao produto de dois sinais ruidosos que ocorre no receptor DCSK. Assim, o desempenho em canal AWGN do DCSK também é função de N .

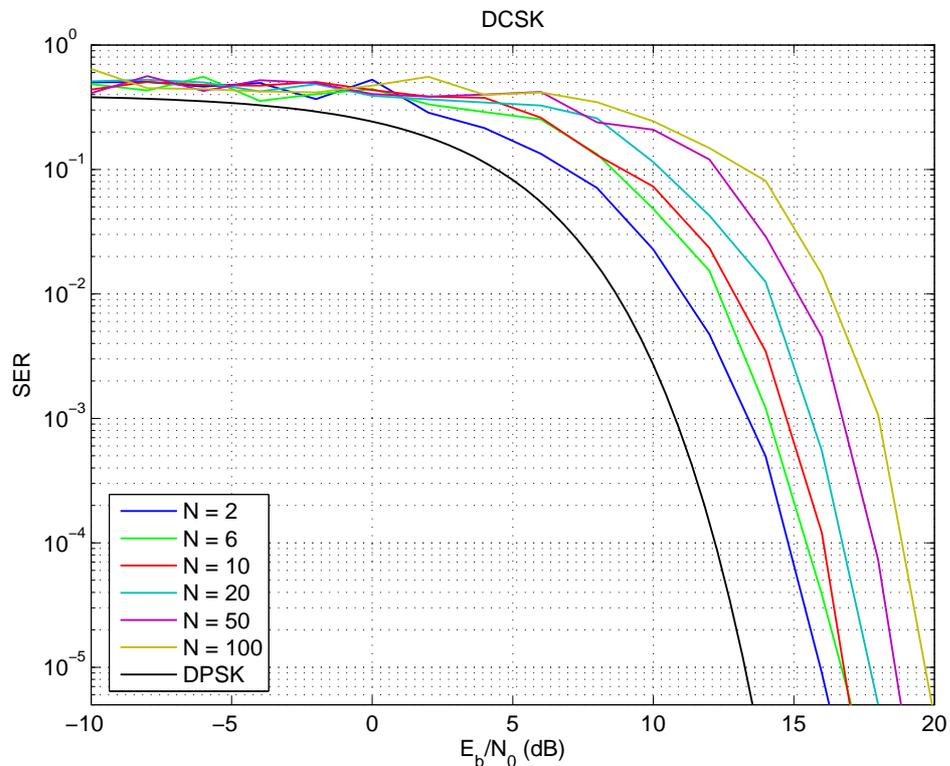


Figura 4.29: Taxas de erro de símbolo do sistema DCSK.

Caso os sinais caóticos fossem substituídos por sinais com energia constante, o sistema equivalente seria um DPSK binário não-coerente com energia por símbolo $E_b/2$ (LATHI, 1998).

Para este sistema, um resultado conhecido é que, utilizando-se um receptor ótimo, tem-se

$$\text{SER}_{\text{DPSK}} = \frac{1}{2} e^{-\frac{1}{2} \frac{E_b}{N_0}} \quad (4.49)$$

A curva de desempenho do receptor ótimo desse sistema, que pode ser considerado o limite de desempenho do DCSK, é mostrada na Figura 4.29. Para atingir $\text{SER} = 10^{-3}$, o DCSK precisa de uma E_b/N_0 cerca de 4dB maior no caso $N = 10$. Os motivos para o melhor desempenho do DPSK convencional são dois:

1. a energia por símbolo é constante, o que exclui a possibilidade da energia utilizada em um dado símbolo ser menor aproximando a variável de observação do nível de decisão;
2. mesmo no caso não-coerente, conhecem-se, a menos da fase, as funções de base no receptor. Isto permite a utilização de filtros casados para melhorar o desempenho do sistema (LATHI, 1998).

A Figura 4.30 mostra curvas de SER para o FM-DCSK. Neste caso, não ocorre mais o problema de imprecisão de estimação já que a energia por símbolo é constante. Porém, devido à correlação entre sinais ruidosos no receptor, permanece a dependência do desempenho com N . Ou seja, o FM-DCSK continua tendo desempenho em canal AWGN pior do que o DPSK devido à vantagem (2) desse último apresentada acima.

O FM-DCSK com $N = 2$ iguala-se em desempenho ao DPSK porque o trecho de informação e o trecho de referência é constituído por um único ponto, tornando a vantagem (2) do DPSK desnecessária.

Finalmente, a Figura 4.31 mostra num mesmo gráfico o desempenho de todos os sistemas analisados para o caso $N = 10$. Como esperado, percebe-se claramente que o FM-DCSK é o que possui melhor desempenho entre eles e isto se deve basicamente a energia por símbolo ser mantida constante neste sistema. Mesmo assim, o desempenho ainda fica abaixo do seu equivalente utilizando portadoras senoidais, o DPSK, pelo não conhecimento da função de base no receptor. Em contrapartida, o FM-DCSK goza das vantagens decorrentes dos sistemas de modulação caótica discutidos no Capítulo 1.

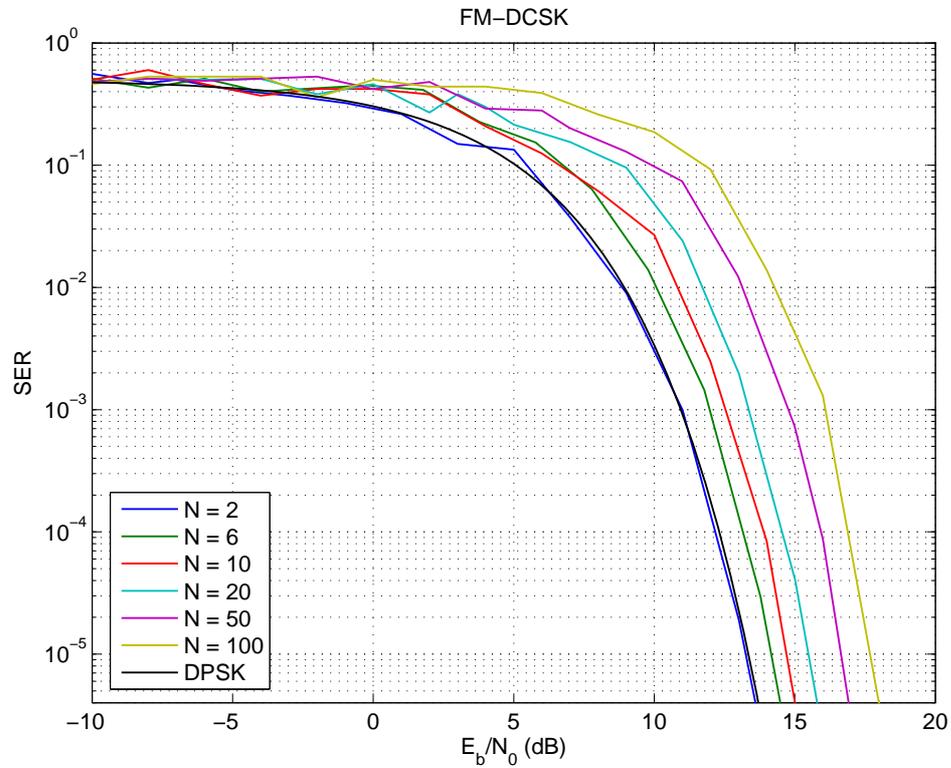


Figura 4.30: Taxas de erro de símbolo do sistema FM-DCSK.

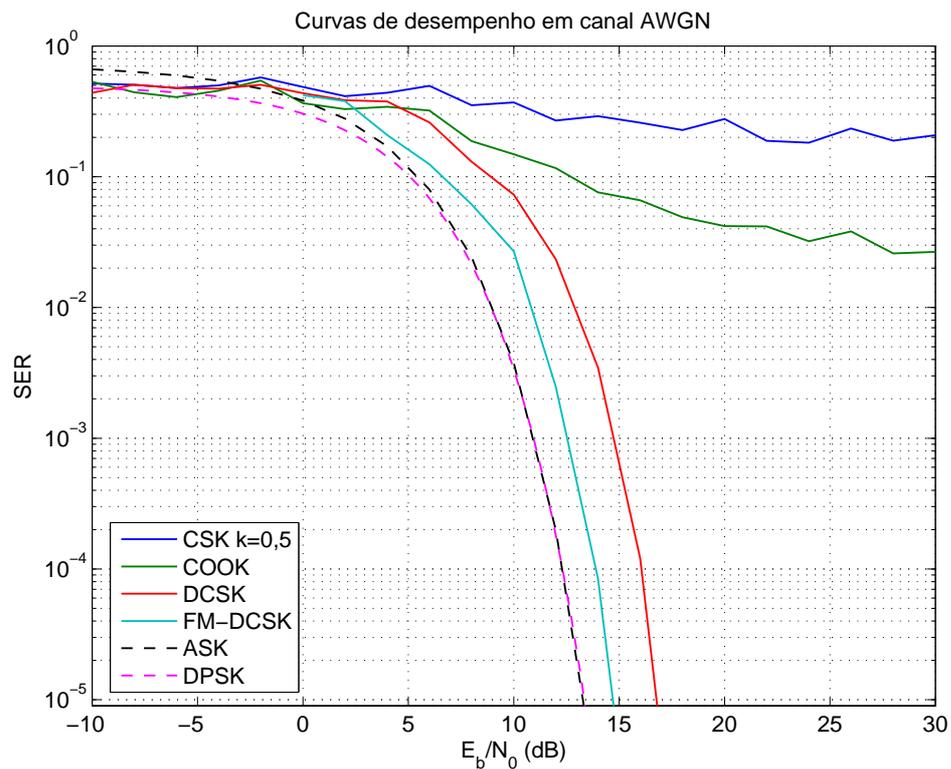


Figura 4.31: Taxas de erro de símbolo em canal AWGN de sistemas de comunicação digital usando sinais caóticos para $N = 10$.

Apesar do FM-DCSK ter as melhores características entre os sistemas caóticos analisados, é importante ressaltar que informações da dinâmica do mapa caótico não são utilizadas no processo de demodulação. Como o conhecimento do mapa utilizado na transmissão pode ser utilizado pelo receptor para melhorar os resultados alcançados será discutido nos próximos dois capítulos.

4.7 Conclusões

Neste capítulo foram apresentados alguns dos sistemas de modulação utilizando sinais caóticos mais estudados na literatura. O desempenho desses em canal AWGN foi analisado qualitativa e quantitativamente.

Como contribuição dessa tese, utilizou-se notação de tempo discreto, de forma consistente com os mapas utilizados na geração de sinais caóticos e, ao mesmo tempo, simplificando as simulações computacionais.

A Tabela 4.1 resume os problemas encontrados nas principais modulações digitais estudadas no capítulo. A coluna **Limiar** diz respeito ao problema da dependência do limiar com a potência do ruído no canal. A coluna **Energia** representa o problema da variabilidade da energia por símbolo. A coluna **Sincronização** significa a necessidade de recuperação das funções de base caóticas no receptor e a última coluna, **Não uso da dinâmica**, o não uso das propriedades do atrator caótico na estimação do símbolo transmitido.

Tabela 4.1: Problemas das modulações caóticas estudadas no capítulo.

Sistema	Limiar	Energia	Sincronização	Não uso da dinâmica
CSK coerente		X	X	
CSK não-coerente	X	X		X
DCSK		X		X
FM-DCSK				X

Entre as modulações estudadas, o FM-DCSK é a modulação que apresenta melhores resultados por não depender da sincronização caótica, ter nível de decisão independente do ruído e energia média por símbolo constante.

Os receptores não-coerentes e diferenciais analisados apresentam uma característica comum: não utilizam nenhuma característica da dinâmica do sistema gerador de sinais caóticos para fazer a demodulação. Essas técnicas limitam-se a estimar características do sinal recebido e compará-las com um nível de decisão, que pode ser adaptativo ou não.

O conhecimento *a priori* dos mapas geradores no receptor poderia ser utilizado de duas formas:

1. por meio de demodulações coerentes utilizando sincronização caótica.
2. melhorando a relação sinal-ruído ou diferenciando-os pelo desenvolvimento de técnicas para estimar os sinais caóticos que chegam ao receptor.

A presença de ruído e distorções no canal trazem resultados pouco satisfatórios quando se usa a sincronização caótica devido à dependência sensível às condições iniciais (EISENCRAFT; GERKEN, 2000; KENNEDY; ROVATTI; SETTI, 2000; LAU; TSE, 2003). Resta assim a segunda alternativa. No próximo capítulo serão exploradas algumas das técnicas de estimação de órbitas e condições iniciais baseadas em maximização de funções de verossimilhança (ABARBANEL, 1996; PAPADOPOULOS; WORNELL, 1993; KAY, 1995). Também são abordados seus limites de desempenho. No Capítulo 6 estas técnicas serão aplicadas a sistemas de modulação utilizando caos.

Capítulo 5

Estimação de sinais caóticos e de suas condições iniciais

Neste capítulo abordam-se técnicas de estimação de sinais caóticos e de suas condições iniciais na presença de ruído branco gaussiano aditivo (AWGN), bem como limites de desempenho dessas técnicas. O desempenho desses estimadores é relacionado a propriedades estatísticas dos sistemas dinâmicos que geram órbitas caóticas, como o número de Lyapunov e a densidade invariante.

Os sistemas de comunicações com recepção não-coerente ou diferencial usando sinais caóticos discutidos no Capítulo 4 fazem pouco uso da dinâmica implícita nos sinais caóticos. Pode-se imaginar que o desempenho daqueles sistemas seria basicamente o mesmo caso fosse usado um sinal aleatório ao invés de sinais caóticos.

As informações *a priori* sobre o sistema dinâmico que gera o sinal caótico, se conhecidas no transmissor e no receptor, podem ser usadas para melhorar a relação sinal-ruído na entrada do receptor e assim diminuir a SER do sistema. Além disso, essas informações podem ser usadas para diferenciar seqüências geradas por mapas diferentes, por meio da estimação do sinal caótico ou de sua condição inicial.

Na Seção 5.1 encontra-se uma das contribuições originais desta tese. É a dedução do limite inferior de Cramér-Rao (CRLB - *Cramér-Rao Lower Bound*) da variância de um estimador não-enviesado da condição inicial de uma órbita caótica gerada por um mapa unidimensional

em função de propriedades desse mapa.

O estimador de máxima verossimilhança (MLE - *Maximum Likelihood Estimator*) para a condição inicial pode ser deduzido facilmente quando a densidade invariante das órbitas de um mapa é uniforme (PAPADOPOULOS; WORNELL, 1993). Esse é o caso, por exemplo, da família de mapas tenda inclinada $f_I(\cdot)$, como visto na Eq. (2.36) da página 62. Como o estimador MLE é assintoticamente eficiente, ou seja é assintoticamente não-enviesado e atinge o CRLB de forma assintótica (KAY, 1993), ele pode ser utilizado para testar os resultados da Seção 5.1, o que é feito na Seção 5.2.

Para mapas genéricos, não é simples obter o MLE. Na Seção 5.3 introduz-se a estimação usando o algoritmo de Viterbi que pode ser aplicado, em princípio, a qualquer mapa, desde que se conheça uma conjugação desse com um mapa com densidade invariante uniforme, obstante o inconveniente dos valores possíveis para a órbita serem em número finito.

O algoritmo de Viterbi baseia-se no proposto por Dedieu e Kisel (1999). Porém, na Seção 5.3 mostra-se numericamente que o algoritmo da forma como foi publicado aplica-se apenas a mapas cuja densidade invariante é uniforme, informação que é deixada de lado pelos autores. Na mesma seção, o algoritmo é generalizado de forma a aplicar-se a mapas que sejam conjugados a um mapa com densidade invariante uniforme, o que constitui uma contribuição original. Por fim, é comparado o desempenho na estimação de órbitas caóticas pelo MLE e pelo algoritmo de Viterbi.

O capítulo termina com um resumo dos resultados que é feito na Seção 5.5.

No próximo capítulo, as técnicas de estimação estudadas serão empregadas na implementação de sistemas de modulação usando sinais caóticos.

5.1 Limites de desempenho para estimadores de condições iniciais

O problema da estimação da condição inicial de uma órbita pode ser colocado como se segue.

Considere uma órbita $s(n, s_0)$ do sistema dinâmico unidimensional

$$s(n+1) = f(s(n)). \quad (5.1)$$

Esse sinal é somado a um processo ruído branco gaussiano $r(n)$ estacionário com variância σ_r^2 e média nula:

$$s'(n) = s(n, s_0) + r(n) \quad (5.2)$$

e estão disponíveis as amostras $s'(n)$ para $0 \leq n \leq N-1$. Para simplificar a notação, definem-se os vetores

$$\mathbf{s} = [s(0), s(1), \dots, s(N-1)]^T, \quad (5.3)$$

$$\mathbf{r} = [r(0), r(1), \dots, r(N-1)]^T \text{ e} \quad (5.4)$$

$$\mathbf{s}' = [s'(0), s'(1), \dots, s'(N-1)]^T. \quad (5.5)$$

Deseja-se determinar o limite ótimo de desempenho de um estimador não-enviesado da condição inicial $s(0) \equiv s_0$. O critério de otimização usado é o limite inferior de Cramér-Rao (CRLB - *Cramér-Rao Lower Bound*) que fornece o menor valor que o erro médio quadrático de um estimador não-enviesado pode assumir (KAY, 1993). Na estimação, admite-se o conhecimento do mapa $f(\cdot)$ e da seqüência corrompida observada \mathbf{s}' . Nas seções seguintes busca-se determinar o CRLB na estimação da condição inicial de uma órbita em função de propriedades dinâmicas do mapa que a gerou, em especial, de seu número de Lyapunov.

Inicialmente, é feita uma breve revisão sobre o CRLB de forma a tornar mais claros os resultados aqui expostos.

5.1.1 O limite inferior de Cramér-Rao (CRLB)

O problema abordado consiste em estimar um parâmetro, no caso a condição inicial s_0 , a partir de um vetor aleatório de dados N -dimensional \mathbf{s}' com distribuição gaussiana. Esses dados podem ser representados pela função densidade de probabilidade $p(\mathbf{s}'; s_0)$. A densidade é parametrizada pelo parâmetro desconhecido s_0 . Assim, fica definida uma classe de funções,

sendo que o que diferencia uma das demais é o valor de s_0 . Quando a densidade é considerada como função do parâmetro desconhecido, com \mathbf{s}' fixo, é chamada *função de verossimilhança*.

Por exemplo, se $N = 1$, s_0 é a média da variável $s'(0)$ com distribuição gaussiana. Desta forma a função densidade de probabilidade desse dado é

$$p(s'(0); s_0) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp \left[-\frac{1}{2\sigma_r^2} (s'(0) - s_0)^2 \right] \quad (5.6)$$

que é exemplificada na Figura 5.1 para diferentes valores de s_0 e $\sigma_r = 0,01$. Note que a função densidade de $s'(0)$ é fortemente afetada por s_0 . Sendo assim, é possível estimar s_0 a partir de $s'(0)$.

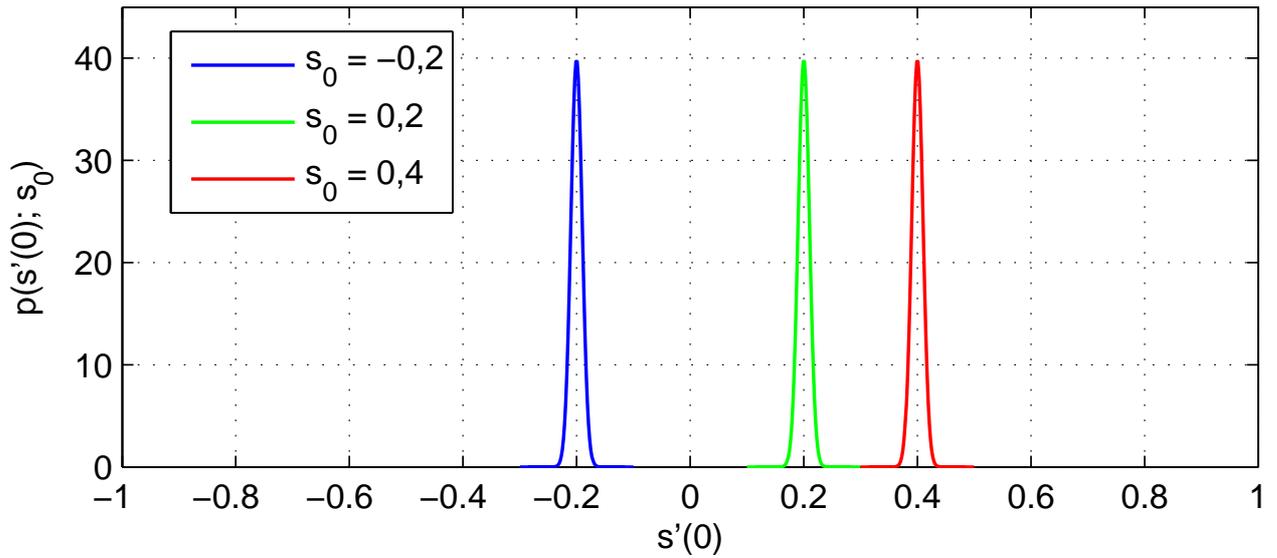


Figura 5.1: Gráficos de $p(s'(0); s_0)$ para alguns valores de s_0 .

Um estimador $\hat{s}_0 = h(\mathbf{s}')$ é uma função que atribui uma estimativa de s_0 para cada vetor de dados disponíveis \mathbf{s}' .

Seja $\hat{\theta}$ um estimador para a variável aleatória θ . Quando se deseja obter estimadores ótimos é necessário escolher um critério de otimização. O critério comumente usado e que é considerado neste trabalho é a minimização do *erro quadrático médio*, definido por

$$\text{mse}(\hat{\theta}) = E \left[(\hat{\theta} - \theta)^2 \right]. \quad (5.7)$$

O $\text{mse}(\hat{\theta})$ representa a média quadrática do desvio do valor obtido pelo estimador em relação

ao valor verdadeiro. Considerando-se apenas estimadores *não-enviesados*, ou seja, aqueles para os quais vale $E[\hat{\theta}] = \theta$,

$$\text{mse}(\hat{\theta}) = E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right] = \text{var}(\hat{\theta}). \quad (5.8)$$

Assim, minimizar o mse equivale, neste caso, a minimizar a variância do estimador $\hat{\theta}$.

Nem sempre é possível encontrar um estimador que tenha variância mínima para todo valor de θ . Quando isto ocorre, este estimador é chamado de estimador *não-enviesado de mínima variância* (MVU - *Minimum Variance Unbiased Estimator*) (KAY, 1993).

Seja $\hat{\theta}$ um estimador não-enviesado de uma variável aleatória θ . O CRLB determina a menor variância e, portanto, o menor mse que $\hat{\theta}$ pode assumir. Se existir um estimador cuja variância iguale o valor do CRLB para qualquer valor de θ , então este estimador é o MVU e é chamado de *eficiente*.

O Teorema 5 a seguir define o CRLB (KAY, 1993).

Teorema 5. *Considere um vetor aleatório \mathbf{x} dependente de um parâmetro escalar θ e uma função densidade de probabilidade $p(\mathbf{x}; \theta)$ que satisfaz a condição de “regularidade”*

$$E\left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta}\right] = 0, \quad \theta \in \Theta, \quad (5.9)$$

sendo Θ o conjunto de valores admissíveis para θ ¹. Neste caso, o mse de qualquer estimador não-enviesado $\hat{\theta}$ satisfaz

$$\text{mse}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2}\right]} \quad (5.10)$$

em que a derivada é tomada no valor verdadeiro de θ e a esperança é tomada com relação a $p(\mathbf{x}; \theta)$, com θ constante. O valor no segundo membro da desigualdade anterior é chamado de limite inferior de Cramér-Rao (CRLB). Além disso, pode ser encontrado um estimador não-enviesado que atinja este limite para todo θ se e somente se existirem funções $h(\mathbf{x})$ e $I(\theta)$ tais

¹Essa condição é geralmente verdadeira exceto quando o conjunto de valores \mathbf{x} para o qual a função densidade é não nula depende do parâmetro sendo estimado, como mostrado, por exemplo, em (KAY, 1993). Como isso não ocorre nos problemas tratados nessa tese, esta condição é considerada sempre satisfeita.

que

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = I(\theta)(h(\mathbf{x}) - \theta). \quad (5.11)$$

Esse estimador, que é o MVU, é $\hat{\theta} = h(\mathbf{x})$ e o mse é $1/I(\theta)$.

Como exemplo ilustrativo desse resultado, considera-se a determinação de um valor constante C a partir de N observações corrompidas pelo vetor \mathbf{r} , sendo AWGN de média nula e variância σ_r^2 ,

$$\mathbf{x} = [C, C, \dots, C]_{1 \times N}^T + \mathbf{r}. \quad (5.12)$$

Deseja-se obter o CRLB na estimação de C . Neste caso,

$$p(\mathbf{x}; C) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp \left[-\frac{1}{2\sigma_r^2} (x(n) - C)^2 \right]. \quad (5.13)$$

Calculando o logaritmo e tomando a derivada, obtém-se

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; C)}{\partial C} &= \frac{\partial}{\partial C} \left\{ -\ln \left[(2\pi\sigma_r^2)^{\frac{N}{2}} \right] - \frac{1}{2\sigma_r^2} \sum_{n=0}^{N-1} (x(n) - C)^2 \right\} \\ &= \frac{1}{\sigma_r^2} \sum_{n=0}^{N-1} (x(n) - C) \\ &= \frac{N}{\sigma_r^2} (\bar{x} - C) \end{aligned} \quad (5.14)$$

em que \bar{x} representa a média temporal dos N componentes do vetor \mathbf{x} . Derivando (5.14), chega-se a:

$$\frac{\partial^2 \ln p(\mathbf{x}; C)}{\partial C^2} = -\frac{N}{\sigma_r^2}. \quad (5.15)$$

Neste caso a derivada obtida em (5.15) não depende de C . Substituindo-se esse resultado em (5.10), conclui-se que o CRLB na estimação de C é

$$\text{mse}(\hat{C}) \geq \frac{\sigma_r^2}{N}. \quad (5.16)$$

Além disso, comparando-se (5.14) com (5.11) observa-se que, neste caso, $I(\theta) = N/\sigma_r^2$ e $h(\mathbf{x}) = \bar{x}$. Assim, a média \bar{x} atinge o CRLB e é portanto um estimador MVU.

5.1.2 O CRLB na estimação da condição inicial de mapas caóticos

O Teorema 5 pode ser aplicado diretamente ao problema de se estimar a condição inicial da órbita corrompida por AWGN da Eq. (5.2). O resultado é expresso pelo Teorema 6.

Teorema 6. *Considere uma órbita $s(n, s_0)$ do sistema dinâmico*

$$s(n+1, s_0) = f(s(n)) \quad (5.17)$$

sendo que o mapa $f(\cdot)$ possui derivada em todos os pontos dessa órbita.

Seja $r(n)$ um processo ruído branco gaussiano de média nula e variância σ_r^2 e o sinal corrompido por ruído observado $s'(n) = s(n) + r(n)$, $0 \leq n < N$. Então, o CRLB na estimação de s_0 dado $s'(n)$ e o mapa $f(\cdot)$ é

$$\text{mse}(\hat{s}_0) \geq \frac{\sigma_r^2}{1 + \sum_{n=1}^{N-1} \left(\prod_{j=0}^{n-1} \frac{df}{ds} \Big|_{s(j, s_0)} \right)^2}. \quad (5.18)$$

Demonstração. Utilizando a notação definida nas Eqs. (5.3 - 5.5), o vetor \mathbf{s}' é definido por

$$\mathbf{s}' = \mathbf{s}(s_0) + \mathbf{r}, \quad (5.19)$$

em que a dependência com s_0 é mostrada de forma explícita. Como \mathbf{r} é formado por variáveis aleatórias gaussianas independentes, a função densidade de probabilidades de \mathbf{s}' é

$$p(\mathbf{s}'; s_0) = \frac{1}{(2\pi\sigma_r^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma_r^2} \sum_{n=0}^{N-1} (s'(n) - s(n, s_0))^2 \right]. \quad (5.20)$$

Tomando o logaritmo de (5.20) e diferenciando com relação a s_0 ,

$$\frac{\partial \ln p(\mathbf{s}'; s_0)}{\partial s_0} = \frac{1}{\sigma_r^2} \sum_{n=0}^{N-1} (s'(n) - s(n, s_0)) \frac{\partial s(n, s_0)}{\partial s_0}. \quad (5.21)$$

Diferenciando novamente, obtém-se

$$\frac{\partial^2 \ln p(\mathbf{s}'; s_0)}{\partial s_0^2} = \frac{1}{\sigma_r^2} \sum_{n=0}^{N-1} \left[(s'(n) - s(n, s_0)) \frac{\partial^2 s(n, s_0)}{\partial s_0^2} - \left(\frac{\partial s(n, s_0)}{\partial s_0} \right)^2 \right]. \quad (5.22)$$

Considerando-se o valor esperado de (5.22):

$$E \left[\frac{\partial^2 \ln p(\mathbf{s}'; s_0)}{\partial s_0^2} \right] = -\frac{1}{\sigma_r^2} \sum_{n=0}^{N-1} \left(\frac{\partial s(n, s_0)}{\partial s_0} \right)^2. \quad (5.23)$$

Para se chegar ao resultado do teorema é necessário exprimir a derivada acima em termos de derivadas do mapa $f(\cdot)$. Usando a regra da cadeia e (5.17), conclui-se que

$$\frac{\partial s(n, s_0)}{\partial s_0} = \frac{df^n}{ds} \Big|_{s_0} = \begin{cases} \frac{df}{ds} \Big|_{s(0)} \cdot \frac{df}{ds} \Big|_{s(1)} \cdot \dots \cdot \frac{df}{ds} \Big|_{s(n-1)}, & n \geq 1 \\ 1, & n = 0. \end{cases}, \quad (5.24)$$

em que $f^n(\cdot)$ é a n -ésima composta de $f(\cdot)$. Em seguida, substituindo (5.24) em (5.23) obtém-se

$$E \left[\frac{\partial^2 \ln p(\mathbf{s}'; s_0)}{\partial s_0^2} \right] = -\frac{1}{\sigma_r^2} \left(1 + \sum_{n=1}^{N-1} \left(\prod_{j=0}^{n-1} \frac{df}{ds} \Big|_{s(j, s_0)} \right)^2 \right). \quad (5.25)$$

Finalmente, utilizando-se (5.10) chega-se à (5.18) e o teorema está provado. \square

Observe que a desigualdade (5.18) é válida para órbitas caóticas ou não, como resulta do enunciado e da demonstração. Esse resultado mostra que, em geral, o CRLB depende de s_0 , que é a condição inicial sendo estimada. No caso de órbitas caóticas, porém, esta dependência torna-se menos importante conforme N cresce já que a órbita $s(n, s_0)$ tende a percorrer todo o atrator caótico (DEVANEY, 2003). Essa característica se manifesta claramente no resultado do Teorema 7 a seguir.

Teorema 7. *São dados os mesmos vetores e condições do Teorema 6. Considerando-se uma órbita $s(n, s_0)$, o limite do CRLB quando $N \rightarrow \infty$ é*

$$\text{mse}(\hat{s}_0) \geq \sigma_r^2 \frac{L^2 - 1}{L^{2N} - 1}, \quad (5.26)$$

em que $L \equiv L(s_0) \neq 1$ é o número de Lyapunov do atrator para o qual a órbita $s(n, s_0)$ converge².

Demonstração. Para n suficientemente grande, pode-se usar a definição de número de Lyapunov, Eq. (2.6) página 48, para calcular a seguinte aproximação:

$$\left(\frac{\partial s(n, s_0)}{\partial s_0}\right)^2 = \left(\frac{df}{ds}\Big|_{s(0)} \cdot \frac{df}{ds}\Big|_{s(1)} \cdots \frac{df}{ds}\Big|_{s(n-1)}\right)^2 \approx L^{2n}(s_0). \quad (5.27)$$

Desta forma,

$$\sum_{n=0}^{N-1} \left(\frac{\partial s(n, s_0)}{\partial s_0}\right)^2 \approx \sum_{n=0}^{N-1} L^{2n}(s_0) = \frac{L^{2N}(s_0) - 1}{L^2(s_0) - 1} \quad (5.28)$$

e usando a Eq. (5.23),

$$E \left[\frac{\partial^2 \ln p(\mathbf{s}'; s_0)}{\partial s_0^2} \right] \approx -\frac{1}{\sigma_r^2} \frac{L^{2N}(s_0) - 1}{L^2(s_0) - 1}, \quad (5.29)$$

para N suficientemente grande.

Assim, considerando-se que a órbita converge para um atrator com número de Lyapunov L e substituindo-se (5.29) na desigualdade (5.10) chega-se a

$$\text{mse}(\hat{s}_0) \geq \sigma_r^2 \frac{L^2 - 1}{L^{2N} - 1}, \quad (5.30)$$

como se desejava demonstrar. □

Cabe ressaltar que a desigualdade (5.26) pode ser utilizada como uma aproximação para o CRLB mesmo para valores baixos de N no caso de órbitas caóticas. Isso se justifica pelo fato de que o número de Lyapunov, sendo uma média geométrica das derivadas do mapa nos pontos de uma órbita é usado para substituir o produto dessas derivadas. Essa simplificação apresenta bons resultados como é mostrado nos exemplos da próxima seção.

Assim como no caso do Teorema 6, esse resultado se aplica a órbitas caóticas ($L > 1$) ou não ($L < 1$). Neste último caso, a soma em (5.28) tende a uma constante a medida que N cresce e, conseqüentemente, o mesmo ocorre com o CRLB da desigualdade (5.26). Esse resultado reflete o fato de que se a órbita converge para um atrator periódico então, a partir de um certo

²Esse resultado é válido mesmo quando o mapa tem mais de um atrator: L é o número de Lyapunov do atrator para o qual a órbita em questão converge.

valor de N pouca informação extra sobre a condição inicial é obtida aumentando-se o número de pontos usados na estimação.

Como observação, no caso $L = 1$, a fórmula da soma de progressão geométrica utilizada em (5.28) não é válida. A somatória nessa equação é igual a N e procedendo-se de forma análoga à demonstração, conclui-se que

$$\text{mse}(\hat{s}_0) \geq \frac{\sigma_r^2}{N}. \quad (5.31)$$

Ou seja, neste caso, que representa órbitas que não se atraem nem se repelem, o CRLB é idêntico ao da estimação de uma constante em AWGN, como se nota por comparação com (5.16).

No caso de órbitas caóticas, a informação sobre a condição inicial sempre aumenta com N : o Teorema 7 mostra que o CRLB decai exponencialmente com N neste caso. A cada nova amostra adicionada à estimação, maior conhecimento obtém-se da condição inicial.

Como mostra a expressão (5.16), a estimação de uma constante decai apenas com $1/N$. Já o decaimento exponencial do CRLB com N no caso caótico da desigualdade (5.26) é muito mais rápido. Teoricamente, portanto, é possível construir estimadores de condições iniciais bastante precisos para mapas caóticos. No entanto, devido à dependência sensível com as condições iniciais, obter s_0 com alta precisão não significa uma reconstrução precisa da órbita $s(n, s_0)$. Se o erro na estimação da condição inicial for ϵ , o erro no cálculo de $s(N-1, s_0)$ por meio de iterações de $f(\cdot)$ a partir de \hat{s}_0 será $O(\max(s(N-1), \epsilon L^{N-1}))$ (ALLIGOOD; SAUER; YORK, 1996).

Desta forma, mapas com número de Lyapunov mais elevado permitem, em princípio, maior precisão na estimação da condição inicial de suas órbitas. No entanto, esta vantagem é reduzida pela grande amplificação do erro quando se calcula pontos subseqüentes da órbita. Assim, se um dado sistema codifica e recupera informação apenas pela escolha de s_0 , é interessante escolher um mapa com alto número de Lyapunov, porém não se deve esperar bons resultados na recuperação da órbita $s(n, s_0)$ inteira.

Uma versão desses teoremas para mapas lineares por partes foi publicada em (PAPADOPOULOS; WORNELL, 1993). Esse problema também é tratado com abordagens diferentes em

(KAY, 1995; KAY; NAGESHA, 1995; ABARBANEL, 1996). Porém em nenhum desses trabalhos são obtidas fórmulas gerais para o CRLB em função do número de Lyapunov de um mapa unidimensional qualquer. Assim, os Teoremas 6 e 7 são resultados originais desta tese.

5.1.3 Exemplos

A seguir são estudados três exemplos de aplicação dos resultados da seção anterior. O primeiro refere-se a um mapa linear por partes com derivada em módulo constante. O segundo é um sistema definido por um mapa também linear por partes mas com derivada não-constante em módulo; por último, um sistema com mapa quadrático.

Primeiro exemplo: mapa tenda

Considere órbitas caóticas do mapa tenda $f_T(\cdot)$ definido na Seção 2.5.1 e repetido aqui por conveniência:

$$f_T(s) = 1 - 2|s|. \quad (5.32)$$

Como $f_T(\cdot)$ é linear por partes com derivada em módulo constante, para uma órbita caótica desse mapa a aproximação da (5.27) torna-se exata com $L(s_0) = L_T = 2$ e a desigualdade (5.26) é válida para qualquer N e não somente para $N \rightarrow \infty$. Sendo assim, o CRLB não depende de s_0 e

$$\text{mse}(\hat{s}_0) \geq \sigma_r^2 \frac{3}{4^N - 1} \quad (5.33)$$

é o CRLB na estimação de s_0 em AWGN para qualquer $N \geq 1$ e qualquer s_0 desde que $s(n, s_0) \neq 0$ para $0 \leq n \leq N - 1$.

No gráfico da Figura 5.2 encontra-se a variação do CRLB em função de N para $\sigma_r = 1$. Para comparação a mesma figura mostra a variação do CRLB na estimação de uma constante em AWGN dada pela desigualdade (5.16).

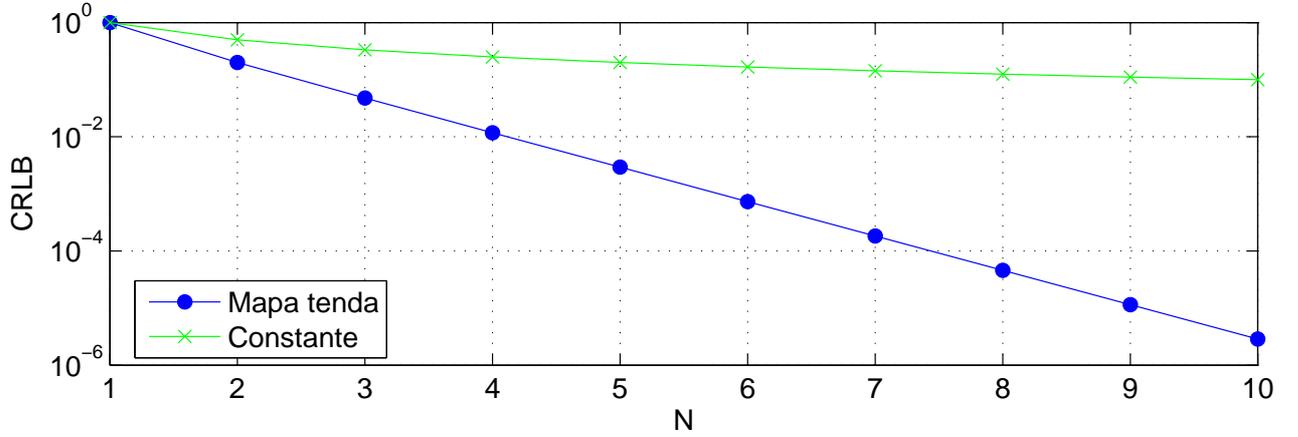


Figura 5.2: CRLB na estimação da condição inicial de uma órbita de $f_T(\cdot)$ em AWGN. Para comparação, também é mostrado o CRLB na estimação de um sinal constante em AWGN.

Segundo exemplo: mapas tenda inclinada

A família de mapas tenda inclinada $f_I(\cdot)$ definida na Seção 2.5.2 é dada por

$$s(n+1) = f_I(s(n)) = \begin{cases} \frac{2}{\alpha+1}s(n) + \frac{1-\alpha}{\alpha+1}, & -1 < s(n) < \alpha \\ \frac{2}{\alpha-1}s(n) - \frac{\alpha+1}{\alpha-1}, & \alpha \leq s(n) < 1 \end{cases} \quad (5.34)$$

O CRLB na estimação da condição inicial desse mapa é função da condição inicial s_0 , do número de pontos usados na estimação N e do parâmetro α que define seu número de Lyapunov L_I .

Na Figura 5.3 observa-se o CRLB obtido na estimação da condição inicial $s_0 = 0$ em função do parâmetro α para $\sigma_r = 1$ e diversos valores de N . Para cada valor de N são mostradas duas curvas: uma obtida através da aplicação do Teorema 6, desigualdade (5.18), e outra tracejada obtida através da aproximação do Teorema 7, desigualdade (5.26). Apesar da aparente descontinuidade, é importante ressaltar que ambas as curvas são determinísticas. Percebe-se claramente que o resultado do Teorema 7 é realmente uma boa aproximação do CRLB mesmo para valores pequenos de N . Observa-se que os menores valores de CRLB concentram-se na região próxima de $\alpha = 0$.

Considerando-se o mapa $f_T(\cdot)$ um particular mapa $f_I(\cdot)$ com $\alpha = 0$, este é o mapa cujas órbitas apresentam maior número de Lyapunov $L_{I_{max}} = 2$ e conseqüentemente o menor valor de CRLB pela inequação (5.26).

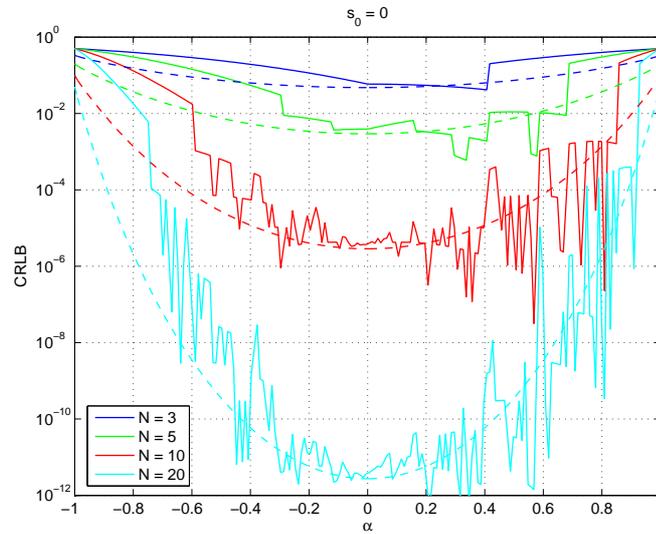


Figura 5.3: CRLB na estimação da condição inicial da órbita $s(n, 0)$ do mapa $f_I(\cdot)$ em função de α para $\sigma_r = 1$ e diversos valores de N . A aproximação da desigualdade (5.26) é mostrada em tracejado.

Note-se que as características acima discutidas não resultam do valor de s_0 escolhido. Na Figura 5.4 são mostradas curvas para $s_0 = 0,6$. Os resultados mostram o mesmo tipo de comportamento para todos os valores testados de $s_0 \in (-1, 1)$.

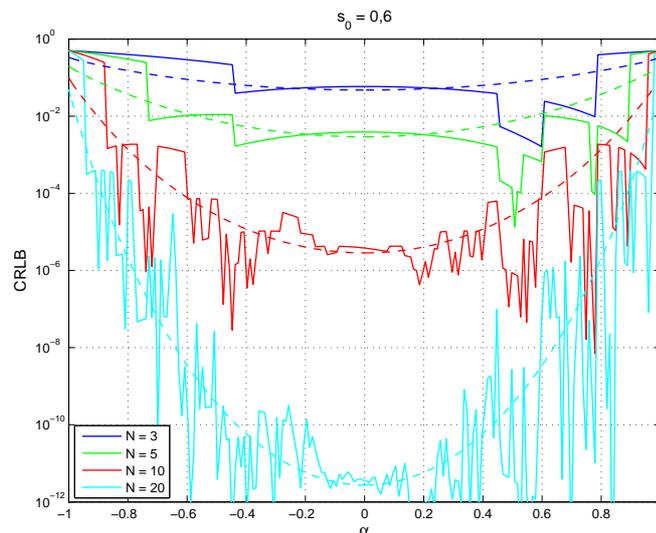


Figura 5.4: CRLB na estimação da condição inicial da órbita $s(n, 0,6)$ do mapa $f_I(\cdot)$ em função de α para diversos valores de N e $\sigma_r = 1$. A aproximação da desigualdade (5.26) é mostrada em tracejado.

Pode-se afirmar que entre os mapas $f_I(\cdot)$, o mapa $f_T(\cdot)$ é o que apresenta melhores resultados em termos de limite de desempenho da estimação da condição inicial de órbitas, que se acentuam a medida que mais pontos são usados na estimação de s_0 .

A Figura 5.5 ilustra a dependência do CRLB com a condição inicial para $\alpha \neq 0$. Em cada curva o CRLB foi calculado para 10000 condições iniciais espalhadas uniformemente no intervalo $[-1, 1]$. No gráfico (a), considera-se $\alpha = 0,8$ e no gráfico (b), toma-se $\alpha = -0,8$.

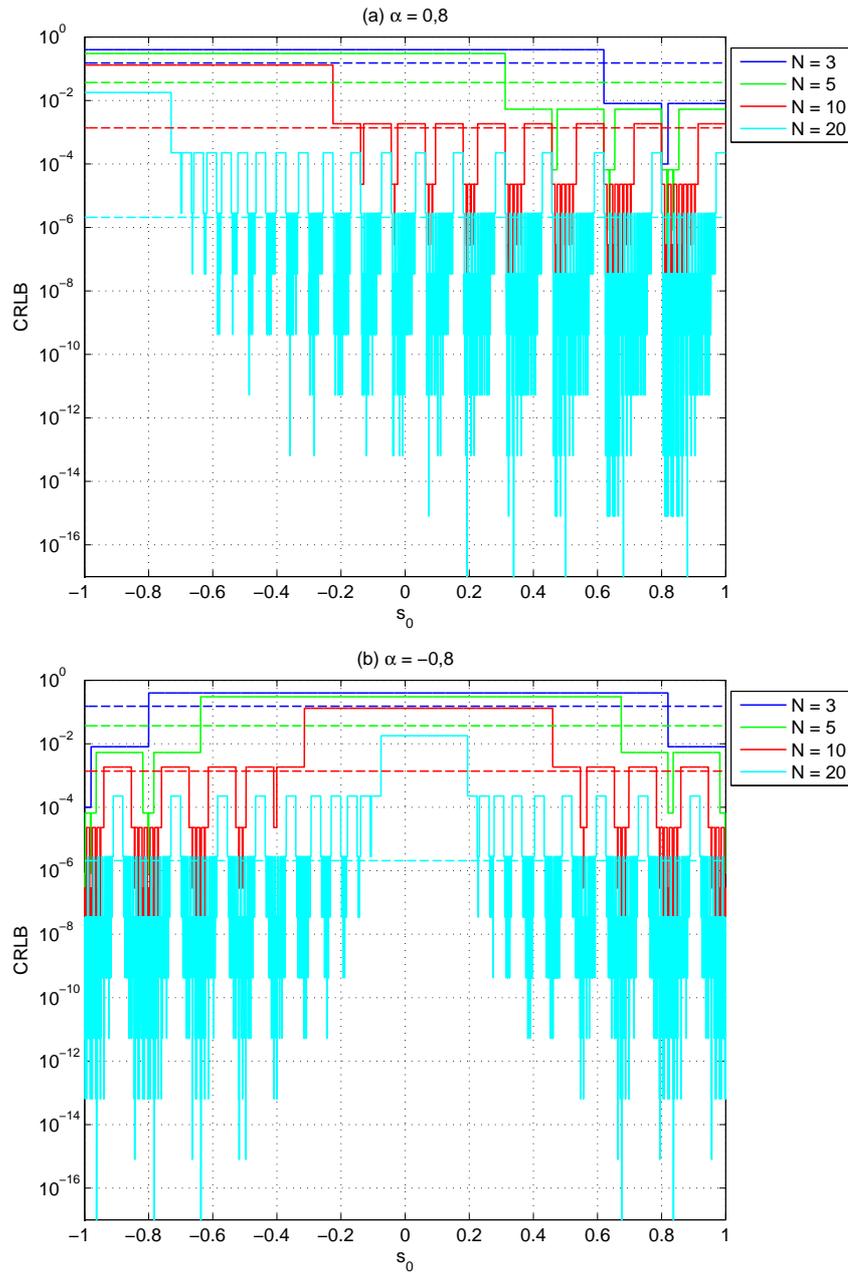


Figura 5.5: CRLB na estimação da condição inicial da órbita $s(n, s_0)$ do mapa $f_I(\cdot)$ com (a) $\alpha = 0,8$ e (b) $\alpha = -0,8$ em função de s_0 para diversos valores de N e $\sigma_r = 1$. A aproximação da desigualdade (5.26) é mostrada em tracejado.

O aspecto descontínuo das curvas é devida à descontinuidade da derivada dos mapas $f_I(\cdot)$. Cada trecho constante corresponde a condições iniciais cujos N primeiros pontos das órbitas percorrem subintervalos em que $f_I(\cdot)$ possui mesmas derivadas, resultando em igual

mse (Eq. (5.18)). Conforme N cresce, estes trechos constantes tornam-se menores e o gráfico assume um aspecto mais complicado.

Para cada valor de N é mostrado em tracejado o CRLB aproximado dado pela desigualdade (5.26). Percebe-se nesses gráficos que a aproximação do Teorema 7 é de fato uma aproximação do valor médio do CRLB em termos das condições iniciais.

Terceiro exemplo: mapa quadrático

Como último exemplo, considere-se o mapa quadrático $f_Q(\cdot)$ e o sistema dinâmico associado definido na Eq. (2.13) da página 54, repetida aqui por conveniência.

$$s(n+1) = f_Q(s(n)) = -2s(n)^2 + 1. \quad (5.35)$$

A Figura 5.6 mostra o CRLB em função da condição inicial s_0 para $N = 3$ amostras e $\sigma_r = 1$ calculado utilizando-se (5.18); note-se a dependência do CRLB com a condição inicial que está sendo estimada. O CRLB atinge valores altos próximo do ponto $s_0 = 0$ em que a derivada de $f_Q(\cdot)$ se anula e de $s_0 = \pm\sqrt{\frac{1}{2}}$ para os quais $f_Q(s_0) = 0$. Próximos deles, as órbitas levam mais iterações para se distanciar. É mais difícil estimar a condição inicial de uma órbita que se inicia próxima dos pontos acima referidos.

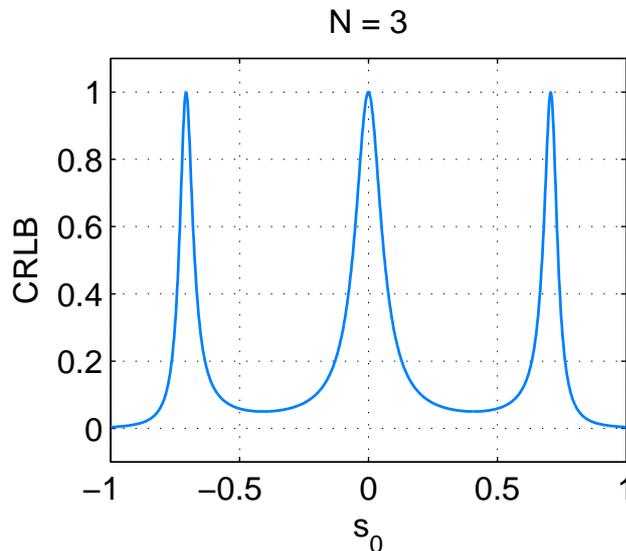


Figura 5.6: CRLB na estimação da condição inicial da órbita $s(n, s_0)$ do mapa $f_Q(\cdot)$ para $N = 3$ e $\sigma_r = 1$.

Na Figura 5.7 é mostrada a dependência do CRLB com s_0 para valores de N crescentes. Diferentemente do que ocorria com os mapas tenda inclinada (Figura 5.5) não aparecem trechos constantes já que a derivada de $f_Q(\cdot)$ varia ponto a ponto. Conforme aumenta o número de pontos usados na estimação, a dependência do CRLB com s_0 torna-se mais complicada com picos espalhados por todo o intervalo $[-1, 1]$. Porém, a relação entre estes picos e o CRLB médio tende a tornar-se cada vez menor, conforme discussão qualitativa feita no comentário seguinte ao Teorema 6.

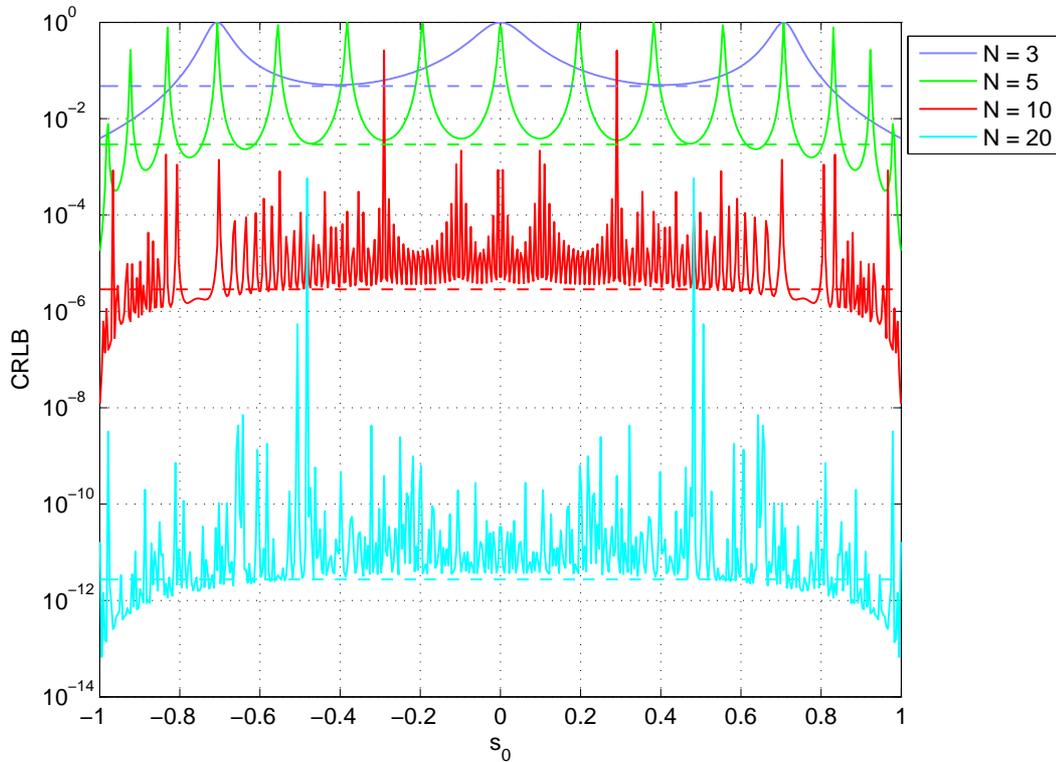


Figura 5.7: CRLB na estimação da condição inicial da órbita $s(n, s_0)$ do mapa $f_Q(\cdot)$ para diversos valores de N e $\sigma_r = 1$. A aproximação da desigualdade (5.26) é mostrada em tracejado.

Como no caso do mapa tenda, aplicando o Teorema 7 e utilizando o fato de que o número de Lyapunov das órbitas caóticas desse mapa é $L_Q = 2$, obtém-se

$$\text{mse}(\hat{s}_0) \geq \frac{3}{4^N - 1} \quad (5.36)$$

para N suficientemente grande. Esse resultado se mostra coerente com os gráficos da Figura 5.7. Por exemplo, para $N = 20$, (5.36) fornece aproximadamente $\text{mse}(\hat{s}_0) \geq 2,7 \cdot 10^{-12}$ o que está de acordo com a figura. Os valores obtidos através da desigualdade (5.36) são mostrados

em tracejado na figura.

5.1.4 Comentários

O Teorema 7 mostra o forte relacionamento entre as propriedades estatísticas do atrator de um mapa, aqui representadas pelo número de Lyapunov e o limite de desempenho de um estimador da condição inicial de uma órbita gerada por este mapa corrompida por AWGN, representado pelo CRLB.

Esse resultado, em princípio, provê um critério para se escolher o melhor mapa unidimensional para ser usado na geração de um sinal caótico do qual será necessário recuperar a condição inicial quando corrompido por AWGN. Assim, se for possível encontrar um estimador eficiente para este problema, quanto maior o número de Lyapunov de um mapa, menor o CRLB desse estimador e portanto melhor as estimativas.

Uma outra conclusão interessante que se pode tirar da desigualdade (5.26) é que mapas conjugados (veja Definição 15 da página 49) têm o mesmo CRLB na estimação de condições iniciais para N grande, já que as órbitas destes mapas possuem mesmo número de Lyapunov. Por exemplo, os mapas $f_T(\cdot)$ e $f_Q(\cdot)$ têm desempenho idêntico neste quesito já que $L_T = L_Q = 2$. Porém, quando N é pequeno, o desempenho depende da condição inicial sendo estimada e a análise é mais complicada.

Apesar dos resultados refletirem apenas limitações de desempenho e não o desempenho em si, eles mostram que a escolha adequada do mapa em conjunto com o uso das propriedades estatísticas deve trazer benefícios aos sistema de comunicação usando sinais caóticos discutidos no Capítulo 4.

Nas próximas seções são apresentadas duas formas de se fazer a estimação das condições iniciais e das órbitas de sistemas caóticos: o MLE para mapas com densidade invariante uniforme e o algoritmo de Viterbi.

5.2 O MLE

Nesta seção é apresentado um exemplo de estimação de máxima verossimilhança para a condição inicial e para uma órbita de um mapa tenda inclinada $f_I(\cdot)$ baseado no artigo (PADOPOULOS; WORNELL, 1993). Primeiramente, é feita uma revisão da definição e das propriedades do estimador de máxima verossimilhança (MLE). Em seguida, na Seção 5.2.2 mostram-se os resultados de sua aplicação na estimação da condição inicial de órbitas dos mapas $f_I(\cdot)$ e seu relacionamento com o CRLB. Na Seção 5.2.3, é abordada a estimação de órbitas inteiras.

5.2.1 O MLE - Definição e propriedades

Lembrando que a função de verossimilhança é a densidade $p(\mathbf{x}; \theta)$ para \mathbf{x} fixo, o MLE de um parâmetro escalar é definido como o valor de θ que maximiza $p(\mathbf{x}; \theta)$ (KAY, 1993). A maximização é realizada sobre todo o conjunto de valores possíveis para a variável θ . Como $p(\mathbf{x}; \theta)$ é também uma função de \mathbf{x} , a maximização produz um estimador $\hat{\theta}$ função de \mathbf{x} . A motivação para a definição do MLE decorre de que $p(\mathbf{x}; \theta)d\mathbf{x}$ dá a probabilidade de se observar \mathbf{x} em um pequeno volume $d\mathbf{x}$ para um dado θ .

Na Figura 5.8 é esboçada a forma de uma função de verossimilhança para $\mathbf{x} = \mathbf{x}_0$ fixo em função de θ . O valor de $p(\mathbf{x} = \mathbf{x}_0; \theta)d\mathbf{x}$ para cada θ descreve a probabilidade de se observar \mathbf{x} em uma região de \mathbb{R}^N centrada em \mathbf{x}_0 com volume $d\mathbf{x}$, para um dado valor de θ . Se $\mathbf{x} = \mathbf{x}_0$ foi de fato observado, não seria razoável inferir que $\theta = \theta_1$ porque se $\theta = \theta_1$, a probabilidade de observar $\mathbf{x} = \mathbf{x}_0$ seria muito pequena. É mais “provável” que $\theta = \theta_2$ seja o valor verdadeiro. Assim, escolhe-se $\hat{\theta} = \theta_2$ ou o valor que maximiza $p(\mathbf{x} = \mathbf{x}_0; \theta)$ sobre o intervalo de valores possíveis de θ , como a estimação.

Como exemplo, considera-se, como feito na seção anterior, o caso de se determinar um valor constante C a partir de N observações corrompidas por AWGN de média nula com variância σ_r^2 , representadas pelo vetor $\mathbf{x} = [C, C, \dots, C]_{1 \times N}^T + \mathbf{r}$. A função densidade de probabilidades neste caso é

$$p(\mathbf{x}; C) = \frac{1}{\left(\sqrt{2\pi\sigma_r^2}\right)^N} \exp \left[-\frac{1}{2\sigma_r^2} \sum_{n=0}^{N-1} (x(n) - C)^2 \right]. \quad (5.37)$$

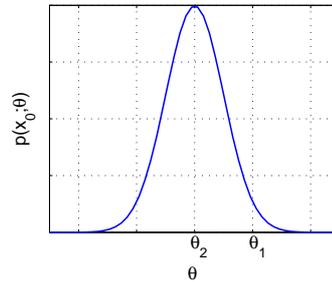


Figura 5.8: Exemplo de função de verossimilhança.

Tomando-se o logaritmo da expressão anterior e derivando-se com relação ao parâmetro C que está sendo estimado, obtém-se

$$\frac{\partial \ln p(\mathbf{x}; C)}{\partial C} = \frac{1}{\sigma_r^2} \sum_{n=0}^{N-1} (x(n) - C) \quad (5.38)$$

que igualando-se a zero leva ao MLE

$$\hat{C} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}. \quad (5.39)$$

Na Seção 5.1.1 mostrou-se que a média temporal \bar{x} é um estimador eficiente neste problema e assim, o MLE é eficiente. Se um estimador eficiente existe, então a estimativa por máxima verossimilhança o produz (KAY, 1993).

Uma propriedade importante do MLE é que ele é assintoticamente não-enviesado e eficiente. Ou seja, o erro quadrático médio se aproxima do limite de Cramér-Rao para um número suficientemente grande de pontos usados na estimação (KAY, 1993).

5.2.2 Estimação da condição inicial com o MLE

Para mapas cuja densidade invariante é uniforme, como no caso da família de mapas tenda inclinada, o MLE foi deduzido em (PAPADOPOULOS; WORNELL, 1993). Neste caso, ele pode ser implementado de forma simples. Quando o mapa tem densidade invariante mais complicada ou mesmo sem uma fórmula fechada conhecida, uma solução é substituir esta densidade por uma matriz de transição de estados e aplicar a técnica de estimação baseada no algoritmo de Viterbi que é descrita na Seção 5.3.

A implementação do MLE possui duas fases: uma filtragem e uma suavização. A filtragem consiste em gerar uma estimativa $\check{s}(n)$ para $s(n)$ utilizando uma ponderação entre a amostra corrompida $s'(n)$ e o valor obtido para $s(n)$ a partir da iteração pelo mapa da estimativa do ponto anterior $\check{s}(n-1)$. A seguir, a suavização percorre a estimativa do passo de filtragem de trás para frente obtendo estimativas $\hat{s}(n)$ a partir de $\hat{s}(n+1)$, utilizando inversas do mapa $f_I(\cdot)$ em cada um dos trechos em que ele é uma bijeção.

As seguintes equações descrevem o processo (PAPADOPOULOS; WORNELL, 1993):

Filtragem

$$\begin{cases} \check{s}(0) = s'(0) \\ \check{s}(n) = \frac{3 \cdot 4^n s'(n) + (4^n - 1) f_I(\check{s}_{n-1})}{4^{n+1} - 1}, \text{ para } n = 1, 2, \dots, N-1 \end{cases} \quad (5.40)$$

Suavização

$$\begin{cases} \hat{s}(N-1) = \check{s}(N-1) \\ \hat{s}(n) = f_I^{-1}(\hat{s}(n+1), \check{s}(n)), \text{ para } n = N-2, N-3, \dots, 0 \end{cases} \quad (5.41)$$

com

$$f_I^{-1}(\hat{s}(n+1), \check{s}(n)) = \begin{cases} \frac{\alpha+1}{2} \hat{s}(n+1) + \frac{\alpha-1}{2}, & \text{se } \check{s}(n) < \alpha \\ \frac{\alpha-1}{2} \hat{s}(n+1) + \frac{\alpha+1}{2}, & \text{se } \check{s}(n) \geq \alpha \end{cases} \quad (5.42)$$

Para se medir o desempenho do estimador MLE, é usual utilizar o *ganho de estimação* definido por

$$G = \frac{\sigma_r^2}{\text{mse}(\hat{s}(0))} = \frac{E[(s'(0) - s(0))^2]}{E[(\hat{s}(0) - s(0))^2]}. \quad (5.43)$$

Os resultados da Seção 5.1 podem ser utilizados para determinar o máximo ganho de estimação possível. Como o MLE é assintoticamente não-enviesado, pelo Teorema 7 o CRLB aproximado dessa estimação é

$$\text{mse}(\hat{s}(0)) \geq \sigma_r^2 \frac{L_I^2 - 1}{L_I^{2N} - 1}. \quad (5.44)$$

Substituindo (5.44) na Eq. (5.43), conclui-se que, para o mapa tenda $f_T(\cdot)$ para o qual $L_I = 2$, o ganho é limitado aproximadamente por

$$G \leq \frac{4^N - 1}{3}. \quad (5.45)$$

O algoritmo foi testado para diversos valores de N variando-se a relação sinal-ruído (SNR_{in}) no canal. Para cada estimativa, essa relação foi calculada como

$$\text{SNR}_{in} = \frac{\sum_{n=0}^{N-1} s(n)^2}{N\sigma_r^2}. \quad (5.46)$$

Para cada valor de SNR_{in} obteve-se o ganho de estimação G tomando-se uma média de 10000 estimações de condições iniciais aleatórias. A Figura 5.9 mostra os resultados obtidos para o mapa tenda $f_T(\cdot)$ sendo que o ganho de estimação foi expresso em decibéis, $G_{dB} = 10 \log(G)$.

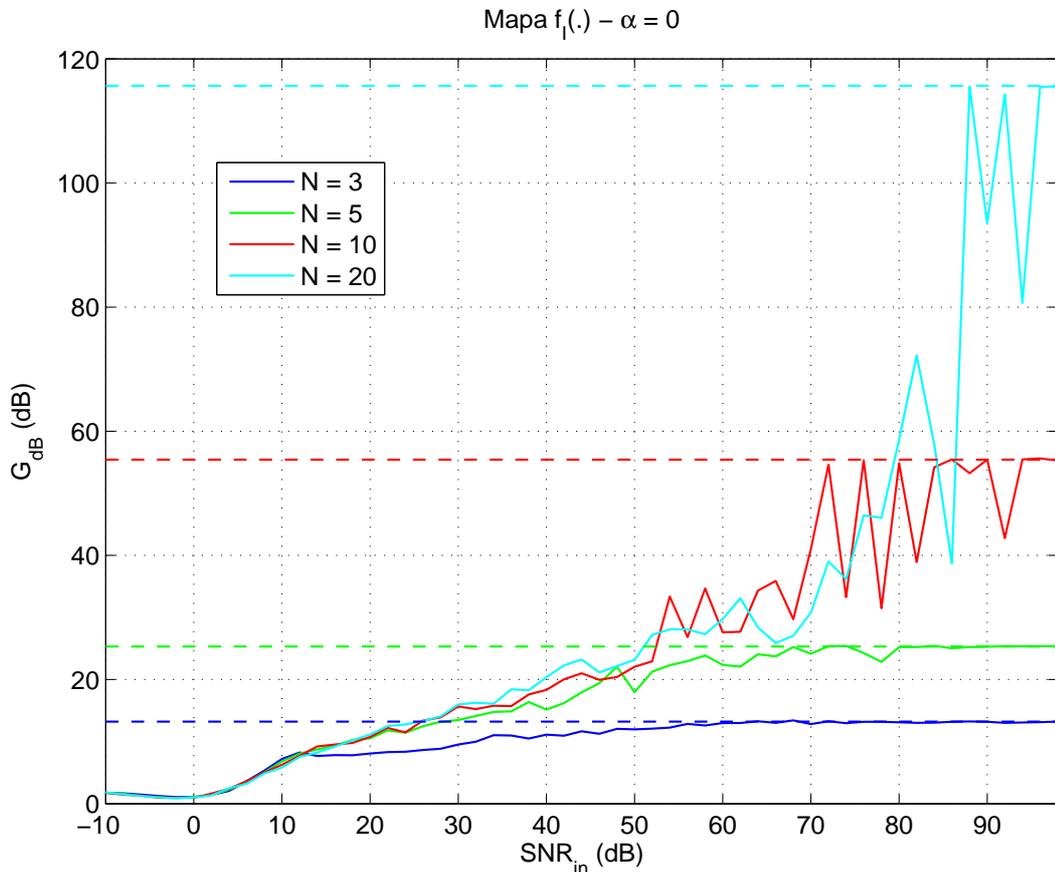


Figura 5.9: Ganho de estimação do MLE para condição inicial do mapa $f_T(\cdot)$. Em tracejado o limite para o ganho dado por (5.45).

Para cada valor de N a Figura 5.9 mostra em tracejado o limite teórico expresso na desigualdade (5.45). Verifica-se que os resultados estão de acordo com os limites previstos, confirmando novamente os resultados da seção anterior.

Os resultados da simulação mostram que o MLE sempre apresenta um ganho positivo (em dB) mesmo para valores de SNR e N baixos. O número de pontos usados na estimação afeta o ganho apenas para valores de SNR mais altos (a partir de 20dB ou 30dB) quando a estimativa para $N = 2$ já se aproxima do limite teórico do CRLB. Ou seja, quando o SNR_{in} é baixo, o aumento do número de pontos usados não beneficia a estimação de s_0 .

Quando $\alpha \neq 0$, pela desigualdade (5.44),

$$G \leq \frac{L_I^{2N} - 1}{L_I^2 - 1}, \quad (5.47)$$

com $L_I = \left(\frac{2}{\alpha+1}\right)^{\frac{\alpha+1}{2}} \left(\frac{2}{1-\alpha}\right)^{\frac{1-\alpha}{2}}$, como descrito no Capítulo 2.

A Figura 5.10 mostra a variação do ganho de estimação com o parâmetro α para uma SNR_{in} de (a) 20dB e (b) 90dB e diversos valores de N . Cada ponto dos gráficos foi obtido como uma média de 10000 estimações de condições iniciais aleatórias. Em tracejado é mostrado o limite assintótico aproximado dado em (5.47). Percebe-se que o melhor desempenho do mapa tenda $f_T(\cdot)$ ($\alpha = 0$) torna-se mais pronunciada para altas relações sinais-ruído, quando o ganho do MLE aproxima-se do limite teórico imposto pelo CRLB. Para SNR mais baixas, como também pode ser visto na Figura 5.9, o aumento do número de pontos não melhora o ganho do estimador da condição inicial.

5.2.3 Estimação de órbitas com o MLE

O algoritmo descrito pelas Eqs. (5.40) - (5.42) pode ser utilizado para se estimar a órbita inteira de um mapa tenda inclinada a partir da órbita corrompida por ruído.

Neste caso, define-se o ganho da estimação G como

$$G = \frac{\sigma_r^2}{e}, \quad (5.48)$$

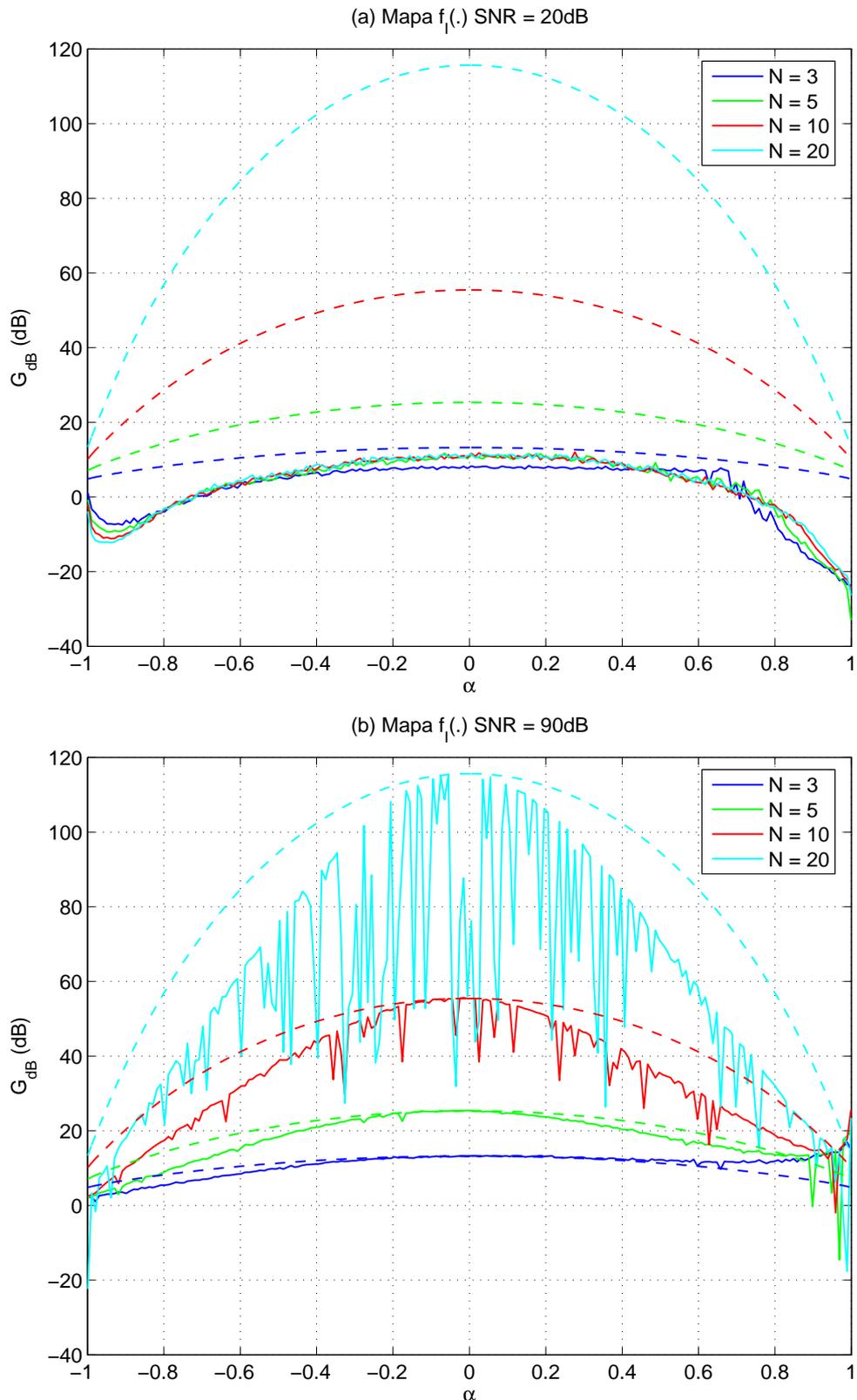


Figura 5.10: Ganho de estimação do MLE para condição inicial do mapa $f_I(\cdot)$ em função de α para (a) $\text{SNR}_{in} = 20\text{dB}$ e (b) $\text{SNR}_{in} = 90\text{dB}$. Em tracejado os limites teóricos da desigualdade (5.47).

sendo $e = \overline{(\hat{s}(n) - s(n))^2}$, com a média temporal calculada para $0 \leq n < N$. Em (PADOPOULOS; WORNELL, 1993), mostra-se que, neste caso,

$$G \leq (N + 1). \quad (5.49)$$

Nas curvas da Figura 5.11 pode-se verificar a variação do ganho de estimação em dB em função de SNR_{in} para diversos valores de N e o mapa tenda. Para cada valor de SNR_{in} obteve-se a estimação de 10000 órbitas com condições iniciais aleatórias. Em cada caso também é mostrado o limite do ganho de estimação dado pela desigualdade (5.49).

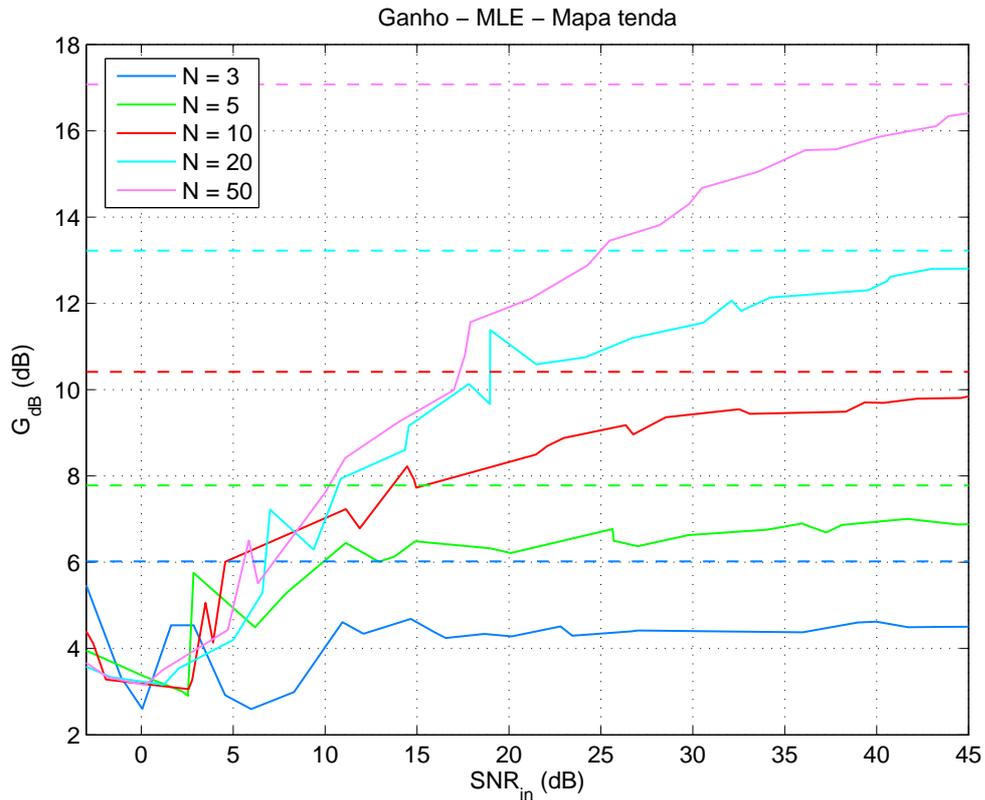


Figura 5.11: Ganho de estimação do MLE para órbitas do mapa $f_T(\cdot)$ para diversos valores de N . Os limites dados pela expressão (5.49) são indicados em tracejado.

Novamente, o ganho em dB é sempre positivo mas não é tão elevado quanto no caso da estimação da condição inicial. A estimação tende a ter maior erro para pontos próximos do final da órbita. Isso ocorre devido à dependência sensível com as condições iniciais das órbitas caóticas.

Na estimativa de órbitas, uma outra curva que é interessante na avaliação do desempenho

é a da SNR_{out} depois da estimação pela SNR_{in} antes da estimação. A SNR_{in} é definida pela Equação (5.46) e a SNR_{out} é definida por

$$SNR_{out} = \frac{\sum_{n=0}^{N-1} s(n)^2}{\sum_{n=0}^{N-1} (s(n) - \hat{s}(n))^2} \quad (5.50)$$

Na Figura 5.12 são traçadas curvas de SNR_{out} por SNR_{in} para os mesmos casos da Figura 5.11. Como esperado percebe-se uma sensível melhora na SNR após a estimativa da órbita.

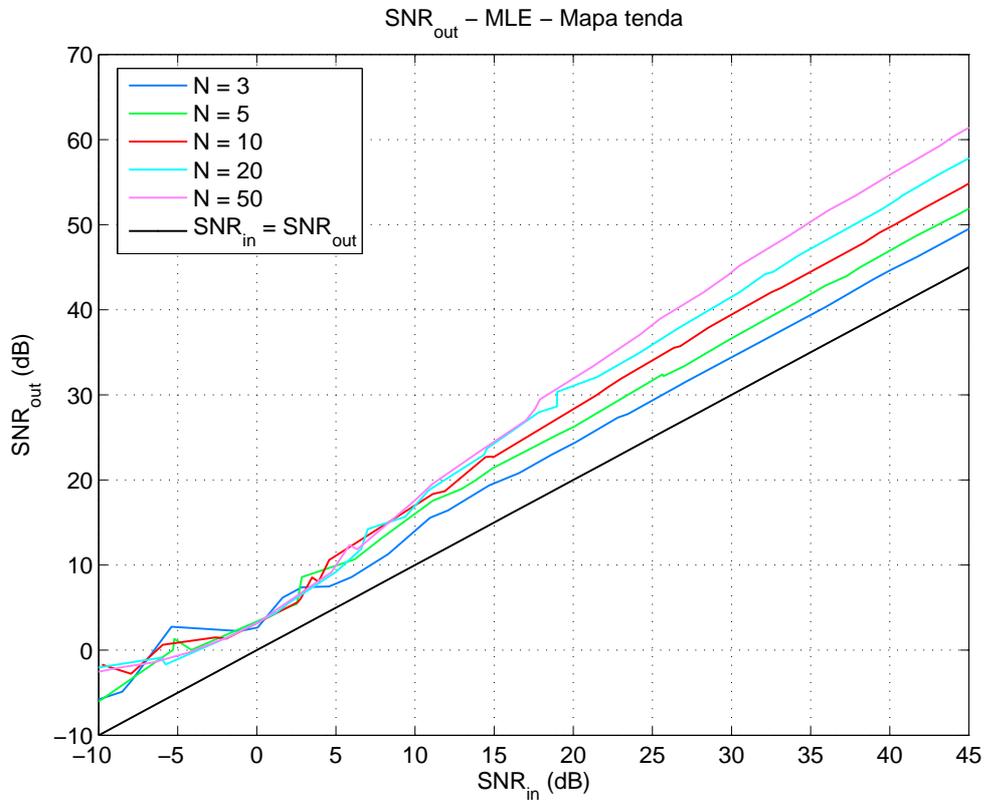


Figura 5.12: SNR na entrada e na saída do MLE para órbitas do mapa $f_T(\cdot)$.

O MLE apresentado só tem funcionalidade para mapas com densidade invariante uniforme. Por isso, antes de aplicar a estimação de órbitas a sistemas de comunicação, é necessário procurar modos de estimar a órbita de sinais caóticos gerados a partir de mapas com densidade invariante não-uniforme. Uma solução para esse problema é a estimação baseada no algoritmo de Viterbi discutida a seguir.

5.3 O algoritmo de Viterbi

A idéia básica subjacente ao algoritmo de Viterbi na estimação de órbitas caóticas é interpretar as seqüências geradas por mapas unidimensionais como um processo de Markov (HAYKIN, 2000) que em cada instante assume um de N_S estados possíveis. Esses estados são definidos a partir da partição do domínio U em N_S subintervalos.

Sejam U_i e U_j subintervalos da partição. Nesse caso, a densidade invariante pode ser substituída por uma matriz de transição de estados cujos elementos a_{ij} definem as probabilidades da amostra $s(n)$ estar no subconjunto U_j dado que $s(n-1)$ está no subconjunto U_i .

Nas próximas seções esses conceitos são introduzidos formalmente e uma série de simulações são utilizadas para ilustrá-los, usando a família de mapas tenda inclinada $f_I(\cdot)$ e o mapa quadrático $f_Q(\cdot)$.

5.3.1 Partição do espaço de estados

Um passo fundamental para aplicação da estimação com o algoritmo de Viterbi é considerar o espaço de estados U como a reunião de intervalos disjuntos U_j , $j = 1, 2, \dots, N_S$ que cubram todos os valores possíveis do sinal $s(n)$, ou seja, $\bigcup_{j=1}^{N_S} U_j = U = [-1, 1]$, isto é, uma partição de U .

Define-se que no instante n o sistema está no *estado* $q(n) = j$ se $s(n) \in U_j$.

O centro do intervalo U_j é denotado por $B(j)$.

Uma forma simples de se realizar a partição é a empregada por Dedieu e Kisel (1999) e consiste em utilizar intervalos de igual comprimento $\Delta_s = 2/N_S$. Assim, define-se

$$U_j = \begin{cases} [(j-1)\Delta_s - 1; j\Delta_s - 1[, & j = 1, 2, \dots, N_S - 1 \\ [1 - \Delta_s; 1], & j = N_S \end{cases} \quad (5.51)$$

Porém, por meio de simulações, observou-se que esta partição só fornece bons resultados na aplicação do algoritmo de Viterbi quando o mapa em questão tem densidade invariante uniforme, como no caso dos mapas $f_I(\cdot)$. Ao tentar-se aplicar o algoritmo proposto por Dedieu e Kisel (1999) ao mapa quadrático $f_Q(\cdot)$, os resultados ficaram muito abaixo do esperado como

será ilustrado na Seção 5.3.3.

A solução encontrada para o caso de mapas com densidade invariante não-uniforme é aplicar uma conjugação (vide Definição 15 do Capítulo 2) entre o mapa em estudo e outro com densidade invariante uniforme às extremidades dos intervalos U_j . Esse processo faz com que os pontos das órbitas distribuam-se uniformemente pelos novos subintervalos.

Por exemplo, no caso do mapa quadrático $f_Q(\cdot)$, aplica-se o mapa de conjugação entre $f_Q(\cdot)$ e $f_T(\cdot)$

$$C_{TQ}(s) = -\cos \frac{\pi(s+1)}{2} \quad (5.52)$$

às extremidades dos subintervalos U_j .

É interessante notar a semelhança desse método com a transformação de um processo estocástico qualquer em um ruído branco, conhecida como *branqueamento*, que aparece em problemas de predição linear (HAYKIN, 1996), representação espectral (PAPOULIS; PILLAI, 2002), modelagem em tempo discreto de canais com interferência intersimbólica (PROAKIS, 1995) entre outros.

Se N_S for suficientemente grande, pode-se considerar o problema de estimação da órbita corrompida por AWGN da seguinte forma: dado o vetor com N observações \mathbf{s}' , encontre a seqüência de estados $\hat{\mathbf{q}} = [\hat{q}(0), \hat{q}(1), \dots, \hat{q}(N-1)]^T$ que “melhor descreve” a órbita \mathbf{s} . Se existir um método de se encontrar esta seqüência “ótima” de subintervalos ocupados pelos pontos do vetor \mathbf{s} , então define-se a órbita estimada $\hat{\mathbf{s}}$ como a seqüência dos centros desses subintervalos mais prováveis, isto é,

$$\hat{\mathbf{s}} = [B(\hat{q}(0)), B(\hat{q}(1)), \dots, B(\hat{q}(N-1))]. \quad (5.53)$$

A determinação do vetor de estados estimado $\hat{\mathbf{q}}$ e da órbita estimada $\hat{\mathbf{s}}$ são tratadas na próxima seção.

5.3.2 O algoritmo de Viterbi para estimar órbitas caóticas

Para uma seqüência genérica \mathbf{q} de N estados, i.e.

$$\mathbf{q} = [q(0), q(1), \dots, q(N-1)]^T, \quad (5.54)$$

faz-se necessário definir algum critério de proximidade para a escolha da seqüência de estados estimada $\hat{\mathbf{q}}$ a partir do sinal observado \mathbf{s}' . Uma maneira é encontrar $\hat{\mathbf{q}}$ que maximize a probabilidade *a posteriori*

$$P(\hat{\mathbf{q}}|\mathbf{s}') = \max_{\mathbf{q}} P(\mathbf{q}|\mathbf{s}') \quad (5.55)$$

em que $P(\mathbf{q}|\mathbf{s}')$ é a probabilidade de que os pontos da órbita original ocupem os estados \mathbf{q} dado que se observou a seqüência \mathbf{s}' .

Pelo Teorema de Bayes (PAPOULIS; PILLAI, 2002),

$$P(\mathbf{q}|\mathbf{s}') = \frac{p(\mathbf{s}'|\mathbf{q})P(\mathbf{q})}{p(\mathbf{s}')}, \quad (5.56)$$

em que $p(\mathbf{s}')$ e $p(\mathbf{s}'|\mathbf{q})$ são a função densidade de probabilidade dos possíveis vetores \mathbf{s}' e a função densidade de probabilidade dos vetores \mathbf{s}' dado que o vetor de estados da órbita é \mathbf{q} , respectivamente. A probabilidade $P(\mathbf{q})$ é a chance de se obter a seqüência de estados \mathbf{q} quando o mapa gerador é iterado.

Assim, dado o vetor de observações \mathbf{s}' , deve-se encontrar o argumento $\hat{\mathbf{q}}$ tal que

$$\hat{\mathbf{q}} = \arg \max_{\mathbf{q}} P(\mathbf{q}|\mathbf{s}') = \arg \max_{\mathbf{q}} p(\mathbf{s}'|\mathbf{q})P(\mathbf{q}). \quad (5.57)$$

Para facilitar a notação, define-se \mathbf{q}_k o vetor dos primeiros $k+1$ estados de \mathbf{q} , isto é

$$\mathbf{q}_k = [q(0), q(1), \dots, q(k)]^T \quad (5.58)$$

e \mathbf{s}'_k as primeiras $k+1$ amostras observadas, ou seja,

$$\mathbf{s}'_k = [s'(0), s'(1), \dots, s'(k)]^T. \quad (5.59)$$

É importante observar que, devido à forma como as órbitas são geradas e a hipótese AWGN no modelo de canal, o processo \mathbf{q}_k é um processo de Markov discreto de primeira ordem³ considerando k como variável temporal. Assim,

$$P(\mathbf{q}_k) = P(q(k)|q(k-1)) P(\mathbf{q}_{k-1}) \quad (5.60)$$

em que $P(q(k)|q(k-1))$ é a probabilidade de transição do estado $q(k-1)$ para o estado $q(k)$.

Além disso, devido à independência entre as amostras de ruído,

$$p(\mathbf{s}'_k|\mathbf{q}_k) = \prod_{n=0}^k p(s'(n)|q(n)) = \prod_{n=0}^k p_r(s'(n) - s(n)) \approx \prod_{n=0}^k p_r(s'(n) - B(q(n))). \quad (5.61)$$

em que $p_r(\cdot)$ é a função densidade de probabilidade do ruído $r(n)$. A expressão (5.61) só é verdadeira se o número de subintervalos N_S for suficientemente grande.

Usando as Eqs. (5.57), (5.60) e (5.61), pode-se expressar $P(\mathbf{q}|\mathbf{s}')$ como um produtório de probabilidades de transição de estados por probabilidades de observação condicionais, ou seja,

$$\hat{\mathbf{q}} = \arg \max_{\mathbf{q}} \prod_{n=1}^{N-1} P(q(n)|q(n-1)) p(s'(n)|q(n)). \quad (5.62)$$

Encontrar \mathbf{q} que maximiza o produtório da Eq. (5.62) é um problema clássico cuja solução eficiente é dada pelo algoritmo de Viterbi (VITERBI, 1967; HAYKIN, 2000), utilizado pela primeira vez para estimação de sinais caóticos por Marteau e Abarbanel (1991). Com o algoritmo de Viterbi, evita-se fazer uma busca exaustiva entre os $(N_S)^N$ seqüências de estados possíveis para uma órbita de comprimento N .

Seja $\gamma(n, j)$ a probabilidade da seqüência de estados mais provável (no sentido de máxima verossimilhança) que termina no estado j , no instante n , dada a seqüência observada \mathbf{s}' . Assim,

$$\gamma(n, j) = \max_{\mathbf{q}_n} P(\mathbf{q}_{n-1}, q(n) = j|\mathbf{s}'). \quad (5.63)$$

³Um processo de Markov é um processo estocástico que assume valores de um conjunto finito a cada instante. Se o valor do processo no instante n depende apenas de seu valor no instante $n-1$ então o processo é chamado de primeira ordem (PAPOULIS; PILLAI, 2002).

Das Eqs. (5.60) e (5.61) segue que $\gamma(n, j)$ pode ser escrita da seguinte forma recursiva:

$$\gamma(n, j) = \max_i [\gamma(n-1, i) a_{ij}] b_j(s'(n)) \quad (5.64)$$

em que

$$a_{ij} = P(q(n) = j | q(n-1) = i) \quad (5.65)$$

e

$$b_j(s'(n)) = p(s'(n) | q(n) = j). \quad (5.66)$$

Os coeficientes a_{ij} são as probabilidades de transição de estados que dependem do mapa $f(\cdot)$ e da partição usada. Os coeficientes $b_j(\cdot)$ são as probabilidades condicionais de observação que dependem apenas da função densidade de probabilidade do ruído $p_r(\cdot)$.

O algoritmo de Viterbi trabalha em duas fases: a de avanço e a de retrocesso.

- **Fase de avanço:** a Eq. (5.64) é usada para calcular $\gamma(n, j)$ no instante n para os N_S nós ($n, j = 1, \dots, N_S$). Dentre os N_S caminhos que podem ligar os nós ($n-1, j = 1, \dots, N_S$) ao nó (n, j), apenas o melhor é mantido, como mostrado no exemplo da Figura 5.13. Ao final dessa fase, no instante $n = N-1$, seleciona-se o nó ($N-1, j^*$) que fornece a maior probabilidade.

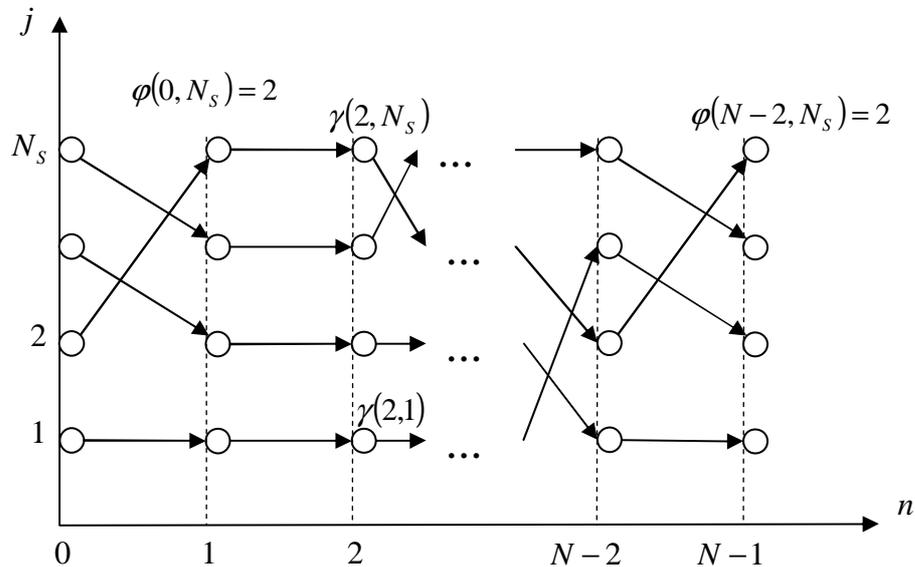


Figura 5.13: Exemplo de estrutura para a decodificação usando o algoritmo de Viterbi. Neste caso, $N_S = 4$.

- **Fase de retrocesso:** para obter a seqüência de estados mais provável, deve-se considerar o argumento i que maximiza a Eq. (5.64) para cada n e j . Isto pode ser feito através da matriz $\varphi(n, j)$ na qual é armazenado o estado no instante $n - 1$ que leva ao nó j no instante n com máxima probabilidade. Exemplos de valores de elementos dessa matriz são mostrados na Figura 5.13.

Seja π_i a probabilidade *a priori* de $s'(0)$ estar no estado i (se a densidade invariante do mapa for uniforme, pode-se considerar $\pi_i = 1/N_S$). O algoritmo de Viterbi completo é o seguinte.

Inicialização

$$\begin{cases} \gamma(0, i) = \pi_i b_i(s'(0)), & 1 \leq i \leq N_S \\ \varphi(0, i) = 0 \end{cases} \quad (5.67)$$

Fase de avanço

$$\begin{cases} \gamma(n, j) = \max_{1 \leq i \leq N_S} [\gamma(n-1, i) a_{ij}] b_j(s'(n)), & 1 \leq n \leq N-1, 1 \leq j \leq N_S \\ \varphi(n, j) = \arg \max_{1 \leq i \leq N_S} [\gamma(n-1, i) a_{ij}] b_j(s'(n)), & 1 \leq n \leq N-1, 1 \leq j \leq N_S \end{cases} \quad (5.68)$$

Término do avanço

$$\hat{q}(N-1) = \arg \max_{1 \leq i \leq N_S} [\gamma(N-1, i)] \quad (5.69)$$

Fase de retrocesso

$$\hat{q}(n) = \varphi(n+1, \hat{q}(n+1)), \quad n = N-2, \dots, 0. \quad (5.70)$$

Obtida a seqüência de estados estimada $\hat{q}(n)$, a órbita estimada é dada pelo centro dos subintervalos mais prováveis, como definida pela Eq. (5.53).

É importante observar que para se aplicar este algoritmo é necessário que o receptor conheça a função densidade de probabilidade do ruído $p_r(\cdot)$, necessária para o cálculo de $b_j(s'(n))$. Porém, Dedieu e Kisel (1999) propõem uma aproximação para o cálculo de $b_j(s'(n))$ que apresenta resultados razoáveis e que é usada nas simulações a seguir.

5.3.3 Simulações e resultados numéricos

Nesta seção ilustra-se a utilização do algoritmo de Viterbi por meio de simulações numéricas para melhorar a SNR de órbitas caóticas imersas em AWGN com média nula. Inicia-se com o cálculo da matriz de transição de estados

$$\mathbf{A} = \{a_{ij}\}_{1 \leq i, j \leq N_S}, a_{ij} = P(q(n) = j | q(n-1) = i). \quad (5.71)$$

A seguir, estuda-se o desempenho do estimador, inclusive estabelecendo seus limites quando se varia: o mapa usado; o número de pontos da órbita N ; e o número de subconjuntos N_S .

Cálculo da matriz de transição de estados

Considerando uma partição do intervalo $U = [-1, 1]$ em N_S subintervalos, um passo fundamental da aplicação do algoritmo de Viterbi para a estimação de órbitas caóticas é o cálculo da matriz de transição de estados \mathbf{A} da Eq. (5.71).

Nas simulações, \mathbf{A} foi calculada de forma numérica: dado um conjunto de condições iniciais distribuídas uniformemente em U , determina-se a porcentagem delas que cai em cada subintervalo U_j , após uma iteração do mapa $f(\cdot)$.

Como exemplo ilustrativo, considera-se o mapa tenda e $N_S = 5$. Assim, os subintervalos neste caso são:

$$U_1 = [-1; -0,6[, U_2 = [-0,6; -0,2[, U_3 = [-0,2; 0,2[, U_4 = [0,2; 0,6[e U_5 = [0,6; 1]. \quad (5.72)$$

Esses intervalos são mostrados na Figura 5.14.

Por exemplo, os pontos do subintervalo U_4 são mapeados por $f_T(\cdot)$ de forma uniforme sob os intervalos U_3 e U_4 . Assim, pela definição da Eq. (5.71), a quarta coluna da matriz de transição de estados A neste caso é formada por

$$a_{14} = 0, a_{24} = 0, a_{34} = 0,5, a_{44} = 0,5 e a_{54} = 0. \quad (5.73)$$

Procedendo-se de forma semelhante com os pontos dos demais intervalos, chega-se à seguinte

matriz que, neste caso, pode ser obtida de forma exata:

$$\mathbf{A} = \begin{bmatrix} 0,5 & 0,5 & 0 & 0 & 0 \\ 0 & 0 & 0,5 & 0,5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0,5 & 0,5 & 0 \\ 0,5 & 0,5 & 0 & 0 & 0 \end{bmatrix}. \quad (5.74)$$

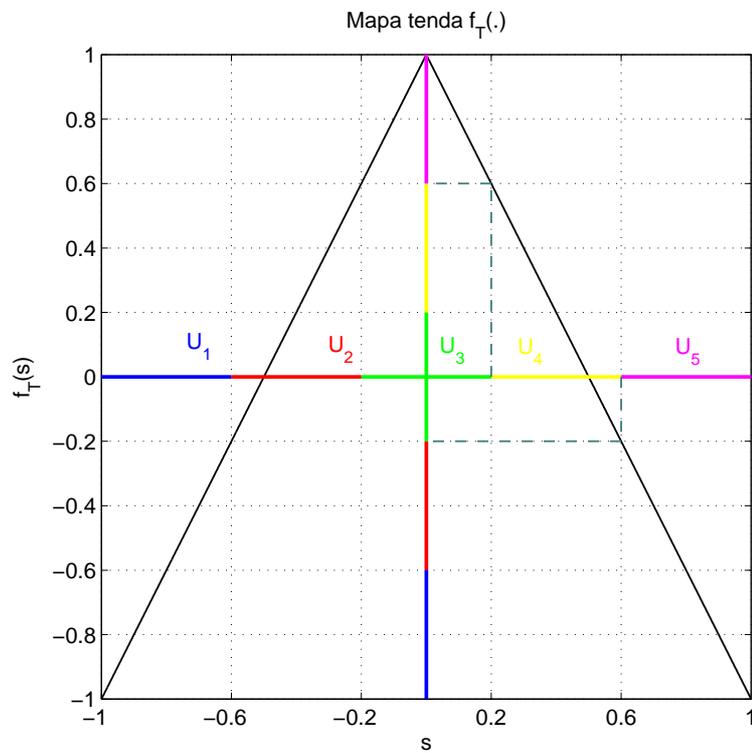


Figura 5.14: Subintervalos para o cálculo da matriz de transição de estados para o mapa $f_T(\cdot)$ com $N_S = 5$.

No caso do mapa quadrático, como já visto na Seção 5.3.1, para que o algoritmo tenha um desempenho bom é necessário aplicar o mapa de conjugação $C_{TQ}(s)$ da Eq. (5.52) na definição dos intervalos U_j . Assim, neste caso, para $N_S = 5$, os intervalos são:

$$\begin{aligned} U_1 &= [-1; -0,8090[, U_2 = [-0,8090; -0,3090[, U_3 = [-0,3090; 0,3090[, \\ U_4 &= [0,3090; 0,8090[e U_5 = [0,8090; 1]. \end{aligned} \quad (5.75)$$

Esses intervalos são mostrados na Figura 5.15.

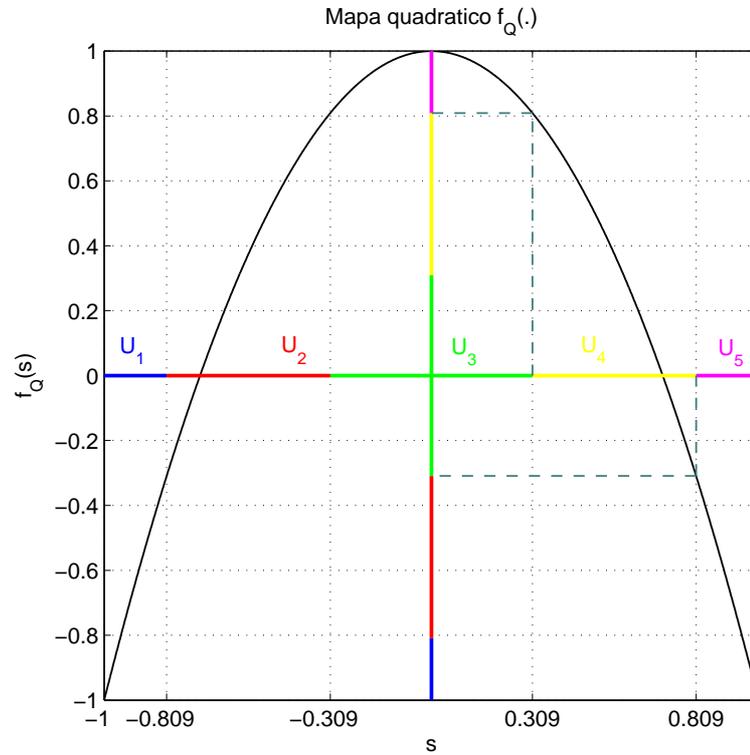


Figura 5.15: Subintervalos para o cálculo da matriz de transição de estados para o mapa $f_Q(\cdot)$ com $N_S = 5$.

Procedendo de maneira idêntica ao exemplo anterior, chega-se à mesma matriz de transição de estados da Eq. (5.74).

Simulações e desempenho do algoritmo de Viterbi

Nesta seção é estudado o desempenho da estimação utilizando o algoritmo de Viterbi por meio de simulações numéricas. Como medidas de qualidade são utilizados o ganho de estimação G definido na Eq. (5.43) e a comparação entre a SNR_{out} na saída com a SNR_{in} na entrada do estimador. Essas grandezas foram definidas nas Equações (5.46) e (5.50) respectivamente.

A Figura 5.16 mostra os resultados da estimação de um trecho de órbita do mapa $f_I(\cdot)$ com $N = 10$ pontos e usando $N_S = 20$ subintervalos. Para cada valor de SNR_{in} tomou-se uma média de 1000 estimações de órbitas com condições iniciais aleatórias. Diversos valores do parâmetro α do mapa foram usados.

O algoritmo de estimação tem um ganho positivo para valores baixos de SNR_{in} , chegando próximos de um valor de saturação e depois decaindo fortemente até ganhos negativos. Ou

seja, para valores acima de aproximadamente 22dB, a SNR na entrada do estimador é maior do que na saída.

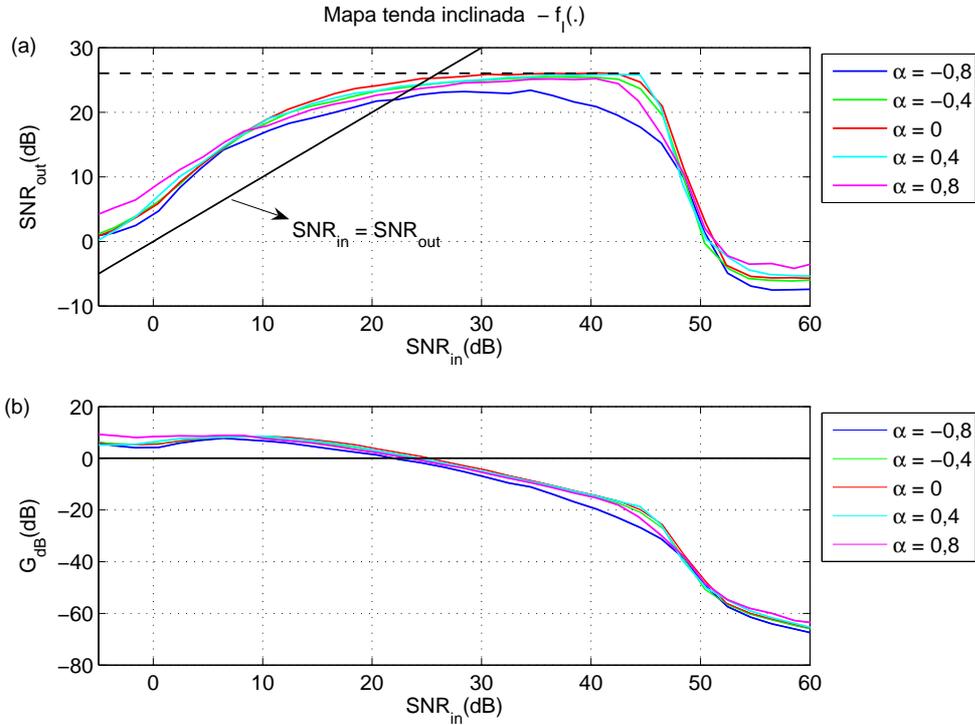


Figura 5.16: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para mapas da família tenda inclinada para diversos valores de α , $N_S = 20$ e $N = 10$. A limitação da Eq. (5.76) é mostrada em tracejado.

Esse fato é justificado pela partição do espaço de estados. Quando a potência do ruído é baixa, o erro introduzido pelo algoritmo ao usar os centros dos intervalos $B(j)$ como pontos da órbita é maior do que o erro devido ao ruído original.

Em relação a variação com α , o algoritmo tem um ganho praticamente idêntico para todos esses mapas, apesar de terem diferentes matrizes de transição de estados e suas órbitas terem diferentes números de Lyapunov. O que eles têm em comum é a densidade invariante uniformemente distribuída no intervalo $[-1, 1]$, como visto na Seção 2.6.

O número de Lyapunov do mapa afeta pouco a qualidade da estimação da órbita ao se utilizar o algoritmo de Viterbi.

O limite para o valor de SNR_{out} indicado pela linha tracejada na Figura 5.16 pode ser calculado admitindo-se que no melhor caso o erro de estimação é causado apenas pela partição. Assim, para cada n , o erro de estimação é uma variável aleatória uniformemente distribuída no

intervalo $[-1/N_S, 1/N_S]$ já que $s(n)$ no intervalo U_j de comprimento $2/N_S$ está sendo aproximado pelo centro do intervalo.

Como a média quadrática de uma variável aleatória uniformemente distribuída no intervalo $[-a, a]$ é $a^2/3$, a média quadrática de $s(n) - \hat{s}(n)$ é, limitada por $1/(3N_S^2)$. Da mesma forma, $s(n)$ está distribuído uniformemente no intervalo $[-1, 1]$, tendo, portanto, média quadrática $1/3$. Assim, se todos os pontos de $\hat{s}(n)$ estiverem nos subintervalos corretos, o valor esperado de SNR_{out} será

$$E[\text{SNR}_{out}] = E \left[10 \log \frac{\sum_{n=0}^{N-1} s(n)^2}{\sum_{n=0}^{N-1} (s(n) - \hat{s}(n))^2} \right] = 10 \log \frac{N/3}{N/(3N_S^2)} = 20 \log N_S. \quad (5.76)$$

Daí o valor limite indicado na Figura 5.16 em $20 \log N_S = 20 \log 20 \approx 26\text{dB}$.

Na Figura 5.17 é apresentado como varia o desempenho da estimação em função de diversos N , para $N_S = 20$ e o mapa tenda $f_T(\cdot)$, ou seja, $\alpha = 0$.

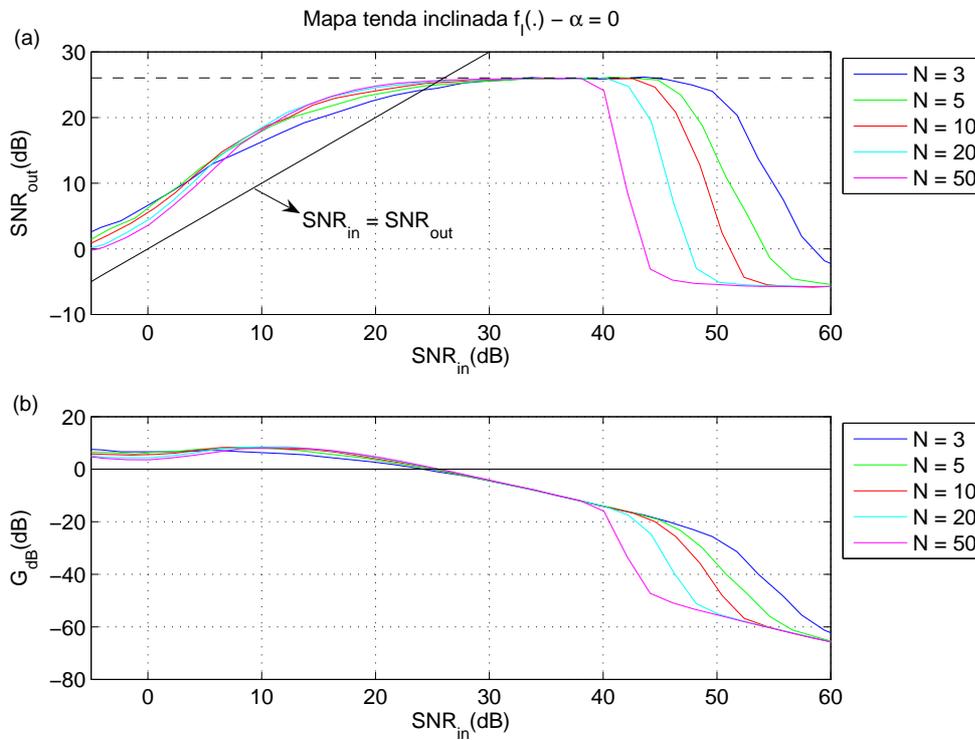


Figura 5.17: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa $f_T(\cdot)$ e diversos comprimentos N da seqüência sendo estimada. $N_S = 20$. A limitação da Eq. (5.76) é mostrada em tracejado.

Ainda na Figura 5.17, na faixa em que o ganho é positivo (relação sinal-ruído no canal

menor do que 25dB) percebe-se que as curvas indicam pouca variação com N . Daí, pode-se concluir que a maior parcela de erro na estimação está concentrada nas primeiras amostras do sinal, bem diferente do que ocorre com o MLE cuja dependência com N é bastante forte conforme mostrado na Figura 5.11.

Para $N = 10$, na Figura 5.18 pode-se ver a variação do desempenho com o número de subconjuntos N_S usados e os respectivos limites dados pela Eq. (5.76). Como esperado, quanto maior N_S , maior a faixa de valores de SNR_{in} em que o ganho da estimação é positivo.

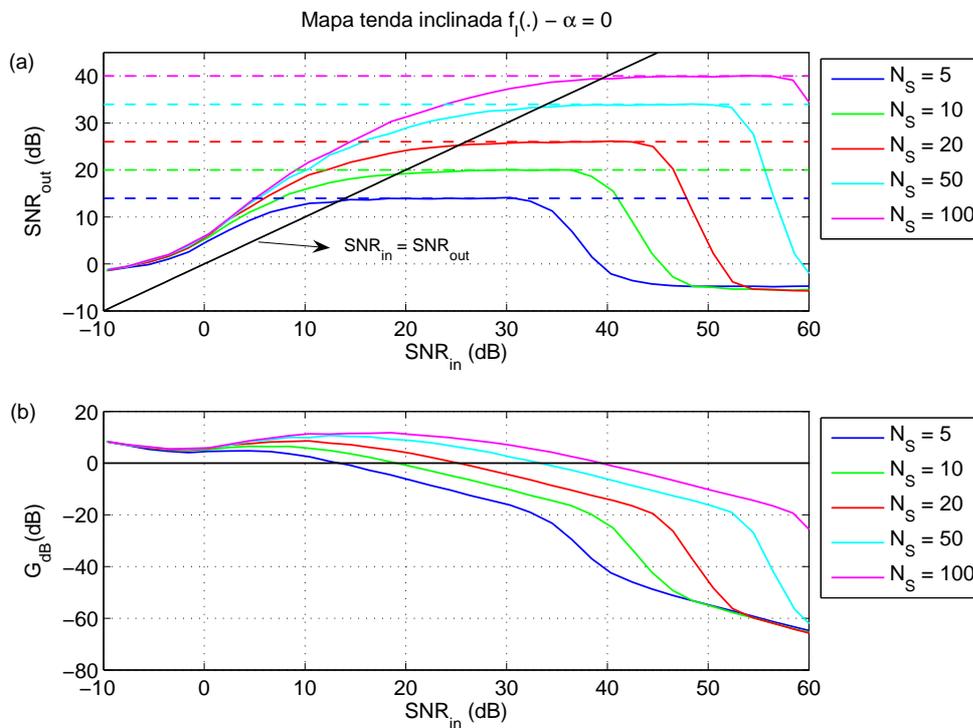


Figura 5.18: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa $f_T(\cdot)$ para diversos valores de N_S , o número de subintervalos em que $[-1, 1]$ foi dividido. As limitações da Eq. (5.76) são mostradas em tracejado.

Para cada valor de N_S , na Tabela 5.1 é indicado o maior valor de SNR_{in} que resulta um ganho positivo do algoritmo. Este valor permite precisar na Figura 5.18(a) a intersecção da reta $SNR_{in} = SNR_{out}$ com as várias curvas parametrizadas por N_S . A referida tabela também apresenta o valor limite de $20 \log N_S$ calculado na Eq. (5.76). As duas colunas apresentam valores muito próximos. A partir desse ponto não compensa aplicar o algoritmo de Viterbi devido aos erros introduzidos pelo partição do intervalo U .

Assim, dada uma SNR_{in} no canal, para que o algoritmo produza ganho positivo deve-se

usar um número de intervalos na determinação da matriz de transição de estados dado por

$$N_S \geq 10^{\text{SNR}_{in}/20}. \quad (5.77)$$

Tabela 5.1: Valores máximos de SNR_{in} para ganho de estimação positivo.

N_S	SNR_{in} máximo (dB)	$20 \log N_S$ (dB)
5	13,4	14,0
10	19,1	20,0
20	25,2	26,0
50	33,3	34,0
100	40,0	40,0

Nas Figuras 5.19 e 5.20 são mostradas curvas semelhantes às das Figuras 5.17 e 5.18 respectivamente, mas para o mapa quadrático $f_Q(\cdot)$ tendo sido usada a transformação nos subintervalos dada pela Eq. (5.52). Os resultados são muito parecidos, apesar de que os limites mostrados agora são apenas aproximados, não tendo sido usados intervalos com comprimento uniforme.

A partir das simulações, chega-se a alguns resultados interessantes deste trabalho no que se refere ao algoritmo de Viterbi:

- se aplicado o mapa de conjugação aos intervalos da partição, os resultados obtidos para os mapas $f_T(\cdot)$ e $f_Q(\cdot)$ são praticamente idênticos, como é visto na comparação entre as Figuras 5.17 e 5.18 com 5.19 e 5.20, respectivamente;
- se a simulação é feita sem aplicação do mapa de conjugação aos intervalos na estimação de órbitas do mapa $f_Q(\cdot)$, os resultados ficam muito aquém em relação aos obtidos com a aplicação, como se pode ver comparando-se, respectivamente, as Figuras 5.19 e 5.20 com 5.21 e 5.22. Por exemplo, sem a conjugação, para SNR_{in} maior do que 15dB, o algoritmo apresenta ganho negativo, independentemente do número de intervalos e pontos utilizados;

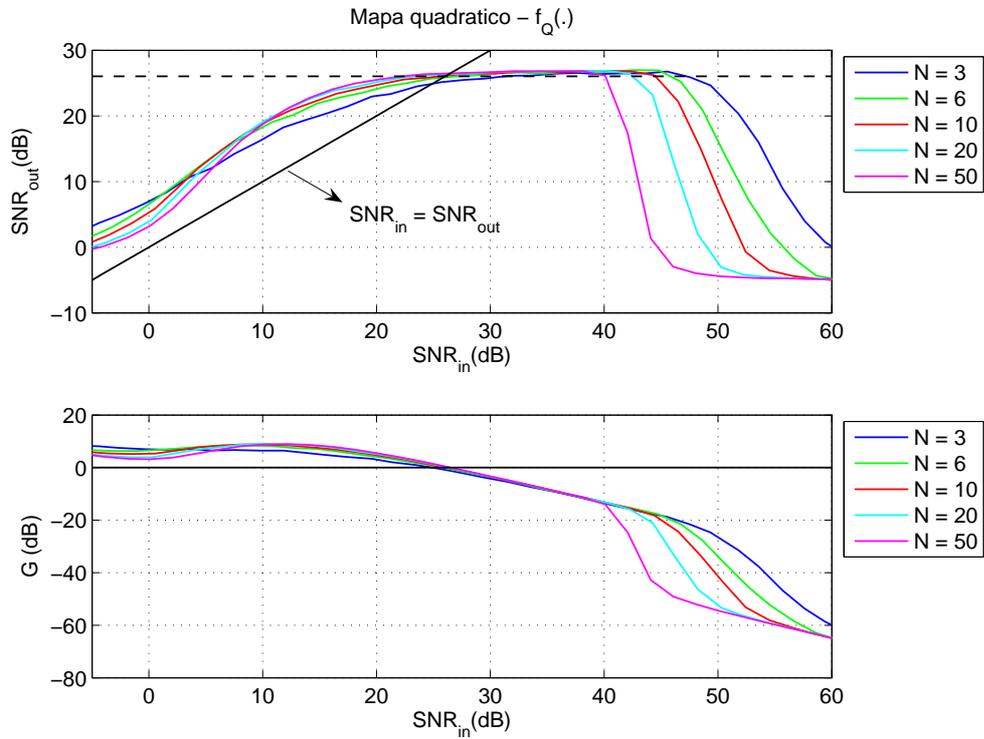


Figura 5.19: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi utilizando a conjugação para o mapa quadrático para diversos comprimentos N da sequência estimada e $N_S = 20$. A limitação da Eq. (5.76) é mostrada em tracejado.

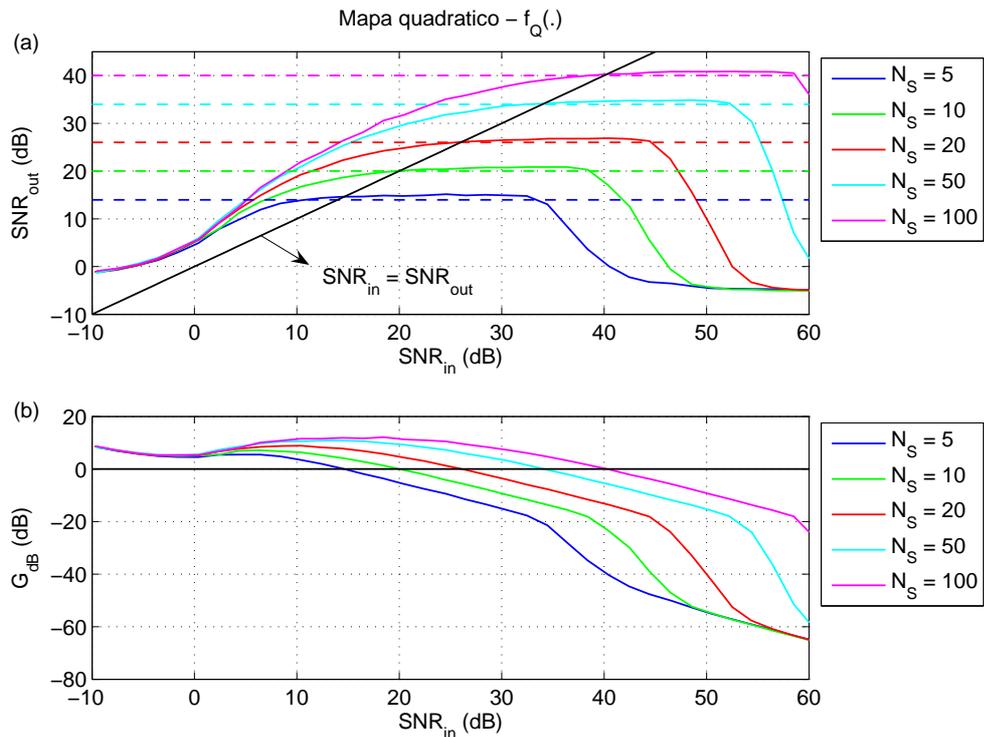


Figura 5.20: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi utilizando a conjugação para o mapa quadrático para diversos valores de N_S e $N = 10$. As limitações da Eq. (5.76) são mostradas em tracejado.

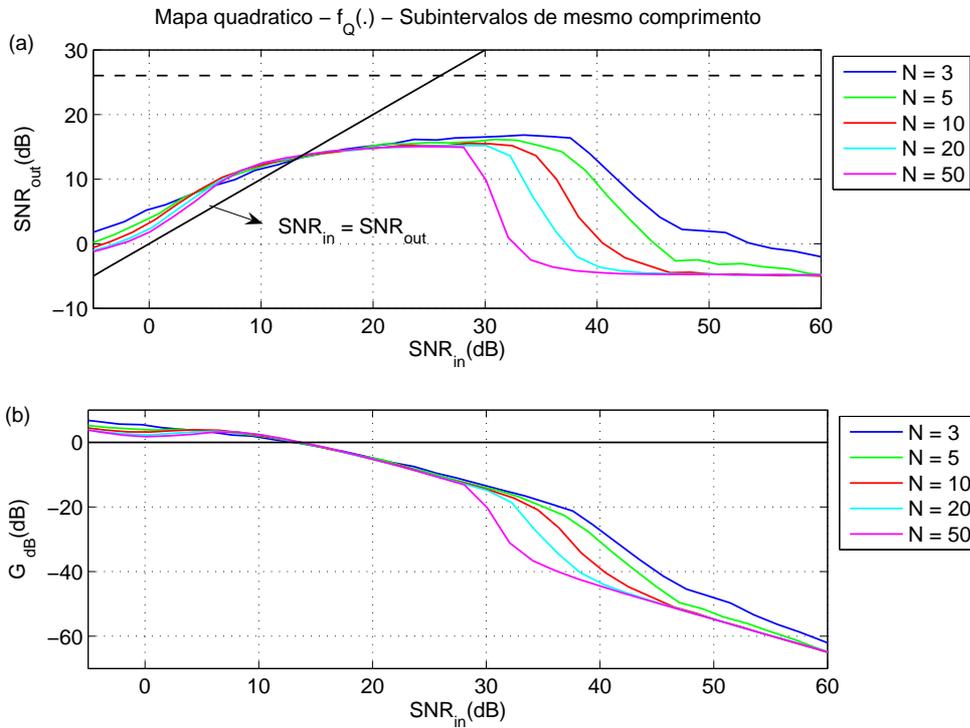


Figura 5.21: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa quadrático para diversos valores de N sem aplicação do mapa de conjugação. A limitação da Eq. (5.76) é mostrada em tracejado.

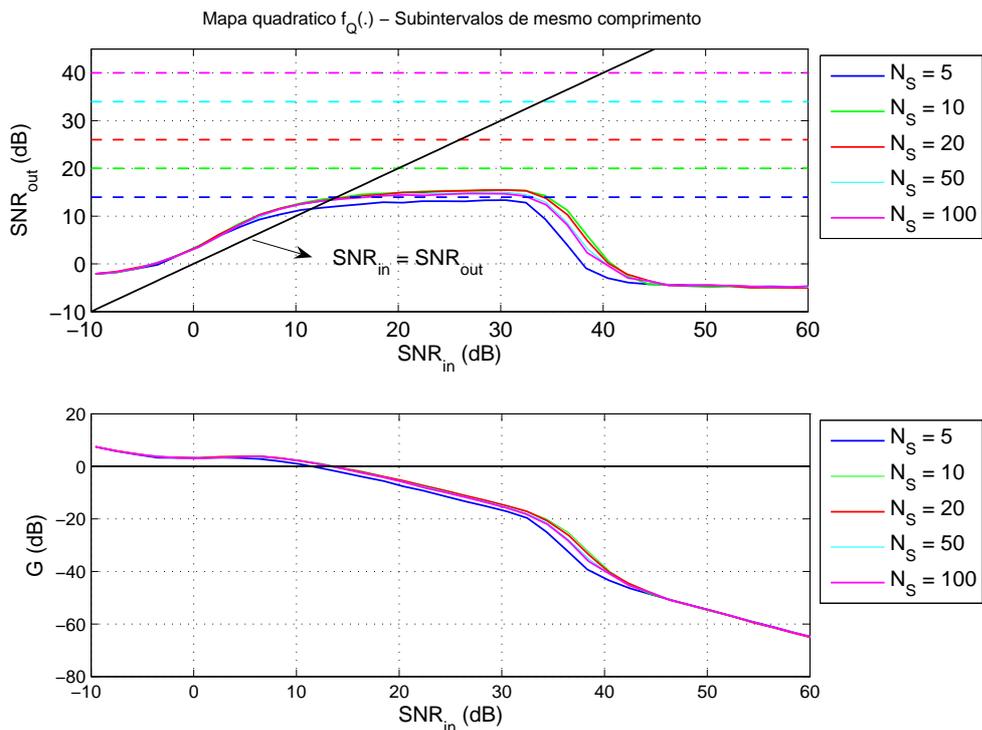


Figura 5.22: (a) SNR_{out} e (b) ganho de estimação G do algoritmo de Viterbi para o mapa quadrático para diversos valores de N_s sem aplicação do mapa de conjugação. As limitações da Eq. (5.76) são mostradas em tracejado.

- quando a densidade invariante do mapa envolvido não é uniforme, a aplicação de uma conjugação entre este mapa e outro com densidade invariante uniforme aos intervalos da partição é importante para garantir resultados adequados para a estimação.

5.4 Comparações entre o algoritmo de Viterbi e o MLE

O desempenho do MLE na estimação de órbitas é fortemente influenciado pelo número de pontos da órbita N , como mostra a desigualdade (5.49). O limite de desempenho do algoritmo de Viterbi é mais sensível ao número N_S de subconjuntos usados na partição, como mostra a expressão (5.76).

O ganho de estimação via MLE apresenta crescimento aproximadamente monotônico com a SNR_{in} a partir de $\text{SNR}_{in} \approx 0$. Já com a aplicação do algoritmo de Viterbi, após o ganho crescer até um valor máximo, ele decai podendo ser negativo.

A curva de ganho para os dois algoritmos para estimação de órbitas do mapa $f_T(\cdot)$ é mostrada na Figura 5.23. No caso do algoritmo de Viterbi escolheu-se apenas $N = 20$ já que, como mostra a Figura 5.17, seu desempenho praticamente se repete para outros valores.

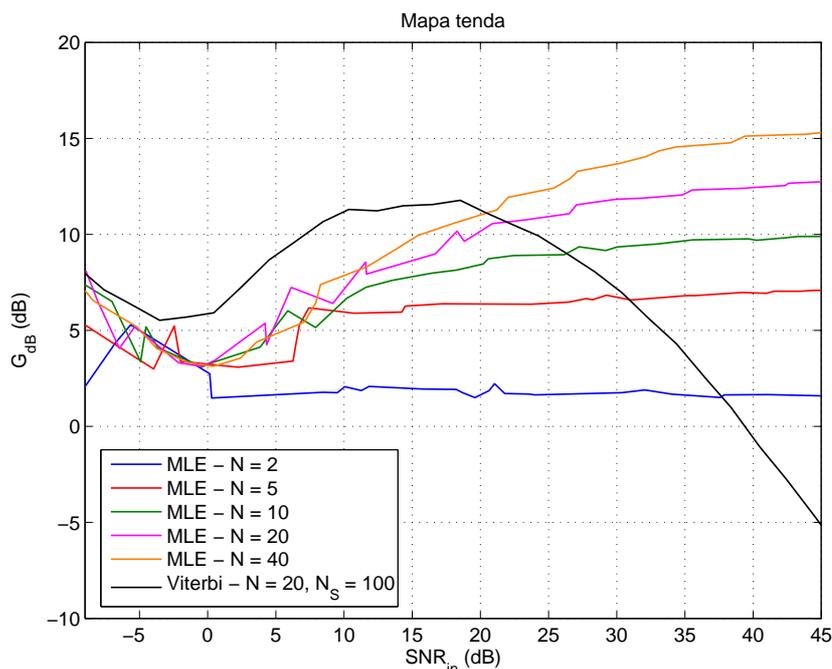


Figura 5.23: Curvas de ganho para o MLE e para o algoritmo de Viterbi para estimação de uma órbita do mapa tenda $f_T(\cdot)$.

Na Figura 5.23 verifica-se que para SNR_{in} entre 0 e 20dB, que é usualmente a condição de interesse, o algoritmo de Viterbi tem um desempenho melhor. Apenas para SNR_{in} acima de 20dB para $N = 50$ o desempenho do MLE é superior; para $N = 3$, a ultrapassagem ocorre perto de 35dB.

Outra distinção entre o MLE e o algoritmo de Viterbi é que o primeiro tem menos complexidade e requer menos memória.

As comparações feitas levam à escolha do algoritmo de Viterbi na implementação dos sistemas de comunicações que são apresentados no próximo capítulo.

5.5 Conclusões

Este capítulo apresentou resultados a respeito da estimação de sinais caóticos e de suas condições iniciais quando imersos em AWGN.

Após um estudo da estimação da condição inicial de órbitas caóticas, mais precisamente, dos limites que estimadores não-enviesados dessas condições podem alcançar, deduziu-se que o CRLB para este problema é função do número de Lyapunov da órbita, desde que seja usado um número suficiente de pontos na estimação. Os Teoremas 6 e 7 expressam esses resultados e, da forma na qual foram enunciados, são um resultado original da tese.

Em seguida, descreveu-se o algoritmo MLE para a família de mapas tenda inclinada que possuem densidade invariante uniforme. Mostrou-se que os resultados são compatíveis com os limites deduzidos no caso da estimação da condição inicial. Esse algoritmo também foi utilizado na estimação de órbitas inteiras resultando nas Figuras 5.11 e 5.12.

Para obter um algoritmo que se aplicasse a estimação de órbitas de mapas mais genéricos, deduziu-se o algoritmo de estimação de Viterbi, nos moldes propostos por Dedieu e Kisel (1999). Porém, esse algoritmo só apresenta bons resultados para mapas com densidade invariante uniforme. Para que se aplique a mapas com densidade invariante diferente, mostra-se que é necessário modificar por mapas de conjugação os intervalos em que o domínio do mapa é subdividido sendo esta outra contribuição original desta tese.

Comparando-se por simulações os dois algoritmos, conclui-se que, na faixa de valores de

relação sinal-ruído em que eles são efetivamente empregados, o algoritmo de Viterbi leva vantagem nos casos em que o MLE pode ser aplicado.

No próximo capítulo, examinam-se sistemas de comunicação que utilizam o algoritmo de Viterbi para obter valores de SER próximos ou melhores do que os descritos no Capítulo 4.

Os enunciados dos Teoremas deste capítulo são listados na Tabela 5.2.

Tabela 5.2: Teoremas do Capítulo 5

Teorema	Enunciado
5	<p>Considere um vetor aleatório \mathbf{x} dependente de um parâmetro escalar θ e uma função densidade de probabilidade $p(\mathbf{x};\theta)$ que satisfaz a condição de “regularidade” $E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$, $\theta \in \Theta$, sendo Θ o conjunto de valores admissíveis para θ. Neste caso, o mse de qualquer estimador não-enviesado $\hat{\theta}$ satisfaz $\text{mse}(\hat{\theta}) \geq \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]}$ em que a derivada é tomada no valor verdadeiro de θ e a esperança é tomada com relação a $p(\mathbf{x};\theta)$, com θ constante. O valor no segundo membro da desigualdade anterior é chamado de <i>limite inferior de Cramér-Rao</i> (CRLB). Além disso, pode ser encontrado um estimador não-enviesado que atinja este limite para todo θ se e somente se existirem funções $h(\mathbf{x})$ e $I(\theta)$ tais que $\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(h(\mathbf{x}) - \theta)$. Esse estimador, que é o MVU, é $\hat{\theta} = h(\mathbf{x})$ e o mse é $1/I(\theta)$.</p>
6	<p>Considere uma órbita $s(n, s_0)$ do sistema dinâmico $s(n+1, s_0) = f(s(n))$ sendo que o mapa $f(\cdot)$ possui derivada em todos os pontos dessa órbita. Seja $r(n)$ um processo ruído branco gaussiano de média nula e variância σ_r^2 e o sinal corrompido por ruído observado $s'(n) = s(n) + r(n)$, $0 \leq n < N$. Então, o CRLB na estimação de s_0 dado $s'(n)$ e o mapa $f(\cdot)$ é $\text{mse}(\hat{s}_0) \geq \frac{\sigma_r^2}{1 + \sum_{n=1}^{N-1} \left(\prod_{j=0}^{n-1} \frac{df}{ds} _{s(j, s_0)}\right)^2}$.</p>
7	<p>São dados os mesmos vetores e condições do Teorema 6. Considerando-se uma órbita $s(n, s_0)$, o limite do CRLB quando $N \rightarrow \infty$ é $\text{mse}(\hat{s}_0) \geq \sigma_r^2 \frac{L^2 - 1}{L^{2N} - 1}$, em que $L \equiv L(s_0) \neq 1$ é o número de Lyapunov do atrator para o qual a órbita $s(n, s_0)$ converge.</p>

Capítulo 6

Modulação digital utilizando estimação de sinais caóticos

Este capítulo propõe três sistemas de modulação digital que utilizam sinais caóticos à luz das técnicas de estimação discutidas no Capítulo 5. Cada sistema é analisado e seus desempenhos são comparados.

Os sistemas são: o ML-CSK (*Maximum Likelihood Chaos Shift Keying* - Chaveamento caótico com máxima verossimilhança) utilizando dois mapas, o ML-CSK utilizando apenas um mapa e o ML-DCSK (*Maximum Likelihood Differential Chaos Shift Keying* - Chaveamento caótico diferencial com máxima verossimilhança). Os dois primeiros baseiam-se nos propostos em (KISEL; DEDIEU; SCHIMMING, 2001), nos quais empregou-se o algoritmo de Viterbi com as modificações discutidas no Capítulo 5, que permitem usar mapas com densidade invariante não-uniforme. O terceiro é uma contribuição original deste trabalho.

Simulações computacionais comparam o desempenho em canal AWGN desses sistemas com os correspondentes descritos no Capítulo 4.

Nas Seções 6.1 e 6.2 apresentam-se os sistemas ML-CSK com um e dois mapas e o ML-DCSK. São discutidas na Seção 6.3 os resultados das simulações com os três novos sistemas e eles são comparados com os do Capítulo 4. Finalmente, as conclusões do capítulo são apresentadas na Seção 6.4.

6.1 O ML-CSK modificado

O ML-CSK na versão binária ($M = 2$), foi apresentado originalmente por (KISEL; DEDIEU; SCHIMMING, 2001). Sua implementação é feita utilizando uma ou duas funções de base caóticas.

6.1.1 O ML-CSK com duas funções de base

Nesse caso, o transmissor é igual ao do CSK com duas funções de base (Seção 4.2.3). Assim, o sinal transmitido pelo ML-CSK é dado por

$$x_m(n) = x_{m1}s_1(n) + x_{m2}s_2(n), m = 1, 2. \quad (6.1)$$

sendo $\mathbf{x}_1 = [\sqrt{E_b}, 0]^T$ e $\mathbf{x}_2 = [0, \sqrt{E_b}]^T$.

Os mapas geradores dos sinais $s_1(n)$ e $s_2(n)$, respectivamente $f_1(\cdot)$ e $f_2(\cdot)$, devem ser escolhidos de maneira que suas matrizes de transição de estados, \mathbf{A}_1 e \mathbf{A}_2 , sejam distintas. Ao se estimar $s_1(n)$ utilizando o algoritmo de Viterbi com a matriz \mathbf{A}_2 o ganho de estimação deve ser baixo ou mesmo negativo. O mesmo deve ocorrer ao se estimar $s_2(n)$ usando \mathbf{A}_1 .

O receptor para o ML-CSK com duas funções de base é mostrado na Figura 6.1.

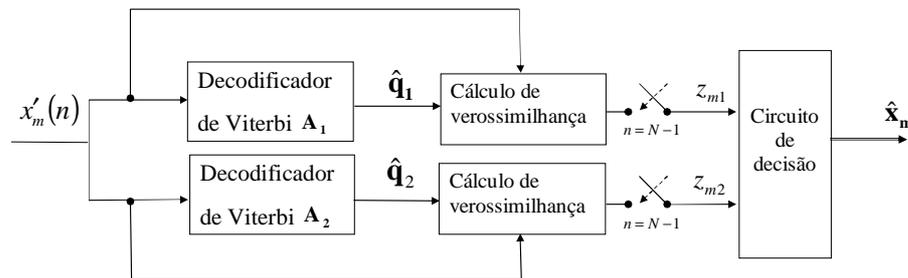


Figura 6.1: Diagrama de blocos do receptor ML-CSK com duas funções de base.

Os decodificadores de Viterbi no receptor tentam recuperar o sinal original $x_m(n)$ a partir das matrizes de transição de estados: ou \mathbf{A}_1 ou \mathbf{A}_2 .

A cada símbolo são obtidas as seqüências de estados estimadas, $\hat{\mathbf{q}}_1$ e $\hat{\mathbf{q}}_2$. Dadas as observações $x'(n)$, as variáveis de observação z_{m1} e z_{m2} são proporcionais à probabilidade de se

obter $\hat{\mathbf{q}}_1$ e $\hat{\mathbf{q}}_2$ respectivamente. Mais precisamente,

$$z_{m1} = \prod_{n=1}^{N-1} P(\hat{q}_1(n)|\hat{q}_1(n-1), \mathbf{A}_1)p(x'(n)|\hat{q}_1(n)), \quad (6.2)$$

$$z_{m2} = \prod_{n=1}^{N-1} P(\hat{q}_2(n)|\hat{q}_2(n-1), \mathbf{A}_2)p(x'(n)|\hat{q}_2(n)). \quad (6.3)$$

Nessas equações utilizou-se a medida de verossimilhança expressa na Eq. (5.62) da página 144. A probabilidade $P(\hat{q}(n)|\hat{q}(n-1), \mathbf{A}_i)$ é lida diretamente da matriz de transição de estados \mathbf{A}_i e $p(x'(n)|\hat{q}(n))$ depende apenas da função densidade de probabilidade do ruído e pode ser aproximada como descrito por Dedieu e Kisel (1999).

A decisão é feita por comparação entre as duas medidas de verossimilhança, z_{m1} e z_{m2} . De acordo com a maior decide-se qual mapa foi utilizado no transmissor, ou seja, se $\hat{\mathbf{x}}_m = \mathbf{x}_1$ ou $\hat{\mathbf{x}}_m = \mathbf{x}_2$, com maior probabilidade.

A dificuldade de projeto que aparece é como obter o mapa $f_2(\cdot)$ com matriz de transição de estados \mathbf{A}_2 que permita discriminar as medidas de verossimilhança das Eqs. (6.2) e (6.3).

Como primeiro enfoque, para mapas lineares por partes no intervalo $U = [-1, 1]$, pode-se utilizar a seguinte regra, adaptada de (KISEL; DEDIEU; SCHIMMING, 2001):

$$f_2(s) = \begin{cases} f_1(s) + 1, & f_1(s) < 0 \\ f_1(s) - 1, & f_1(s) \geq 0 \end{cases} \quad (6.4)$$

A Figura 6.2 mostra a construção do mapa $f_2(\cdot)$ para o mapa $f_1(\cdot) = f_T(\cdot)$. Da forma proposta, $f_1(s)$ e $f_2(s)$ mapeiam s em pontos distantes de uma unidade.

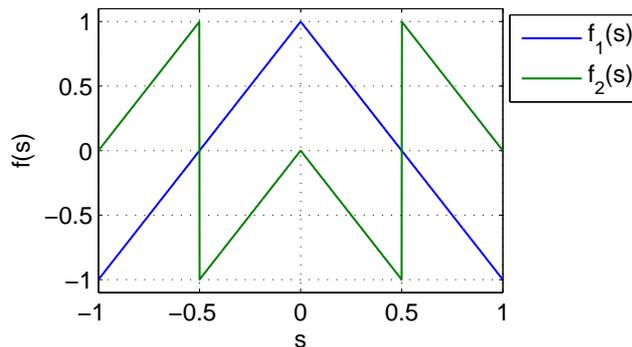


Figura 6.2: Construção do mapa $f_2(\cdot)$ para $f_1(\cdot) = f_T(\cdot)$.

Nesse caso, utilizando partição uniforme e o procedimento detalhado na Seção 5.3.3, chega-se às seguintes matrizes para $N_S = 5$:

$$\mathbf{A}_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}. \quad (6.5)$$

As órbitas geradas por esse $f_2(\cdot)$ têm número de Lyapunov $L = 2$, já que a derivada em cada ponto continua tendo módulo constante e igual a 2. Simulações mostram que elas também são aperiódicas garantindo-se assim que sejam caóticas.

Note-se porém que esse método não é necessariamente ótimo e deve ser aplicado com cautela. Não há nenhuma garantia de que as órbitas do mapa $f_2(\cdot)$, gerado dessa forma, têm comportamento caótico no caso geral.

Por exemplo, caso fosse aplicada a mesma estratégia para o mapa quadrático obter-se-ia o mapa $f_2(\cdot)$ mostrado na Figura 6.3, cujas órbitas convergem para um ponto fixo estável superatrator (DEVANEY, 2003) em $s = 0$, que atrai todas as órbitas no intervalo $[-1, 1]$, ou seja, não seriam produzidas órbitas caóticas.

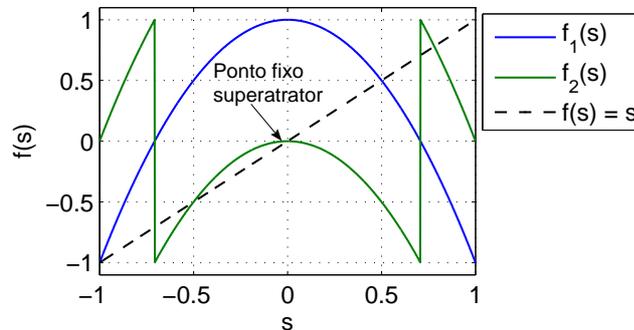


Figura 6.3: Tentativa de construção do mapa $f_2(\cdot)$ para $f_1(\cdot) = f_Q(\cdot)$ usando a regra da Eq. (6.4). Em destaque o ponto fixo superatrator que aparece nesse caso.

Para efeitos de simulação computacional, utilizou-se como base o mapa $f_2(\cdot) = -f_Q(\cdot)$ mostrado na Figura 6.4. Esse mapa possivelmente não é ótimo, já que os pontos próximos das raízes de $f_1(\cdot)$ e $f_2(\cdot)$ são mapeados próximos pelas duas funções. As matrizes de transição de

estados para esses dois mapas para $N_S = 5$ utilizando a partição (5.75) da página 148 são:

$$\mathbf{A}_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \quad (6.6)$$

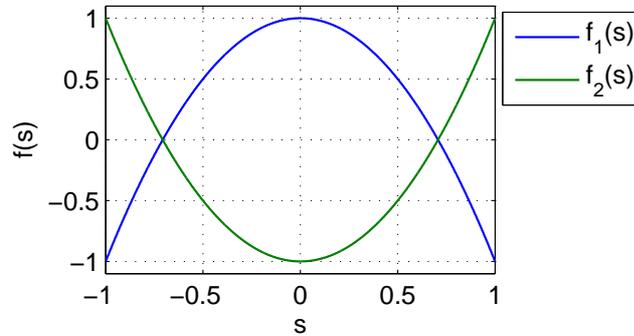


Figura 6.4: Construção do mapa $f_2(\cdot)$ utilizado nas simulações para $f_1(\cdot) = f_Q(\cdot)$.

Com essa escolha de $f_2(\cdot)$, as órbitas resultam aperiódicas e encontra-se um número de Lyapunov igual a 2 para suas órbitas, garantindo sua condição caótica. Porém, repara-se que nas posições a_{23} e a_{43} existem probabilidades de transição não-nulas simultaneamente em ambas as matrizes o que gera erros na detecção do ML-CSK. Como isto não acontece em nenhuma das posições das matrizes da Eq. (6.5), espera-se que o par de mapas quadráticos resulte num sistema com pior desempenho.

Encontrar o mapa $f_2(\cdot)$, dado o mapa $f_1(\cdot)$, que apresenta propriedades ótimas de separação entre as matrizes \mathbf{A}_1 e \mathbf{A}_2 é um problema em aberto. Como mostra o exemplo anterior, é necessário impor ainda que o mapa $f_2(\cdot)$ gere órbitas caóticas.

A Figura 6.5 mostra os exemplos de sinal transmitido $x(n)$ pelo ML-CSK utilizando (a) o mapa tenda e (b) o mapa quadrático. Essa figura pode ser comparada com as Figuras 4.4, 4.13 e 4.22 que mostram os sinais equivalentes para o CSK, o DCSK e o FM-DCSK respectivamente.

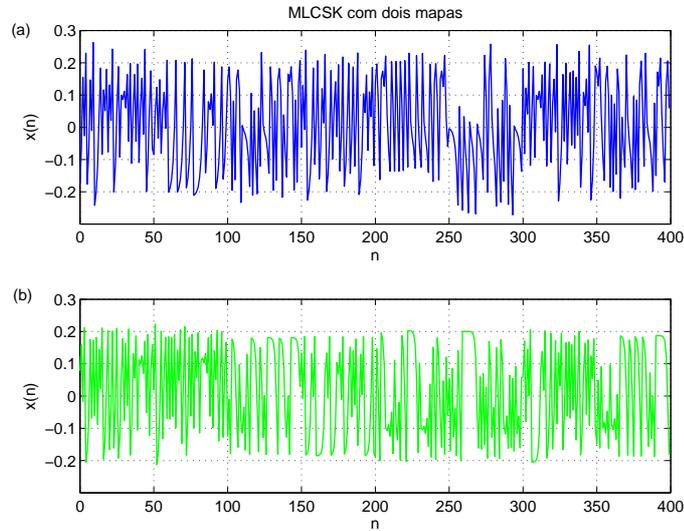


Figura 6.5: Sinais ML-CSK com duas funções de base para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$: (a) mapa tenda; (b) mapa quadrático. Em ambos os casos $N = 50$ e $E_b = 1$. Os símbolos “1” e “0” são transmitidos por $x_1(n)$ e $x_2(n)$ respectivamente.

6.1.2 O ML-CSK com uma função de base

Também é possível construir um sistema de modulação com recepção baseada na estimação pelo algoritmo de Viterbi utilizando apenas uma função de base. Nesse caso, de acordo com o símbolo que se deseja enviar, transmite-se o sinal caótico diretamente ou aplica-se uma transformação invertível a essa seqüência. Essa transformação deve fazer com que a seqüência transformada não seja mais um trecho de órbita válido do mapa utilizado. Dessa forma, evita-se o problema de encontrar um mapa $f_2(\cdot)$, como no caso de duas seqüências de base.

Para o caso binário, para mapas que não sejam ímpares, essa transformação pode ser, por exemplo, $T(s) = -s$ que pode ser desfeita multiplicando-se novamente a seqüência por -1 . Nesse caso, o transmissor ML-CSK com um mapa é igual ao do CSK bipolar com uma função de base, ou seja,

$$x_m(n) = x_{m1}s_1(n), m = 1, 2. \quad (6.7)$$

sendo $x_{11} = \sqrt{E_b}$ e $x_{12} = -\sqrt{E_b}$.

O demodulador para esse sistema é mostrado na Figura 6.6. As duas variáveis de observação z_{m1} e z_{m2} são calculadas conforme a Eq. (6.2). Porém, no cálculo de z_{m2} , $x'_m(n)$ é substituído por $-x'_m(n)$.

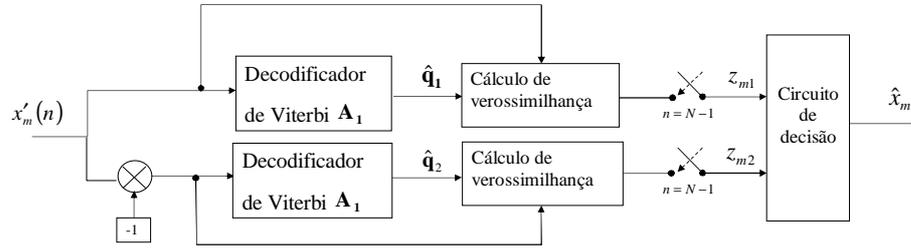


Figura 6.6: Diagrama de blocos do receptor ML-CSK com uma função de base.

Se $x_1(n)$ é transmitido, não há a multiplicação por -1 no transmissor e a verossimilhança expressa por z_{m1} deve ser maior do que z_{m2} já que $-x_1(n)$ não é uma órbita do mapa $f_1(\cdot)$. Já se $x_2(n)$ é transmitido, $-x_2(n)$ é uma órbita do mapa $f_1(\cdot)$ e z_{m2} deve resultar maior que z_{m1} .

Esse esquema de modulação-demodulação é mais simples do que o anterior, já que a decodificação se baseia apenas em uma matriz de transição de estados.

É interessante ressaltar que esse esquema pode ser generalizado para uma modulação M -ária, $M > 2$. Nesse caso, basta considerar outras transformações invertíveis e ainda é necessário apenas uma matriz de transição de estados.

Note-se que a transformação apresentada é apenas ilustrativa. A escolha da transformação invertível não constitui objeto deste trabalho.

A Figura 6.7 mostra exemplos de sinais modulados utilizando o ML-CSK com um mapa para $f_1(\cdot) = f_T(\cdot)$ e $f_1(\cdot) = f_Q(\cdot)$.

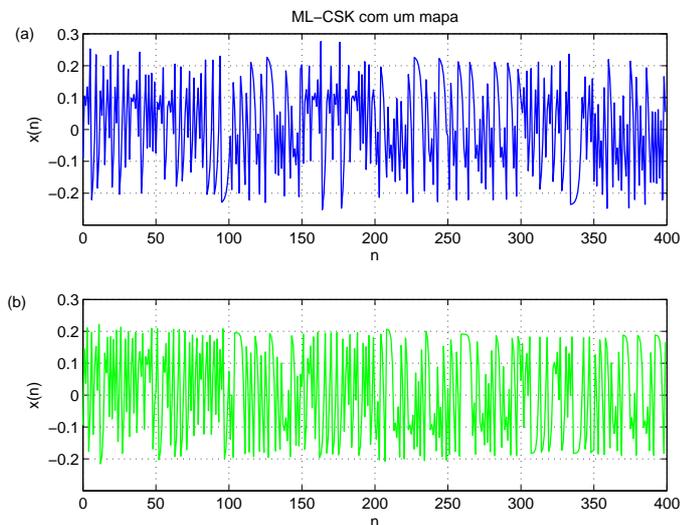


Figura 6.7: Sinais ML-CSK com uma função de base para a seqüência $\{1, 1, 0, 1, 0, 0, 1, 0\}$: (a) mapa tenda; (b) mapa quadrático. Em ambos os casos $N = 50$ e $E_b = 1$. Os símbolos “1” e “0” são transmitidos por $x_1(n)$ e $x_2(n)$ respectivamente.

6.2 O ML-DCSK

Esse sistema é uma modificação do DCSK descrito no Capítulo 4. O que os distingue é a introdução de um estimador de Viterbi no receptor, de forma a melhorar a relação sinal-ruído do trecho de referência do sinal DCSK, que é sempre o mesmo independente do símbolo transmitido.

Assim, o transmissor do ML-DCSK é o mesmo do DCSK apresentado na Figura 4.12 da página 91. Seu receptor é mostrado na Figura 6.8.

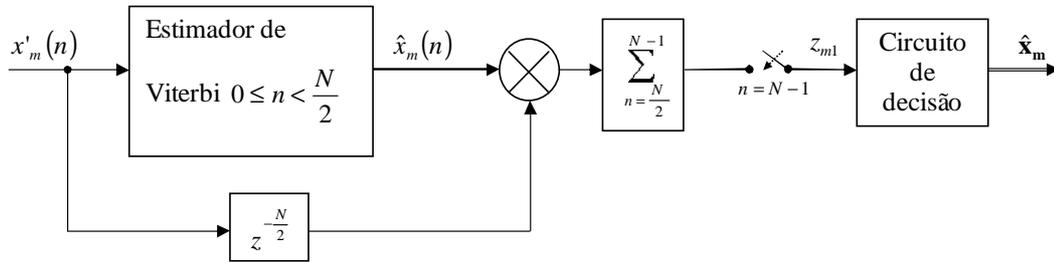


Figura 6.8: Diagrama de blocos do receptor ML-DCSK. O estimador de Viterbi fornece uma estimativa dos primeiros $N/2$ pontos de $x'_m(n)$.

A idéia básica do ML-DCSK é obter um trecho de referência mais próximo do que foi transmitido. Se o número de subintervalos utilizados N_S for suficientemente grande, $\hat{x}_m(n) \approx x_m(n)$ para $0 \leq n < N/2$ e, nesse caso, usando a notação adotada até aqui,

$$z_{m1} = \sum_{n=\frac{N}{2}}^{N-1} \hat{x}_m \left(n - \frac{N}{2} \right) [x_m(n) + r(n)] \approx (-1)^{m+1} E_b \sum_{n=\frac{N}{2}}^{N-1} s^2(n) + \sqrt{E_b} \sum_{n=\frac{N}{2}}^{N-1} s \left(n - \frac{N}{2} \right) r(n). \quad (6.8)$$

Comparando a expressão (6.8) com a equivalente para o DCSK (Eq. (4.47) da página 105), repetida aqui por conveniência:

$$z_{m1} = (-1)^{m+1} E_b \sum_{n=\frac{N}{2}}^{N-1} s^2(n) + \sqrt{E_b} \sum_{n=\frac{N}{2}}^{N-1} s \left(n - \frac{N}{2} \right) r(n) + (-1)^{m+1} \sqrt{E_b} \sum_{n=\frac{N}{2}}^{N-1} s(n) r \left(n - \frac{N}{2} \right) + \sum_{n=\frac{N}{2}}^{N-1} r(n) r \left(n - \frac{N}{2} \right), \quad (6.9)$$

nota-se a ausência das duas últimas parcelas o que deve reduzir a variância de z_{m1} desde que

$\hat{x}_m(n) \approx x_m(n)$ para $0 \leq n < N/2$.

A utilização da estimação de Viterbi também no ramo com o atraso foi testada. Nesse caso, são necessários três estimadores: um estimando o trecho de referência e dois para o trecho de informação, que tenta estimar $x_1(n)$ e $x_2(n)$ no intervalo $N/2 \leq n < N$. Assim, as variáveis de observação são as correlações entre a estimativa do trecho de referência e cada uma das estimativas do trecho de observação. A correlação que resulta maior indica o símbolo transmitido. Porém, a idéia foi descartada por não apresentar ganhos consideráveis em relação ao ML-DCSK proposto, além de aumentar muito a complexidade do receptor.

6.3 Curvas de desempenho em canal AWGN

Nessa seção são mostrados alguns resultados de simulações computacionais dos sistemas ML-CSK descritos, buscando avaliar seu desempenho em termos de SER em canais AWGN. Utilizam-se os mesmos procedimentos e métodos do Capítulo 4 a fim de facilitar comparações com os sistemas lá descritos e com os sistemas de modulação convencionais.

6.3.1 Curvas de desempenho do ML-CSK modificado

Na Figura 6.9 são mostradas curvas de SER em função de E_b/N_0 para o ML-CSK utilizando dois mapas. Nesse caso, é utilizado o mapa tenda como $f_1(\cdot)$ e $N_S = 100$. Para comparação, são mostradas as taxas de erro de símbolo do sistema COOK em tracejado. Verifica-se que as curvas para o ML-CSK dependem do número de pontos utilizados por símbolo N , pois a qualidade da estimativa do algoritmo de Viterbi não depende de N para uma mesma SNR_{in} , como se percebe pela Figura 5.17 da página 151. Como $E_b/N_0 = \text{SNR}_{in} \cdot N$, quanto menor N , maior a SNR_{in} para uma mesma E_b/N_0 , tornando melhores a qualidade da estimação do desempenho do ML-CSK. Em contrapartida, quanto menor N , mais difícil fica a discriminação entre os dois mapas, o que tende a piorar o desempenho do ML-CSK. A junção desses dois efeitos causa um comportamento complicado das taxas de erro com N .

O número de subintervalos escolhidos para o algoritmo de Viterbi, $N_S = 100$ garante que a partição não afeta de forma significativa o desempenho do sistema na faixa de E_b/N_0

considerada.

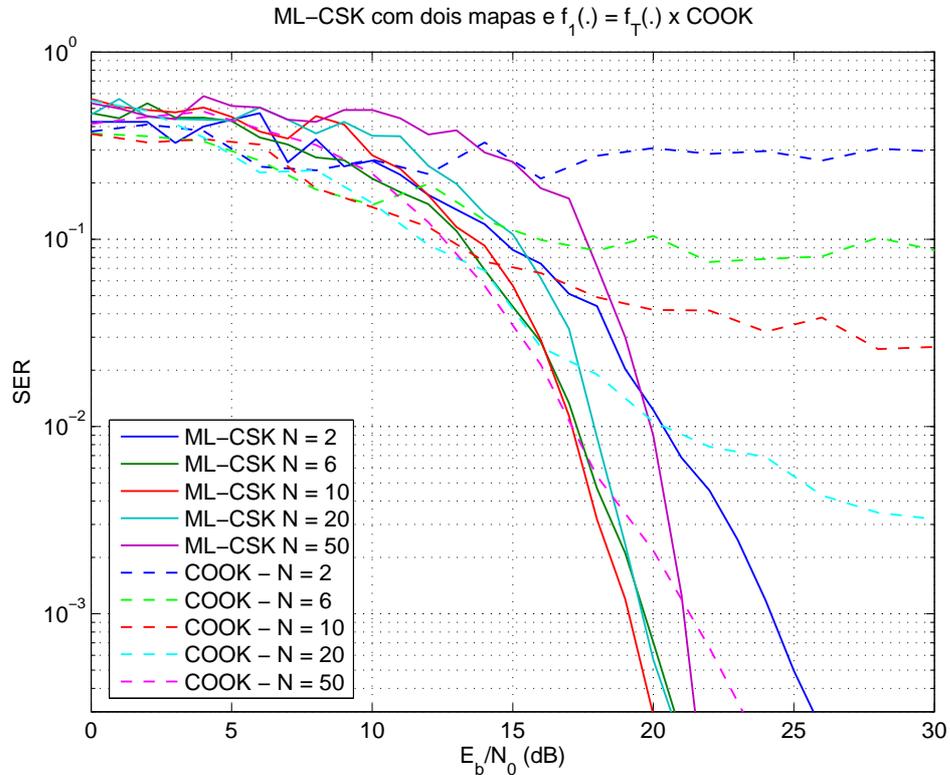


Figura 6.9: Taxas de erro de símbolo do ML-CSK com duas funções de base utilizando $f_1(\cdot) = f_T(\cdot)$.

Compara-se o desempenho do ML-CSK com dois mapas com o do sistema COOK, que tem o melhor desempenho entre os sistemas com recepção baseada em correlação não-coerente, conforme a Seção 4.5.1. Claramente, o ML-CSK oferece melhores resultados ao custo de maior complexidade no receptor. Além disso, o ML-CSK realmente emprega a informação para alterar as características caóticas do sinal e não sua energia.

Ao contrário do COOK, o ML-CSK depende fortemente das características dos mapas utilizados na modulação, através das matrizes de transição de estados, que devem ser conhecidas tanto pelo transmissor quanto pelo receptor para permitir a demodulação. Assim, pode-se supor uma maior dificuldade de detecção não-autorizada em relação ao COOK.

A Figura 6.10 mostra o desempenho obtido ao se utilizar $f_Q(\cdot)$ como $f_1(\cdot)$ mantidas as demais condições. Apesar dos níveis de SER ainda estarem abaixo dos do COOK para altas E_b/N_0 , seu desempenho é levemente inferior ao caso do mapa tenda. Com o mapa $f_Q(\cdot)$, as matrizes \mathbf{A}_1 e \mathbf{A}_2 (Eq. (6.6)) apresentam simultaneamente posições não-nulas, o que aumenta

a probabilidade de erro na decodificação. Esses resultados ilustram a importância da escolha judiciosa dos mapas para o funcionamento adequado desse sistema.

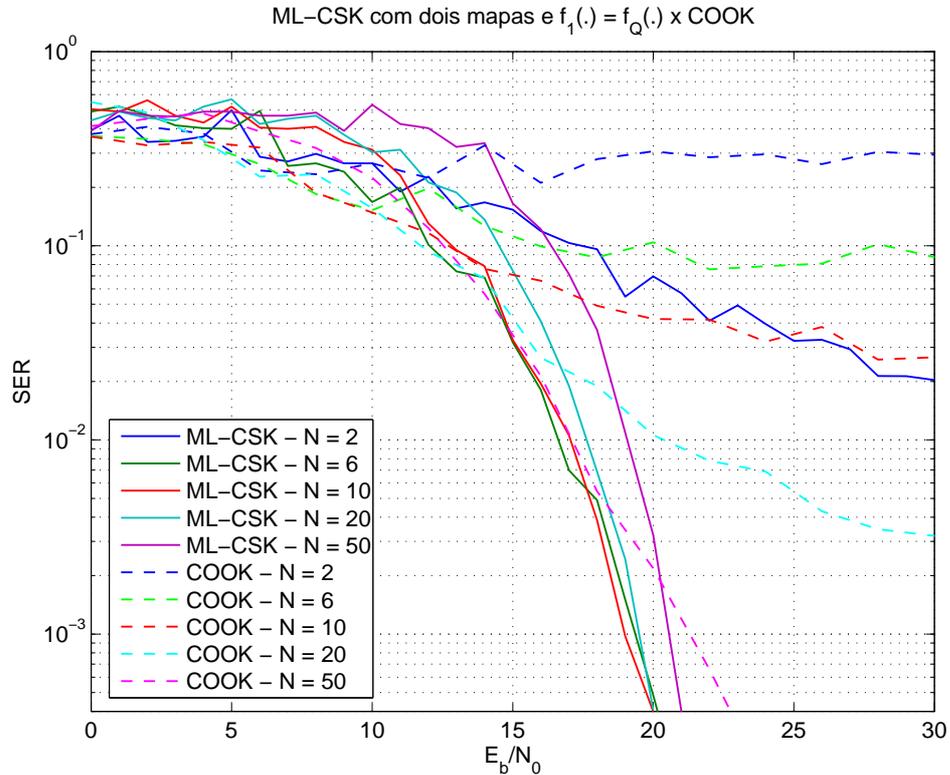


Figura 6.10: Taxas de erro de símbolo do ML-CSK com duas funções de base utilizando $f_1(\cdot) = f_Q(\cdot)$.

As Figuras 6.11 e 6.12 mostram curvas de desempenho para o ML-CSK utilizando um mapa e a transformação $T(s) = -s$. Na Figura 6.11, é utilizado o mapa tenda, $f_1(\cdot) = f_T(\cdot)$ e na Figura 6.12 é utilizado o mapa $f_1(\cdot) = f_Q(\cdot)$. Novamente, o COOK é utilizado como parâmetro de comparação.

Considerada no seu conjunto, as simulações mostram um desempenho levemente superior dos sistemas com um único mapa quando comparados aos com dois mapas. Note-se que os receptores dos sistemas com um único mapa são mais simples pois necessitam armazenar apenas uma matriz.

O melhor resultado do mapa $f_T(\cdot)$ em relação ao mapa $f_Q(\cdot)$ novamente confirma a importância da escolha do mapa e da transformação a ser empregada. Encontrar mapas e transformações que otimizem o desempenho desses sistemas é um problema complicado ainda em aberto.

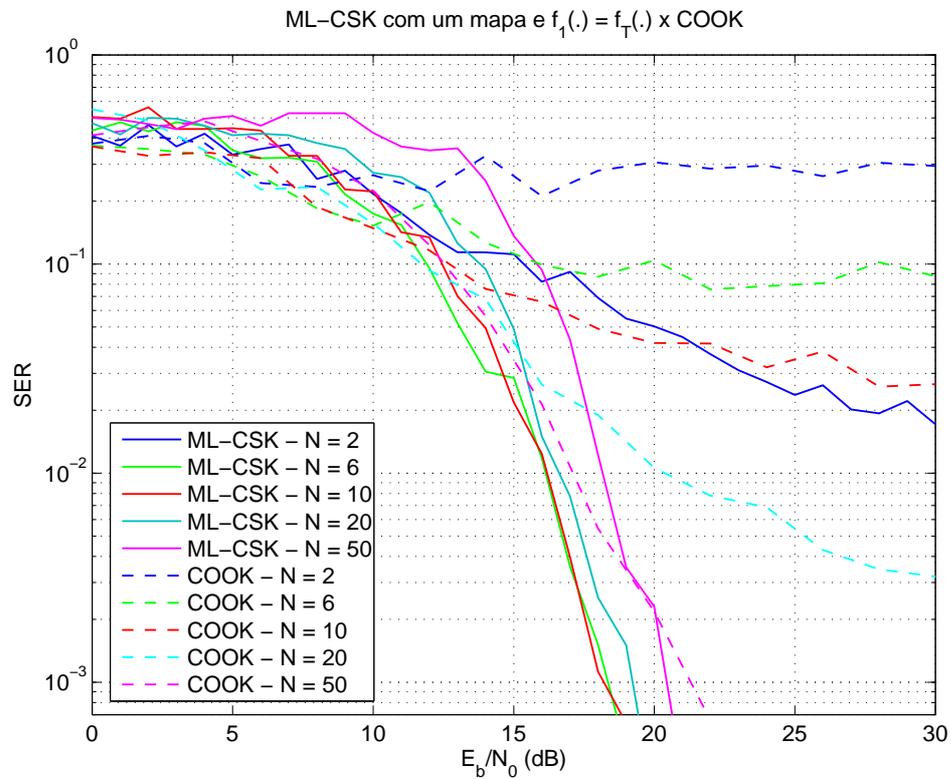


Figura 6.11: Taxas de erro de símbolo do ML-CSK com uma função de base utilizando $f_1(\cdot) = f_T(\cdot)$.

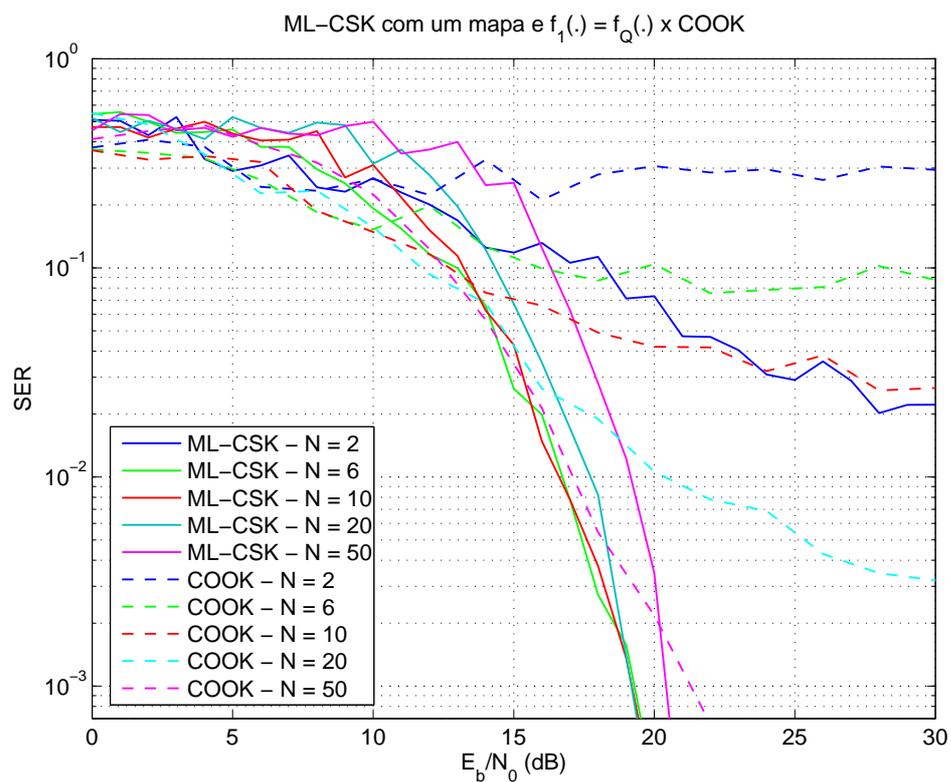


Figura 6.12: Taxas de erro de símbolo do ML-CSK com uma função de base utilizando $f_1(\cdot) = f_Q(\cdot)$.

A Figura 6.13 permite comparar a SER obtida com o ML-CSK com um ou dois mapas e com $f_1(\cdot) = f_T(\cdot)$ ou $f_1(\cdot) = f_Q(\cdot)$ para $N = 10$. Entre todos, o ML-CSK com um mapa e $f_1(\cdot) = f_T(\cdot)$ é o que apresenta melhor desempenho.

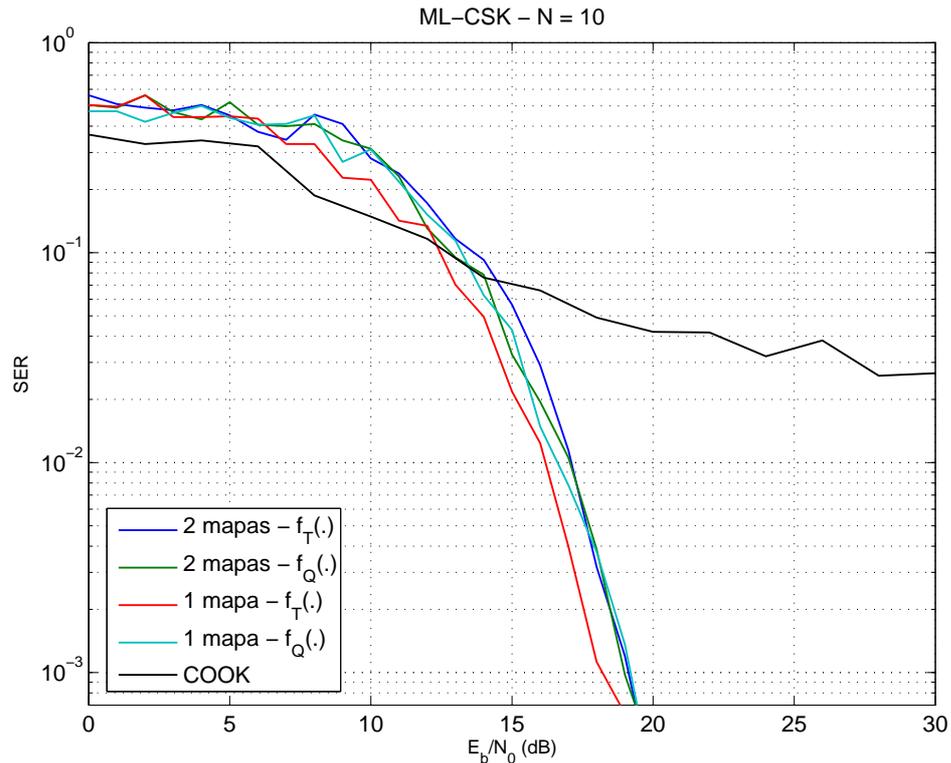


Figura 6.13: Taxas de erro de símbolo para os sistemas ML-CSK testados ($N = 10$).

6.3.2 Curvas de desempenho do ML-DCSK

Estão mostradas na Figura 6.14 curvas de erro de símbolo para o ML-DCSK. Foi utilizado o mapa tenda $f_T(\cdot)$ e $N_S = 100$. As curvas do sistema DCSK também são mostradas em tracejado para facilitar a comparação.

Pela análise da taxa de erro de símbolo, verifica-se que fixado uma relação E_b/N_0 , a SER do ML-DCSK é sempre maior do que a do DCSK. Os resultados mostram que a variância de z_{m1} do ML-DCSK em relação ao do DCSK sem estimação de máxima verossimilhança não decaiu e a aproximação da expressão (6.8) não se verificou nas simulações.

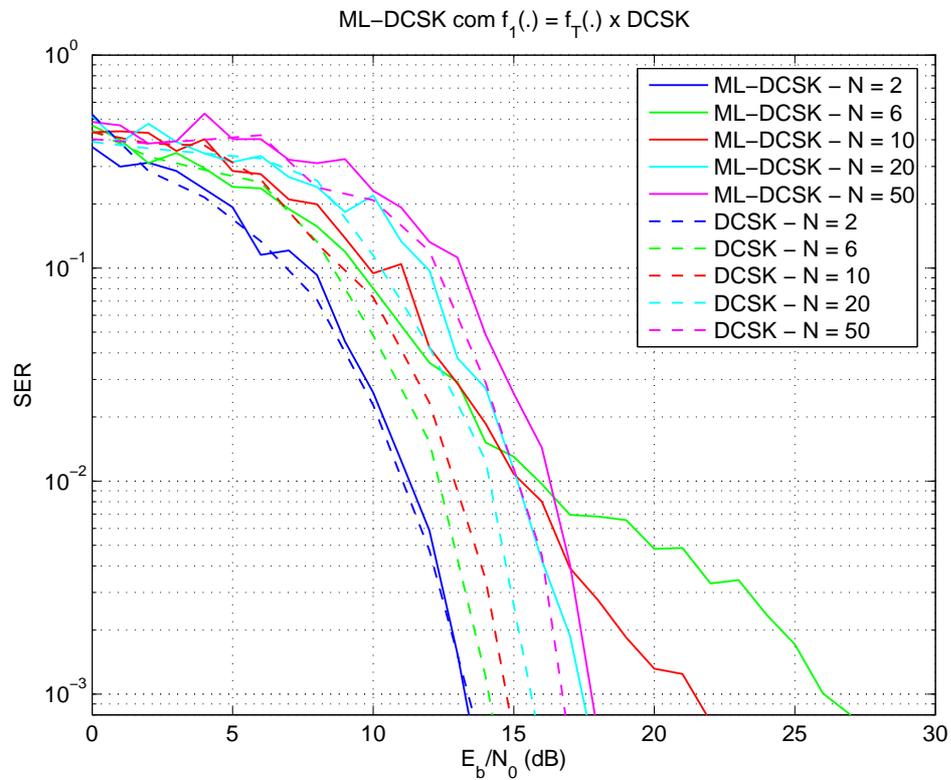


Figura 6.14: Taxas de erro de símbolo do ML-DCSK utilizando o mapa $f_T(.)$ como gerador e $N_S = 100$.

6.4 Conclusões

O capítulo descreveu três sistemas que empregam estimação pelo algoritmo de Viterbi na tentativa de aperfeiçoar o desempenho dos sistemas descritos no Capítulo 4.

Como já foi visto, o COOK é o sistema baseado em recepção não-coerente com melhor desempenho (página 102). Comparado com ele, o ML-CSK com um ou dois mapas apresentou desempenho melhor em canal AWGN. O custo dessa melhora sem dúvida é a maior complexidade do receptor. Em contrapartida, a informação é codificada na dinâmica do sistema caótico e não em propriedades facilmente mensuráveis, como a energia, tornando assim mais difícil a detecção não-autorizada. É necessário o conhecimento das matrizes de transição de estados no receptor para realizar a demodulação.

São necessárias mais pesquisas para se encontrar mapas com matrizes de transição de estados que tornem, no caso do ML-CSK com dois mapas, as medidas de verossimilhanças das Eqs. (6.2-6.3) mais discriminantes. Um caminho viável poderá ser o uso de algum método de

otimização. No caso com um único mapa, a transformação invertível utilizada também deve ser melhor estudada.

Já, no caso da recepção diferencial, o ML-DCSK apresentou resultados mais modestos do que o DCSK, mostrando ser necessário mais estudo para aplicação eficiente dos métodos de estimação de órbitas caóticas, melhorando o desempenho desse tipo de recepção.

Capítulo 7

Conclusões e trabalhos futuros

Desde a publicação de alguns trabalhos pioneiros como, por exemplo, (YAMADA; FUJISAKA, 1983; PECORA; CARROLL, 1990; CUOMO; OPPENHEIM, 1993b; HAYES; GREBOGI; OTT, 1993), surgiu um interesse crescente na utilização de sinais caóticos em sistemas de modulação.

Muitos dos primeiros trabalhos tinham como base a possibilidade de sincronização entre sistemas caóticos com condições iniciais arbitrárias, demonstrada em (YAMADA; FUJISAKA, 1983; PECORA; CARROLL, 1990). Porém, apesar de funcionarem em condições ideais, esses sistemas mostraram-se bastante vulneráveis quando submetidos a ruído aditivo ou quando havia diferenças de parâmetros entre transmissor e receptor (EISENCRAFT; GERKEN, 2000; EISENCRAFT, 2001; WILLIAMS, 2001).

Mesmo com avanços nas técnicas de sincronização caótica nos últimos anos (MILLERIOUX; DAAFOUZ, 2001; BOUTAYEB; DAROUACH; RAFARALAH, 2002), boa parte dos trabalhos na área da Engenharia de Telecomunicações que tratavam de modulação utilizando sinais caóticos passaram a considerar com maior ênfase os sistemas de modulação digital não-coerentes e diferenciais, ou seja, aqueles que não necessitam da recuperação das seqüências de base para realizar a demodulação. Diversas modulações digitais foram propostas nessa linha, destacando-se o CSK e o DCSK e suas variações (KOLUMBÁN; KENNEDY; CHUA, 1998). Porém, na maior parte desses sistemas a informação não é codificada na dinâmica caótica e sim em propriedades mais simples dos sinais transmitidos, como sua energia ou a correlação

entre seus trechos. Informações sobre a dinâmica dos mapas geradores de sinais caóticos não são utilizadas no processo de demodulação.

Nesta tese, foram abordadas técnicas de estimação de sinais caóticos gerados por sistemas unidimensionais de tempo discreto e de suas condições iniciais, visando aplicá-las para melhorar as características dos sistemas CSK e DCSK. Neste sentido: determinou-se limites de desempenho na estimação de condições iniciais de órbitas caóticas em função de propriedades estatísticas do seu atrator; foi obtida uma generalização do algoritmo de Viterbi proposto por Dedieu e Kisel (1999) de forma que ele pudesse ser aplicado a mapas mais genéricos e não apenas a mapas com densidade invariante uniforme; e estudaram-se sistemas que utilizam este algoritmo na demodulação. Além disso, estendeu-se a notação equivalente passa-baixas de tempo discreto, usual em sistemas com portadoras senoidais, para sistemas com portadoras caóticas.

Contribuições e conclusões do trabalho são apresentadas a seguir juntamente com propostas de trabalhos futuros.

7.1 Contribuições

No Capítulo 1 discutiram-se os problemas que limitam a aplicação de modulações caóticas em sistemas de comunicações práticos:

- I. determinação teórica da largura de banda dos sinais caóticos;
- II. sincronização caótica;
- III. avaliação em condições não ideais de canal;
- IV. integração com a tecnologia existente;
- V. adequação e uniformidade de notação;
- VI. escolha dos mapas geradores a serem utilizados;
- VII. exploração mais intensa das propriedades da dinâmica caótica em receptores não-coerentes e diferenciais.

Desses, nesta tese foram estudados aspectos relacionados aos problemas (V), (VI) e (VII) que são respectivamente a adequação e uniformidade de notação, a escolha dos mapas geradores a serem utilizados e a exploração mais intensa das propriedades da dinâmica caótica em receptores não-coerentes e diferenciais.

As principais contribuições dessa tese encontram-se esquematizadas na Tabela 7.1 e são descritas de forma resumida em seguida.

Tabela 7.1: Principais contribuições da tese

Capítulo	Contribuições
4	A. Notação de equivalente passa-baixas de tempo discreto para sistemas de modulação digital com portadoras caóticas
5	B. CRLB na estimação de condições iniciais de órbitas caóticas em função do número de Lyapunov (Teoremas 6 e 7) C. Generalização do algoritmo de Viterbi para estimação de órbitas caóticas de mapas com densidade invariante não-uniforme D. Comparações entre MLE e algoritmo de Viterbi na estimação de órbitas
6	E. Avaliação do desempenho do ML-CSK utilizando mapas com densidade invariante não-uniforme F. ML-DCSK

A. Notação de equivalente passa-baixas de tempo discreto para sistemas de modulação digital com portadoras caóticas

Tendo-se em mente o problema (V), utilizou-se uma notação comum para sistemas de modulação digital usando portadoras senoidais ou caóticas fazendo análises exclusivamente em tempo discreto, que são mais convenientes para se trabalhar com mapas geradores de sinais caóticos. A inspiração para a notação unificada são os artigos (KOLUMBÁN; KENNEDY; CHUA, 1997, 1998) que utilizam tempo contínuo.

B. CRLB na estimação de condições iniciais de órbitas caóticas em função do número de Lyapunov (Teoremas 6 e 7)

Para mapas unidimensionais quaisquer, deduziram-se limites de desempenho, mais precisamente, o limite inferior de Cramér-Rao (CRLB), na estimação de condições iniciais de órbitas. Mostrou-se que, quando o número de pontos utilizados na estimação tende a infinito, esse limite depende apenas do número de Lyapunov do atrator da órbita. Esse resultado pode ser interessante à luz do problema (VI).

C. Generalização do algoritmo de Viterbi para estimação de órbitas caóticas de mapas com densidade invariante não-uniforme

Considerando os problemas (VI) e (VII), propôs-se uma generalização do algoritmo de estimação de Viterbi para órbitas caóticas de Dedieu e Kisel (1999). Essa extensão tornou o novo algoritmo aplicável a mapas com densidade invariante não-uniforme desde que se conheça uma conjugação de tal mapa com um mapa com densidade invariante uniforme. Utiliza-se um processo análogo ao filtro branqueador que aparece em diversos problemas como os de predição linear (HAYKIN, 1996), de representação espectral (PAPOULIS; PILLAI, 2002), de modelamento de canais com interferência intersimbólica (PROAKIS, 1995), entre outros.

D. Comparações entre MLE e algoritmo de Viterbi na estimação de órbitas

Tendo em vista o problema (VII), compararam-se os desempenhos de duas formas de estimação para órbitas caóticas: o estimador de máxima verossimilhança proposto em (PAPADOPOULOS; WORNELL, 1993) e o algoritmo de Viterbi citado acima em (C). Além de se aplicar, com a generalização que foi exposta, a uma categoria maior de mapas, o algoritmo de Viterbi apresentou maiores ganhos de estimação numa faixa considerável de relação sinal-ruído desde que se escolhesse um número de intervalos de partição conveniente.

E. Avaliação do desempenho do ML-CSK utilizando mapas com densidade invariante não-uniforme

Com a contribuição (C) descrita acima, pôde-se avaliar também para mapas com densidade invariante não-uniforme, o desempenho dos sistemas ML-CSK com um e dois mapas propostos em (KISEL; DEDIEU; SCHIMMING, 2001). Tais sistemas tiveram desempenho melhor do que seus equivalentes que utilizam correlação não-coerente. Porém, mais pesquisas são necessárias para escolher mapas e transformações que tornem o desempenho desses sistemas ótimos.

F. O sistema ML-DCSK

Ainda com relação ao problema (VII) propôs-se o sistema DCSK com estimação de máxima verossimilhança (ML-DCSK - *Maximum Likelihood DCSK*), um sistema baseado em recepção diferencial semelhante ao DCSK, mas que utiliza um estimador de Viterbi para melhorar a relação sinal-ruído do trecho de referência. Os resultados das simulações indicaram a necessidade de mais estudo para melhorar o desempenho do DCSK.

Como resultado da contribuição (B) foi publicado o artigo

- EISENCRAFT, M.; BACCALÁ, L. A. The impact of the Lyapunov number on the Cramér-Rao lower bound for the estimation of chaotic signals. In: INTERNATIONAL WORKSHOP ON TELECOMMUNICATIONS, 2004, Santa Rita do Sapucaí. *Proceedings of IWT'2004*, p. 176-179.

Além disso, foi submetido um artigo à *1st IFAC Conference on Analysis and Control of Chaotic Systems* (<http://www.univ-reims.fr/Labos/LAM/chaos06/>) à ser realizada de 28 a 30 de junho de 2006 em Reims, França.

A contribuição (D) gerou um resumo submetido à *IX Experimental Chaos Conference* (<http://www.lac.inpe.br/ecc9>) a ser realizada de 29 de maio a 1º de junho de 2006 em São José dos Campos, São Paulo.

7.2 Conclusões

Seguem-se as principais conclusões a que se chegou no desenvolvimento desta tese.

7.2.1 Modulações digitais usando sinais caóticos baseadas em correlação

No Capítulo 4 analisou-se qualitativa e quantitativamente o desempenho em canal AWGN de sistemas de modulação digital utilizando portadoras caóticas e recepção baseada em correlação.

Concluiu-se que:

- Os receptores com correlação coerente, apesar de ótimos quando as funções de base são conhecidas no receptor (LATHI, 1998), apresentam um grave empecilho: dependem da sincronização caótica, que, pelo menos nas formas propostas até hoje, é sensível ao ruído no canal e descasamento de parâmetros entre transmissor e receptor, impossibilitando sua aplicação em canais em que o ruído é uma limitação (EISENCRAFT; GERKEN, 2000; WILLIAMS, 2001). Assim, os estudos aqui concentraram-se em sistemas com recepção não-coerente e diferencial.
- As modulações COOK e CSK unipolar com recepção não-coerente apresentam taxas de erro de símbolos altas comparadas aos sistemas equivalentes convencionais com portadora senoidais (ASK não-coerente). Isto deve-se a dois fatores: a variabilidade da energia dos trechos de sinais caóticos e a ausência de um sinal de referência sem ruído no receptor. Esse último obriga a realização de um produto de sinais somados a ruído tornando o nível de decisão dependente da potência do ruído no canal e a taxa de erro dependente do número de amostras utilizado para representar cada símbolo.
- No DCSK com recepção diferencial, o nível de detecção independe da potência de ruído. Esse sistema apresenta desempenho melhor do que o CSK não-coerente porém ainda inferior, em canal AWGN, ao convencional DPSK binário.
- Entre as modulações caóticas estudadas, o FM-DCSK com recepção diferencial é o que apresenta menor taxa de erro para um valor fixo de E_b/N_0 em canal AWGN. Essa superioridade deve-se a três fatores: consegue-se manter a energia por símbolo constante; a independência do nível de decisão com a potência do ruído; e a desnecessidade da sincronização caótica características da demodulação diferencial. Porém, o desempenho

ainda fica abaixo do DPSK binário convencional devido à ausência de um sinal de referência no receptor.

- Não se deve utilizar o pior desempenho dessas modulações digitais caóticas em relação às convencionais em canal AWGN para descartá-las. Em primeiro lugar porque a característica banda larga dos sinais caóticos pode ser vantajosa em algumas situações, como canais com multipercurso (KENNEDY et al., 2000). Além disso, as recepções não-coerentes e diferenciais estudadas não utilizam informações sobre a dinâmica do mapa. Essas podem ser usadas para diminuir a taxa de erro de símbolo através da estimação dos sinais caóticos no receptor.

7.2.2 Estimação de condições iniciais de sinais caóticos

Com o Teorema 6 (Capítulo 5) pôde-se obter limites de desempenho (o CRLB) para estimadores não-enviesados das condições iniciais de órbitas a partir de um trecho de N pontos corrompido por AWGN. Quando N é suficientemente grande, o Teorema 7 apresentou um limite que depende apenas de N e do número de Lyapunov do atrator da órbita, caótica ou não. A partir destes resultados, concluiu-se que:

- Caso deseje-se estimar apenas a condição inicial s_0 de uma órbita e não a órbita como um todo, é mais interessante, em princípio, usar mapas com número de Lyapunov elevado. Deve-se ressaltar que essa escolha torna mais difícil a reconstrução da órbita inteira a partir de s_0 já que o número de Lyapunov é uma taxa média de separação entre órbitas próximas.
- No caso de órbitas caóticas, a dependência do CRLB com a condição inicial que está sendo estimada diminui com N . Isto se deve à transitividade, ou seja, a propriedade das órbitas caóticas percorrerem as vizinhanças de todos os pontos do seu atrator (DEVANEY, 2003).
- O CRLB cai exponencialmente conforme N aumenta para órbitas caóticas. Assim, a cada novo ponto utilizado na estimação, muita informação é acrescentada. Esse fato é consequência da geração de informação característica de sistemas caóticos (FERRARA;

PRADO, 1994). Para uma órbita assintoticamente periódica, por exemplo, o CRLB tende a uma constante conforme N cresce.

7.2.3 Estimação de sinais caóticos

No Capítulo 5, descreveram-se dois métodos de estimação de órbitas caóticas corrompidas por AWGN: o MLE e o algoritmo de Viterbi. O primeiro foi deduzido apenas para mapas que possuem densidade invariante uniforme. Já o segundo abrange casos mais genéricos desde que se consiga uma conjugação entre o mapa gerador da órbita e um mapa com densidade invariante uniforme. A partir das simulações computacionais descritas naquele capítulo, chegou-se às seguintes conclusões:

- Escolhido o número de intervalos de partição adequado, a estimação pelo método de Viterbi apresenta melhores ganhos de estimação do que o MLE para valores de SNR baixos e moderados, como mostrado no exemplo da Figura 5.23. Além disso, a generalização do método de Viterbi se aplica a uma classe maior de mapas. Por essas razões, ela foi escolhida para a implementação dos sistemas ML-CSK e ML-DCSK no Capítulo 6.
- O método de Viterbi apresenta ganho de estimação positivo para mapas com densidade invariante uniforme quando o número de subintervalos N_S escolhido é maior ou igual a $10^{\text{SNR}_{\text{in}}/20}$, sendo SNR_{in} a relação sinal-ruído em dB do sinal à entrada do estimador.
- Simulações computacionais indicam que, para os mapas testados, o ganho de estimação pelo algoritmo de Viterbi tem uma dependência fraca com o número de Lyapunov da órbita.
- A mudança no algoritmo de Viterbi de Dedieu e Kisel (1999) para que ele possa ser aplicado a mapas com densidade invariante não-uniforme é efetivamente necessária. Esse processo, baseado em conjugação de mapas é análogo ao processo de branqueamento utilizado em problemas de predição linear (HAYKIN, 1996), de representação espectral (PAPOULIS; PILLAI, 2002), de modelamento de canais com interferência intersimbólica (PROAKIS, 1995) entre outros.

7.2.4 Sistemas de modulação envolvendo estimação de sinais caóticos

No Capítulo 6, foram estudados três sistemas de modulação caótica que diferem dos descritos no Capítulo 4 por utilizarem a estimação de máxima verossimilhança de órbitas caóticas. As simulações computacionais realizadas com esses sistemas permitiram as seguintes conclusões:

- Estimação de órbitas no receptor é uma forma de utilizar informações sobre a dinâmica dos mapas geradores dos sinais caóticos e assim obter taxas de erros menores do que as dos receptores não-coerentes. Essa informação expressa-se pelas matrizes de transição de estados.
- Os desempenhos em canal AWGN dos sistemas ML-CSK com um e dois mapas foram comparados ao do sistema COOK, que tem os melhores resultados entre os sistemas baseados em modulação caótica com recepção por correlação não-coerente. De forma geral, os sistemas baseados em estimação foram superiores, com leve vantagem do ML-CSK com um único mapa. Porém, espera-se que estes resultados possam ser melhorados escolhendo-se mapas e transformações invertíveis com os quais se consiga maior discriminação entre as medidas de verossimilhança no receptor.
- O ML-DCSK é baseado no DCSK. Nesse sistema se faz uma estimação do trecho de referência para melhorar a relação sinal-ruído desse trecho; entretanto, ele não apresentou desempenho satisfatório em relação ao DCSK original. Comparado ao FM-DCSK, seu desempenho deixa a desejar. É preciso mais pesquisa para aproveitar melhor os processos de estimação de órbitas em receptores diferenciais.

7.3 Trabalhos Futuros

A pesquisa realizada durante a elaboração desta tese, levou a uma série de possibilidades de trabalhos futuros:

- determinação da largura de banda de frequências utilizada em cada sistema de modulação descrito;

- análise estatística teórica da energia de trechos de sinais caóticos permitindo assim calcular com mais precisão níveis de detecção em receptores não-coerentes;
- generalização dos sistemas propostos para transmissão de M símbolos com $M > 2$;
- comparação entre os sistemas de modulação propostos utilizando modelos de canais mais complexos e realistas;
- estudo de formas de multiplexação envolvendo sinais caóticos viabilizando a utilização multiusuário desses sistemas;
- generalização dos limites de Cramér-Rao obtidos para o caso multidimensional;
- cálculo teórico da matriz de transição de estados de um mapa qualquer;
- análise e comparação da complexidade computacional dos estimadores descritos;
- otimização da escolha dos mapas utilizados no ML-CSK com 2 mapas para garantir máxima discriminação entre os estimadores do receptor;
- otimização da escolha da transformação invertível usada no ML-CSK com 1 mapa de forma a garantir máxima discriminação entre os estimadores do receptor;
- proposta de um sistema de modulação que permita recepção diferencial com uso de estimação de órbitas caóticas com desempenho melhor do que o FM-DCSK.

Referências Bibliográficas

- ABARBANEL, H. D. I. *Analysis of observed chaotic data*, New York: Springer Verlag, 1996.
- ALLIGOOD, K. T.; SAUER, T. D.; YORKE, J. A. *Chaos - an introduction to dynamical systems*, New York: Springer, 1996.
- ARGYRIS, A.; SYVRIDIS, D.; LARGER, L.; ANNOVAZZI-LODI, V.; COLET, P; FISCHER, I.; GARCÍA-OJALVO J.; MIRASSO C. R.; PESQUERA, L.; SHORE, K. A. Chaos-based communications at high bit rates using commercial fiber-optic links. **Nature**, v. 437, p. 343-346, 2005.
- BOLLT, E. M. Review of chaos communication by feedback control of symbolic dynamics. **International Journal of Bifurcation and Chaos**, v. 13, n. 2, p. 269-285, 2003.
- BOUTAYEB, M.; DAROUACH, M.; RAFARALAHY H. Generalized state-space observers for chaotic synchronization and secure communications. **IEEE Transactions on Circuits and Systems-I**, v. 49, n. 3, p. 345-349, 2002.
- CESARI, L. *Asymptotic behavior and stability problems in ordinary differential equations*, Berlin: Springer-Verlag, 1971.
- CUOMO, K. M.; OPPENHEIM, A.V. Circuit implementation of synchronized chaos with applications to communications. **Physical Review Letters**, v. 71, n.1, p. 65-68, 1993a.
- CUOMO, K. M.; OPPENHEIM, A. V. Chaotic signals and systems for communications. In: IEEE CONFERENCE ON ACOUSTIC, SPEECH AND SIGNAL PROCESSING, Minneapolis, 1993. *Proc. ICASSP'93*, v. 3, p. 137-140, 1993b.

- CUOMO, K. M.; OPPENHEIM, A. V.; STROGATZ, S. H. Synchronization of Lorenz-based chaotic circuits with applications to communications. **IEEE Transactions on Circuits and Systems-II**, v. 40, n. 10, p. 626-633, 1997.
- DEDIEU, H; KENNEDY, M. P.; HASLER, M. Chaos shift keying: modulation and demodulation of a chaotic carrier using self-synchronized Chua's circuits. **IEEE Transactions on Circuits and Systems - II**, v. 40, n. 10, p. 634-642, 1993.
- DEDIEU, H.; KISEL, A. Communications with chaotic time series: probabilistic methods for noise reduction. **International Journal of Circuit Theory and Applications**, v. 27, p. 577-587, 1999.
- DEVANEY, R. L. *An introduction to chaotic dynamical systems*, 2nd edition, Boulder: Westview Press, 2003.
- DRAKE, D. F. Information's role in the estimation of chaotic signals. 1998. Thesis (Doctor in Philosophy in Electrical Engineering) - Georgia Institute of Technology, 1998.
- EISENCRAFT, M.; GERKEN, M. Comunicação utilizando sinais caóticos: influência de ruído e limitação em banda. In: SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, 18., Gramado, 2000. 1 CD-ROM.
- EISENCRAFT, M. *Sistemas de comunicação utilizando sinais caóticos*. 2001. 115p. Dissertação (Mestrado em Engenharia) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2001.
- FEIGENBAUM, M. J. Quantitative universality for a class of nonlinear transformations. **Journal of Statistical Physics**, v. 19, p. 25-52, 1978.
- FERRARA, N. F.; PRADO, C. P. C. *Caos - uma introdução*, São Paulo: Edgard Blücher, 1994.
- GALIAS, Z.; MAGGIO, G. M. Quadrature chaos shift keying. In: IEEE SYMPOSIUM ON CIRCUITS AND SYSTEMS, Sidney, 2001. *Proc. ISCAS'01*, v. 3, p. 313-316, 2001.

- GUCKENHEIMER, J.; HOLMES, P. *Nonlinear oscillations, dynamical systems and bifurcation of vector fields*, New York: Springer, 2002.
- HAYES, S.; GREBOGI, C.; OTT, E. Communicating with chaos. **Physical Review Letters**, v. 70, n. 20, p. 1781-1784, 1993.
- HAYES, S.; GREBOGI, C.; OTT, E.; MARK, A. Experimental control of chaos for communication. **Physical Review Letters**, v. 73, n. 13, p. 3031-3034, 1994.
- HAYKIN, S. S. *Adaptive filter theory*, 3th edition, Upper Saddle River: Prentice Hall, 1996.
- HAYKIN, S. S. *Communication systems*, 4th edition, New York: Wiley, 2000.
- KAY, S. M., *Fundamentals of statistical signal processing*. New Jersey: Prentice Hall PTR, 1993.
- KAY, S. Asymptotic maximum likelihood estimator performance for chaotic signals in noise. **IEEE Transactions on Signal Processing**, v. 43, n. 4, 1995.
- KAY, S.; NAGESHA, V. Methods for chaotic signal estimation. **IEEE Transactions on Signal Processing**, v. 43, n. 8, 1995.
- KENNEDY, M. P. Three steps to chaos - part I: evolution. **IEEE Transactions on Circuits and Systems - I**, v. 40, n. 10, p. 640-656, 1993.
- KENNEDY, M.P.; KOLUMBÁN, G. Digital communications using chaos. **Signal Processing**, v. 80, p. 1307-1320, 2000.
- KENNEDY, M. P.; KOLUMBÁN, G.; KIS, G.; JÁKÓ, Z. Performance evaluation of FM-DCSK modulation in multipath environments. **IEEE Transactions on Circuits and Systems - I**, v. 47, n. 12, p. 1702-1711, 2000.
- KENNEDY, M. P.; ROVATTI, R.; SETTI, G. *Chaotic electronics in telecommunications*, Boca Raton: CRC Press, 2000.

- KISEL, A.; DEDIEU, H.; SCHIMMING, T. Maximum likelihood approaches for noncoherent communications with chaotic carriers. **IEEE Transactions on Circuits and Systems - I**, v. 48, n. 5, p. 533-542, 2001.
- KOLUMBÁN, G.; KENNEDY, M. P.; CHUA, L. O. The role of synchronization in digital communications using chaos - part i: fundamentals of digital communications. **IEEE Transactions on Circuits and Systems - I**, v. 44, n. 10, p. 927-946, 1997.
- KOLUMBÁN, G.; KENNEDY, M. P.; CHUA, L. O. The role of synchronization in digital communications using chaos - part II: chaotic modulation and chaotic synchronization. **IEEE Transactions on Circuits and Systems - I**, v. 45, n. 11, p. 1129-1140, 1998.
- KOLUMBÁN, G.; KENNEDY, M. P.; KIS, G.; JÁKÓ Z. FM-DCSK: a novel method for chaotic communications. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, Monterey, 1998. *Proc. ISCAS'98*, v. 4, p. 477-480, 1998.
- LASOTA, A.; MACKEY, M. *Probabilistic properties of deterministic systems*, Cambridge: Cambridge University, 1985.
- LATHI, B. P. *Modern digital and analog communication systems*, 3rd edition, New York: Oxford University Press, 1998.
- LAU, F. C. M.; TSE, C. K. *Chaos-based digital communication systems*, Berlin: Springer, 2003.
- LAU, F. C. M.; TSE, C. K.; HAU, S. F. Coexistence of chaos-based and conventional digital communication systems of equal bit rate. **IEEE Transactions on Circuits and Systems - I**, v. 51, n. 2, p. 391-408, 2004.
- LI, T. Y.; YORKE, J. A. Period three implies chaos. **American Mathematics Monthly**, v. 82, p. 985-992, 1975.
- LORENZ, E. Deterministic non-periodic flows. **Journal of the Atmospheric Sciences**, v. 20, p. 130, 1963.

- MANDAL, S.; BANERJEE, S. Analysis and CMOS implementation of a chaos-based communication system. **IEEE Transactions on Circuits and Systems - I**, v. 51, n. 9, p. 1708-1722, 2004.
- MANDELBROT, B. B. *The fractal geometry of nature*, San Francisco: Freeman, 1982.
- MARTEAU, P. F.; ABARBANEL, H. D. I. Noise reduction in chaotic time series using scaled probabilistic method. **Jornal of Nonlinear Science**, v. 1, p. 313-343, 1991.
- MAY, R. M. Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos. **Science**, v. 186, p. 645-647, 1974.
- MILLERIOUX, G.; DAAFOUZ, J. Global chaos synchronization and robust filtering in noisy context. **IEEE Transactions on Circuits and Systems - I**, v. 48, n. 10, p. 1170-1176, 2001.
- MONTEIRO, L. H. A. *Sistemas dinâmicos*, São Paulo: Livraria da Física, 2002.
- NEWTON, I. *The Principia : mathematical principles of natural philosophy*. Translator I. B. Cohen, Berkeley: University of California, 1999.
- PALIS J.; MELO W. *Introdução aos sistemas dinâmicos*, São Paulo: Edgard Blücher, 1978.
- PALIS, J.; MELO, W. *Geometric theory of dynamical systems: an introduction*, New York: Springer-Verlag, 1982.
- PAPADOPOULOS, H. C.; WORNELL, G. W. Optimal detection of a class of chaotic signals. In: IEEE CONFERENCE ON ACOUSTIC, SPEECH AND SIGNAL PROCESSING, Minneapolis, 1993. *Proc. ICASSP'93*, v. 3, p. 117-120, 1993.
- PAPOULIS, A.; PILLAI, S. U. *Probability, random variables and stochastic processes*, 4th edition, New York: McGraw-Hill, 2002.
- PARLITZ, U.; CHUA, L. O.; KOCAREV, L. J.; HALLE, K. S.; SHANG A. Transmission of digital signals by chaotic synchronization. **International Journal of Bifurcation and Chaos**, v. 2, n. 4, p. 973-977, 1992.

- PECORA, L. M.; CARROLL, T. L. Synchronization in chaotic systems. **Physical Review Letters**, v. 64, n. 8, pp. 821-824, 1990.
- POINCARÉ, H. Sur les équations de la dynamique et le problème de trois corps. **Acta Mathematica**, v. 13, p. 1-270, 1890.
- PROAKIS, J. G. *Digital communications*, New York: McGraw-Hill, 1995.
- ROVATTI, R.; MAZZINI, G.; SETTI, G. On the ultimate limits of chaos-based asynchronous DS-CDMA-I: basic definitions and results. **IEEE Transactions on Circuits and Systems - I**, v. 51, n. 7, p. 1348-1364, 2004.
- SETTI, G.; MAZZINI, G.; ROVATTI, R.; CALLEGARI, S. Statistical modeling of discrete-time chaotic process - basic finite-dimensional tools and applications. **Proceedings of the IEEE**, v. 90, n. 5, 2002.
- SKLAR, B. *Digital communications: fundamentals and applications*, Upper Saddle River: Prentice Hall PTR, 2001.
- SMALE, S. Differentiable dynamical systems. **Bulletin of the American Mathematical Society**, v. 73, p. 747-817, 1967.
- STROGATZ, S. H. *Nonlinear dynamics and chaos with applications to Physics, Biology, Chemistry and Engineering*, Reading: Addison-Wesley, 1998.
- SUSHCHIK, M.; TSIMRING, L. S.; VOLKOVSKII, A. R. Performance analysis of correlation-based schemes utilizing chaos. **IEEE Transactions on Circuits and Systems - I**, v. 47, n. 12, p. 1684-1691, 2000.
- VITERBI, A.J. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. **IEEE Transactions on Information Theory**, v. 13, p. 260-269, 1967.
- WADE, G. *Coding techniques: an introduction to compression and error control*, New York: Palgrave, 2000.

- WILLIAMS, C. Chaotic communications over radio channels. **IEEE Transactions on Circuits and Systems - I**, v. 48, n. 12, p. 1394-1404, 2001.
- WOZENCRAFT, J. M.; JACOBS, I. M. *Principles of communication engineering*, New York: John Wiley and Sons, 1965.
- WU, C. W.; CHUA, L. O. A simple way to synchronize chaotic systems with applications to secure communication systems. **International Journal of Bifurcation and Chaos**, v. 3, n. 6, p. 1619-1627, 1993.
- YAMADA, T.; FUJISAKA, H. Stability Theory of Synchronized Motion in Coupled-Oscillator Systems - II The Mapping Approach. **Progress of Theoretical Physics**, v. 70, n. 5, 1983.