

MÉTODOS BAYESIANOS PARA RESTAURAÇÃO DE SINAIS DE ÁUDIO
COM DISTORÇÕES NÃO-LINEARES

Flávio Rainho Ávila

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Luiz Wagner Pereira Biscainho

Rio de Janeiro
Novembro de 2012

MÉTODOS BAYESIANOS PARA RESTAURAÇÃO DE SINAIS DE ÁUDIO
COM DISTORÇÕES NÃO-LINEARES

Flávio Rainho Ávila

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Prof. Marcello Luiz Rodrigues de Campos, Ph.D.

Prof. Dani Gamerman, Ph.D.

Prof. José Antonio Apolinário Junior, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2012

Ávila, Flávio Rainho

Métodos Bayesianos para Restauração de Sinais de
Áudio com Distorções Não-Lineares/Flávio Rainho Ávila.
– Rio de Janeiro: UFRJ/COPPE, 2012.

XVIII, 136 p.: il.; 29, 7cm.

Orientador: Luiz Wagner Pereira Biscainho

Tese (doutorado) – UFRJ/COPPE/Programa de
Engenharia Elétrica, 2012.

Referências Bibliográficas: p. 125 – 130.

1. Processamento bayesiano. 2. Distorções não-
lineares. 3. MCMC. I. Biscainho, Luiz Wagner Pereira.
II. Universidade Federal do Rio de Janeiro, COPPE,
Programa de Engenharia Elétrica. III. Título.

Aos meus pais

Agradecimentos

Ao meu orientador Luiz Wagner, pelas ideias magistrais, aulas excepcionais, conversas informais e rimas sensacionais.

Aos meus pais, Luciana e João Carlos, por me apoiarem nos piores momentos, até mesmo quando decidi fazer pós-graduação.

Aos meus colegas do Laboratório de Processamento de Sinais, por me tratarem com o respeito e a consideração que eu não mereço.

Ao CNPq e à FAPERJ pelo apoio financeiro.

Aos membros da banca, pelas críticas pertinentes e sugestões valiosas, que ajudaram a melhorar esta tese.

Um agradecimento especial àquela que tem estado ao meu lado nos últimos anos, me apoiando e incentivando em todos os momentos. Mesmo sem entender muito bem meu trabalho, ela foi uma das maiores responsáveis por sua conclusão.

Diana: obrigado pelo apoio e pelos *cookies*.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MÉTODOS BAYESIANOS PARA RESTAURAÇÃO DE SINAIS DE ÁUDIO COM DISTORÇÕES NÃO-LINEARES

Flávio Rainho Ávila

Novembro/2012

Orientador: Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

Distorções não-lineares estão presentes em inúmeras aplicações de áudio e voz, causando artefatos que impactam severamente a qualidade dos sinais, o que justifica a elaboração de técnicas para sua compensação. Devido à dificuldade em se tratar esse tipo de distorção de forma genérica, atacamos separadamente três categorias principais. Para cada uma, propusemos métodos de restauração de diferentes níveis de complexidade e acurácia, produzindo assim um conjunto amplo de ferramentas, adequadas a diferentes necessidades.

Para sinais com limitação de amplitude, obtivemos ganhos substanciais em termos de SNR, especialmente para degradação severa, através de uma solução bayesiana que respeita as restrições de amplitude inerentes ao problema. Propomos ainda variações no algoritmo para lidar com a presença de ruído aditivo, que pode incidir antes ou depois da distorção não-linear. Para distorções invertíveis sem memória, resolvemos o problema de identificação da não-linearidade por sua expansão em série de Taylor e também através de um modelo linear por partes, além de considerarmos a presença de ruído aditivo. Para o caso mais geral de distorção invertível com memória, propomos métodos de identificação cega de um sistema não-linear com estrutura adequada para a descrição de um conjunto amplo de distorções encontradas em sinais reais.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

BAYESIAN METHODS FOR RESTORATION OF NONLINEARLY
DISTORTED AUDIO SIGNALS

Flávio Rainho Ávila

November/2012

Advisor: Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

Nonlinear distortions appear in many speech and audio applications, causing artifacts that severely impact the signals quality, thus calling for techniques to compensate for them. Because of the difficulty of providing a general treatment for this type of distortions, we tackle separately three main categories, and, for each one, we propose methods of different levels of complexity and accuracy, thus producing a wide set of tools suitable to different needs.

For signals with amplitude limitation, we achieved substantial gains in terms of SNR, specially for severe degradation, by proposing a Bayesian solution that respects the amplitude constraints inherent to the problem. We also propose modifications of the basic algorithm to deal with the presence of noise, which can be added before or after the nonlinear distortion. For memoryless invertible distortions, we solve the problem of identification of the nonlinearity via its Taylor series expansion, as well as a piece-wise linear model, and we also consider the presence of additive noise. For the more general case of nonlinearity with memory, we propose methods for identification of a nonlinear system whose structure is suitable for the description of a wide range of nonlinear distortions found in real signals.

Sumário

Lista de Figuras	xii
1 Introdução	1
1.1 Restauração digital de áudio	2
1.2 Estrutura da tese	6
2 Inferência Bayesiana	8
2.1 Algumas interpretações de probabilidade	10
2.2 Problema geral de inferência estatística e algumas soluções	11
2.2.1 Teorema de Bayes	12
2.2.2 Alguns critérios de otimização	13
2.2.3 Exemplo	13
2.3 Modelo bayesiano hierárquico	14
2.4 Distribuição <i>a priori</i>	15
2.4.1 Prioris conjugadas	16
2.4.2 Priori não-informativa	16
2.5 Eliminação de parâmetros	17
2.6 Comentários conclusivos	17
3 Técnicas Numéricas de Estimação	19
3.1 Técnicas de Monte Carlo	20
3.2 Método da Rejeição	20
3.3 Cadeias de Markov	21
3.3.1 Conceituação	21
3.3.2 Distribuição invariante	23
3.3.3 Ergodicidade	24

3.3.4	Cadeias de Markov para espaço de estados contínuo	24
3.4	Amostrador de Gibbs	25
3.5	Algoritmo de Metropolis-Hastings	26
3.6	Análise de convergência	27
3.7	Conclusões	29
4	Modelagem de Sinais de Áudio com Distorções Não-Lineares	30
4.1	Modelagem de sinais de áudio através de modelo autorregressivo	31
4.1.1	Estimação dos parâmetros do modelo AR	33
4.1.2	Incorporando um modelo para o ruído	37
4.2	Modelagem da Distorção Não-linear	39
4.2.1	Resumo das soluções propostas	43
5	Restauração de Sinais com Limitação em Amplitude	45
5.1	Formulação do Problema	47
5.2	Solução Bayesiana	48
5.2.1	Algoritmo proposto	49
5.2.2	Resultados	50
5.3	Modelagem de ruído aditivo	51
5.3.1	Ruído seguido de não-linearidade	52
5.3.1.1	Cálculo da distribuição <i>a posteriori</i>	54
5.3.1.2	Algoritmo proposto	56
5.3.2	Não-linearidade seguida de ruído	58
5.3.2.1	Cálculo da distribuição <i>a posteriori</i>	58
5.3.2.2	Algoritmo proposto	58
5.3.3	Resultados	61
5.4	Conclusões	67
6	Distorção Não-Linear Invertível Sem Memória	70
6.1	Trabalhos anteriores	73
6.2	Estimativa inicial	73
6.2.1	Método A	74
6.2.2	Método B	75

6.2.3	Estimativa dos coeficientes	76
6.2.4	Simulações	76
6.3	Cálculo da verossimilhança	76
6.4	Maximização da verossimilhança	78
6.4.1	Solução pelo método do Gradiente	79
6.4.2	Solução pelo método de Gauss-Newton	80
6.5	Solução bayesiana	81
6.6	Modelo linear por partes	89
6.7	Modelagem de ruído aditivo	93
6.8	Conclusões	98
7	Restauração de Sinais Corrompidos com Distorção Não-linear com Memória	103
7.1	Formulação do problema	105
7.2	Cálculo da verossimilhança	106
7.3	Maximização da verossimilhança	106
7.4	Solução bayesiana	108
7.5	Conclusões	113
8	Conclusões e Trabalhos Futuros	119
8.1	Limitação de amplitude	119
8.2	Não-linearidade invertível sem memória	121
8.3	Distorção não-linear com memória	121
8.4	Trabalhos futuros	122
8.4.1	Solução no domínio do tempo	122
8.4.2	Solução no domínio da transformada	123
8.4.3	Elementos psicoacústicos	124
	Referências Bibliográficas	125
A	Algumas Demonstrações	131
A.1	Produto de gaussianas multivariáveis	131
A.2	Cálculo do Jacobiano para o método de Gauss-Newton	132

A.3 Cálculo do gradiente e da hessiana para o modelo ICM	133
B Testes com Sinais Reais	135

Lista de Figuras

1.1	Cadeia genérica de degradação e restauração de áudio.	3
1.2	Sinal contaminado por ruído impulsivo.	6
2.1	Exemplo de modelo bayesiano hierárquico. Os conjuntos de dados x_1, x_2, \dots, x_n , dependem de parâmetros $\theta_1, \theta_2, \dots, \theta_n$, respectivamente, que por sua vez são instâncias de variáveis aleatórias descritas por parâmetros ϕ	15
3.1	Evolução de um parâmetro de uma cadeia de Markov a partir de diferentes valores iniciais. Vemos que as amostras tendem a variar na mesma região a partir de certo ponto, independente da inicialização, evidenciando a convergência do algoritmo.	28
4.1	Modelo de sinal de áudio com distorções não-lineares.	30
4.2	Trecho de sinal gerado por modelo AR.	35
4.3	Evolução dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs.	36
4.4	Evolução da variância de excitação do modelo AR obtida pelo amostrador de Gibbs.	36
4.5	Histograma da variância de excitação do modelo AR, obtida pelo amostrador de Gibbs.	37
4.6	Histogramas de cada elemento dos coeficientes do modelo AR, obtidos a partir das amostras geradas pelo amostrador de Gibbs.	37
4.7	Sinal gerado por modelo AR, contaminado com ruído.	39
4.8	Evolução dos coeficientes do modelo AR para sinal com ruído.	40

4.9	Evolução da variância de excitação do modelo AR, obtidos pelo amostrador de Gibbs, para sinal com ruído.	41
4.10	Histograma da variância de excitação do modelo AR, obtidos pelo amostrador de Gibbs, para sinal com ruído.	41
4.11	Histogramas de cada elemento dos coeficientes do modelo AR, obtidos pelo amostrador de Gibbs, para sinal com ruído.	42
4.12	Evolução da variância de excitação do modelo AR obtidos pelo amostrador de Gibbs a partir de um sinal ruidoso, mas sem modelar o ruído.	42
4.13	Evolução dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs a partir de um sinal ruidoso, mas sem modelar o ruído.	43
5.1	Comparação entre trechos de sinal original, saturado e restaurado pelo método proposto. O sinal saturado foi limitado a uma amplitude de 25 % da amplitude máxima do sinal original.	51
5.2	Evolução dos coeficientes do modelo AR, obtidos com o amostrador de Gibbs a partir de um sinal com saturação. A inicialização com base no método de mínimos quadrados já está bem próxima dos valores corretos. O amostrador de Gibbs corrige essa polarização.	52
5.3	Evolução da variância de excitação do modelo AR, obtidos com o amostrador de Gibbs a partir de um sinal com saturação. Como a inicialização com base no método dos mínimos quadrados minimiza a variância do ruído, essa variável tende a ser subestimada.	53
5.4	Histograma da variância excitação do modelo AR obtido a partir das amostras geradas pelo amostrador de Gibbs a partir de um sinal com saturação.	53
5.5	Histogramas de cada elemento dos coeficientes do modelo AR, obtidos a partir das amostras geradas pelo amostrador de Gibbs a partir de um sinal com saturação.	54
5.6	Histograma da amostra 340 (numa região corrompida) do sinal restaurado obtido a partir do amostrador de Gibbs a partir de um sinal com saturação.	54

5.7	Comparação entre o método proposto (bayesiano) e o método baseado em mínimos quadrados, usado como referência, em termos de SNR. Ambos os métodos aumentam a SNR em relação ao sinal saturado. Para saturação muito severa, o método proposto oferece aumentos maiores de SNRs, mas os resultados são muito próximos para saturação moderada ou leve.	55
5.8	Diagrama de blocos descrevendo o modelo para sinais com saturação contaminado com ruído. (a) Ruído antes da saturação. (b) Ruído depois da saturação.	56
5.9	Comparação entre sinais original, saturado com ruído (antes da saturação) e restaurado pelo algoritmo proposto.	62
5.10	Evolução dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).	62
5.11	Evolução da variância de excitação do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).	63
5.12	Histograma da variância de excitação do modelo AR obtida através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).	63
5.13	Histogramas de cada elemento dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).	64
5.14	Histograma da amostra 192 do sinal restaurado obtido através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).	64
5.15	Comparação entre sinais original, saturado com ruído (depois da saturação) e restaurado.	65
5.16	Evolução dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).	65

5.17	Evolução da variância de excitação do modelo AR obtido através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).	66
5.18	Histograma da variância excitação do modelo AR obtida através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).	66
5.19	Histogramas de cada elemento dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).	67
5.20	Evolução do limite superior de amplitude c_+ para um sinal com saturação e ruído (após a saturação).	67
5.21	Evolução de amostras do limite inferior de amplitude c_- obtidas a partir do algoritmo proposto com base em um sinal com saturação e ruído (após a saturação).	68
5.22	Histograma do limite superior c_+ obtida com o método proposto com base em um sinal com saturação e ruído (após a saturação).	68
5.23	Histograma do limite superior c_- , obtida com o método proposto com base em um sinal com saturação e ruído (após a saturação).	69
5.24	Histograma da amostra 120 do sinal restaurado obtidas através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).	69
6.1	Diagrama de blocos de sinal distorcido com saturação gradual, representando uma distorção não-linear sem memória.	70
6.2	Sinal original contra sinais restaurados obtidos com métodos A e B, comparados com curva ideal (identidade) e curva. Ambos os métodos reduzem sensivelmente a não-linearidade do sinal, mas deixam distorções significativas em altas amplitudes.	77
6.3	Evolução de coeficientes da expansão de Taylor pelo método de Gauss-Newton.	81
6.4	Evolução dos coeficientes do modelo AR pelo método de Gauss-Newton.	82
6.5	Comparação entre sinais original, saturado e restaurado pelo método de Gauss-Newton.	82

6.6	Sinal original contra sinal restaurado obtido com método de Gauss-Newton, comparados com curva ideal (identidade).	83
6.7	Comparação entre sinais original, saturado e restaurado por método bayesiano.	86
6.8	Sinal original contra sinal restaurado obtido com método de bayesiano, comparados com curva ideal (identidade).	86
6.9	Evolução dos coeficientes do modelo AR através de amostrador de Gibbs.	87
6.10	Evolução da variância de excitação do modelo AR através de amostrador de Gibbs.	87
6.11	Evolução dos coeficientes da curva não linear através de amostrador de Gibbs.	88
6.12	Histogramas de cada elemento dos coeficientes do modelo AR, pelo amostrador de Gibbs.	88
6.13	Histograma da excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs.	89
6.14	Histogramas dos coeficientes da curva não-linear, obtida pelo amostrador de Gibbs.	90
6.15	Ilustração do modelo linear por partes.	91
6.16	Evolução dos coeficientes do modelo AR através de amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.	92
6.17	Evolução da variância de excitação do modelo AR através de amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.	92
6.18	Evolução dos coeficientes das retas do modelo linear por partes, através de amostrador de Gibbs.	93
6.19	Histograma da variância de excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.	93
6.20	Histogramas dos coeficientes das retas do modelo linear por partes.	94

6.21	Histogramas de cada elemento dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.	94
6.22	Modelo de distorção não-linear sem memória com observação ruidosa.	95
6.23	Comparação entre sinais original, saturado com ruído e restaurado por método bayesiano.	98
6.24	Sinal original contra sinal restaurado obtido com método Bayesiano, a partir de sinal distorcido com ruído, comparado com curva ideal (identidade).	99
6.25	Evolução dos coeficientes do modelo AR através de amostrador de Gibbs, obtidos a partir de sinal saturado com ruído.	100
6.26	Evolução da variância de excitação do modelo AR através de amostrador de Gibbs, obtida a partir de sinal saturado com ruído.	100
6.27	Evolução dos coeficientes da curva não linear através de amostrador de Gibbs, obtidos a partir de sinal saturado com ruído.	101
6.28	Histogramas dos coeficientes da curva não-linear obtida pelo amostrador de Gibbs, a partir de sinal saturado com ruído.	101
6.29	Histogramas de cada elemento dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs, a partir de um sinal saturado com ruído. .	102
6.30	Histograma da excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs, a partir de um sinal saturado com ruído.	102
7.1	Modelo de distorção não-linear com memória.	105
7.2	Evolução dos coeficientes do filtro linear $B(z)$ pelo método de Gauss-Newton.	109
7.3	Evolução dos parâmetros da não-linearidade pelo método de Gauss-Newton.	109
7.4	Evolução dos coeficientes do modelo AR pelo método de Gauss-Newton.	110
7.5	Comparação entre sinais original, degradado e restaurado pelo método de Gauss-Newton.	110
7.6	Sinal original contra sinal restaurado obtido com método de Gauss-Newton, comparados com curva ideal (identidade).	111

7.7	Comparação entre sinais original, degradado e restaurado por método bayesiano.	113
7.8	Sinal original contra sinal restaurado obtido com método de bayesiano, comparados com curva ideal (identidade).	114
7.9	Evolução dos coeficientes do filtro linear através de amostrador de Gibbs.	114
7.10	Evolução dos coeficientes da curva não linear através de amostrador de Gibbs.	115
7.11	Evolução dos coeficientes do modelo AR através de amostrador de Gibbs.	115
7.12	Evolução da variância de excitação do modelo AR através de amostrador de Gibbs.	116
7.13	Histogramas dos coeficientes do filtro linear, obtida pelo amostrador de Gibbs.	116
7.14	Histogramas dos coeficientes da curva não-linear, obtida pelo amostrador de Gibbs.	117
7.15	Histogramas de cada elemento dos coeficientes do modelo AR, pelo amostrador de Gibbs.	117
7.16	Histograma da excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs.	118

Capítulo 1

Introdução

Distorções não-lineares ocorrem em diversas aplicações de áudio e voz, provocando artefatos audíveis que comprometem a qualidade e inteligibilidade dos sinais. Causas desse tipo de distorção incluem a codificação de voz em modernos sistemas de telecomunicações, não-linearidade de amplificadores e artefatos gerados por algoritmos de remoção de ruído, além de danos em meios de armazenamento de gravações antigas. Assim sendo, algoritmos para compensação de distorções não-lineares seriam de grande importância para provedores de serviços de telecomunicações, bem como para manutenção de arquivos de conteúdo histórico de interesse público.

Apesar de importante e comum, esse tipo de distorção recebeu pouca atenção na literatura de restauração de áudio, possivelmente devido à inerente dificuldade do problema. Com exceção de métodos para o tratamento de limitação de amplitude, as técnicas propostas até o momento para casos mais gerais de distorção não-linear falham quando aplicadas a sinais com distorções reais, ora porque assumem modelos pouco realistas para o sinal original de áudio, ora porque não são robustas a outros tipos de defeito que concomitantemente distorcem o sinal.

Inserindo-se na área de processamento estatístico de sinais, o objetivo desta tese é a elaboração de algoritmos capazes de superar o estado-da-arte em restauração de sinais de áudio distorcidos por diversas formas de não-linearidade. Devido à generalidade desse tipo de distorções, daremos um tratamento separado às seguintes categorias: (1) limitação de amplitude, (2) distorção não-linear invertível sem memória, e (3) distorção não-linear invertível com memória.

Para identificar a distorção e recuperar o sinal original em cada uma dessas

categorias, propomos algoritmos de variados níveis de desempenho e complexidade. Embora o foco principal da tese seja em algoritmos bayesianos baseados em modelos, em que tanto o sinal de áudio quanto as distorções são modelados estatisticamente, propomos também técnicas mais simples, que poderiam ser usadas isoladamente quando o tempo de processamento for um fator crítico, ou como ponto de partida favorável para os algoritmos mais intensivos computacionalmente.

Apesar de termos em mente aplicações de áudio e voz, podemos vislumbrar a aplicação das ideias contidas nos algoritmos propostos em diversas outras áreas, em particular controle não-linear e equalização de canal em sistemas de telecomunicações, nas quais a identificação de sistemas não-lineares é de grande interesse. Embora esse assunto seja foco intensivo de pesquisa, até onde sabemos, poucos trabalhos foram publicados em identificação cega de sistemas não-lineares, isto é, sem acesso ao sinal de entrada. Nosso trabalho fornece uma contribuição significativa ao propor um arcabouço para realizar essa tarefa usando o paradigma bayesiano, a qual permite lidar de forma natural com ruído colorido na entrada, ordem do sistema não-linear desconhecida, além de modelagem de ruído de medição na saída do sistema.

Continuamos o capítulo apresentando uma visão geral de sistemas de restauração digital de áudio, e o concluímos com uma descrição da estrutura da tese.

1.1 Restauração digital de áudio

Sistemas de restauração de áudio têm como objetivo reduzir ou eliminar defeitos audíveis presentes em sinais de áudio, introduzidos pelos mecanismos de gravação e reprodução, ou resultantes de deterioração ou desgaste dos meios de gravação [1], [2]. As distorções causadas pelos sistemas de gravação e reprodução são em alguns casos aproximadamente conhecidas, o que torna possível sua compensação, ainda que imperfeitamente. Já os defeitos introduzidos nos meios de armazenamento decorrem da ação do tempo e do uso, e são em grande medida aleatórios, sendo por isso mais desafiadores para a restauração.

A Figura 1.1 mostra um diagrama de blocos de uma cadeia genérica de degradação e restauração de áudio. O sistema de restauração recebe como entrada o sinal

distorcido e, usando o conhecimento prévio sobre características do sinal original de áudio, bem como sobre as distorções incidentes no sinal, deve gerar na saída um sinal auditivamente similar ao sinal original.

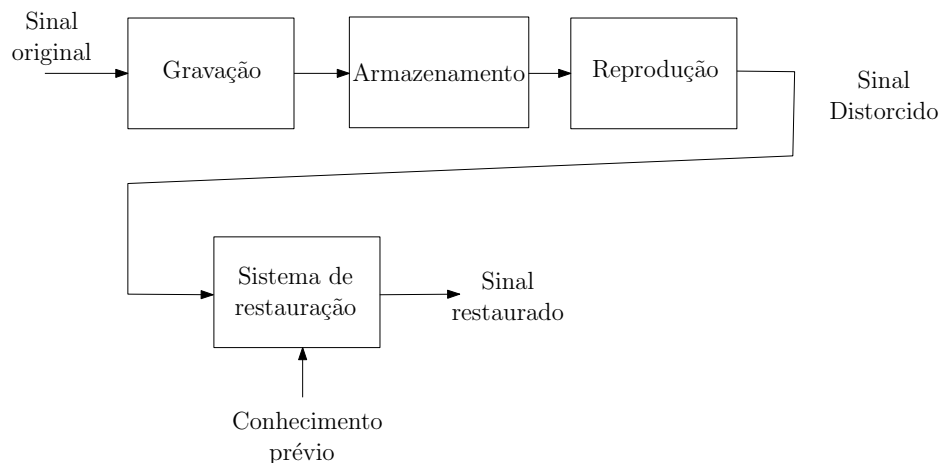


Figura 1.1: Cadeia genérica de degradação e restauração de áudio.

Os defeitos são genericamente classificados como localizados, em que apenas algumas amostras estão corrompidas, ou globais, em que todas as amostras sofreram algum tipo de distorção. Como exemplos mais comuns do primeiro tipo podemos citar o ruído impulsivo (*clicks*), gerado por partículas de poeira ou arranhões na superfície de um disco de vinil, e pulsos longos, caracterizados por oscilações de longa duração e baixa frequência, usualmente gerados por quebra ou grandes arranhões na superfície do disco. Como exemplos comuns de degradação global podemos citar o ruído de fundo, de origem variada, e os defeitos de variação de *pitch*, causados por alteração na velocidade de reprodução do sinal. Também podemos classificar as distorções como lineares, em que diferentes frequências presentes no sinal são alteradas diferentemente, ou não-lineares, em que novas frequências são introduzidas no sinal.

Métodos eficazes para remoção de diversos dos defeitos mencionados acima têm recebido grande atenção na literatura de processamento de áudio, variando desde técnicas heurísticas simples e computacionalmente eficientes até métodos estatísticos sofisticados e computacionalmente complexos, capazes de obter estimativas para o sinal original com alta qualidade sob o ponto de vista perceptivo.

Diversos aspectos precisam ser levados em conta na elaboração de um sistema de restauração de áudio. A seguir, descrevemos aqueles que consideramos mais

importantes.

O primeiro é a descrição estatística do sinal original que se deseja recuperar. Sabemos que sinais de áudio e voz, devido aos mecanismos físicos que o geram, possuem propriedades que os tornam até certo ponto preditíveis, isto é, conhecidas algumas de suas amostras é possível estimar, com variado grau de acurácia, amostras desconhecidas. Naturalmente, o modelo será tão mais acurado quanto mais se souber sobre o tipo de sinal, os instrumentos presentes numa música, características do falante, condições acústicas no ambiente de gravação etc. Por outro lado, se desejamos um sistema que seja geral o bastante para lidar com diversos tipos de defeitos, temos que aceitar resultados em geral piores do que se obteria por soluções particulares.

Diversos modelos e métodos para descrever sinais de áudio foram propostos e vêm sendo usados com sucesso em muitas aplicações. O mais popular é o modelo autorregressivo (AR) [2], em que o valor esperado da amostra atual é estimado por uma combinação linear das amostras em instantes anteriores. Usualmente, o erro de predição é modelado por ruído branco gaussiano, embora em [3] se argumente que uma excitação do tipo *Student-t* seria mais robusta para certos tipos de sinais.

Para lidar com a não-estacionaridade de sinais de áudio, podemos realizar o processamento separadamente em trechos curtos de sinal, atribuindo parâmetros diferentes para o modelo em cada um deles, ou, alternativamente, usar o modelo autorregressivo variante no tempo (TVAR – *Time-Varying Autorregressive*) [4], que permite que os coeficientes do modelo AR variem continuamente no tempo.

Caso o instrumento que gerou o sinal de áudio seja conhecido, podemos adotar um modelo mais realista, tendo como base a análise física do mecanismo de geração do som. Encontramos essa abordagem em [5], no qual os autores propõem o uso de modelos físicos para restauração e melhoramento de gravações de violão.

O segundo aspecto importante é a forma como usamos o conhecimento sobre o mecanismo físico responsável pela geração de distorções no sinal de áudio na elaboração do modelo do sinal degradado. Existe, em geral, um compromisso entre o realismo e a simplicidade do modelo, sendo comum sacrificar sua acurácia a fim de torná-lo matematicamente tratável. As técnicas mais simples (heurísticas), tratam a distorção como um elemento presente no sinal degradado que não pode ser

bem descrito pelo modelo atribuído ao sinal original. Os métodos estatísticos baseados em modelos [2] vão um passo além, ao atribuírem explicitamente um modelo probabilístico para a degradação. Em particular, os métodos bayesianos permitem que se incorpore todo o conhecimento disponível sobre os parâmetros do modelo na forma de suas distribuições *a priori*, produzindo assim modelos bastante complexos que requerem técnicas numéricas computacionalmente intensivas para obtenção das distribuições de interesse. Dentre essas técnicas, as baseadas em MCMC (*Markov Chain Monte Carlo*) são provavelmente as mais populares e serão empregadas extensivamente nesta tese.

Outro aspecto relevante é a avaliação de qualidade do sinal restaurado. Sabemos que critérios tradicionais para avaliação do nível de distorção, tais como a SNR (*Signal-to-Noise Ratio*) e a THD (*Total Harmonic Distortion*) são inadequadas para avaliar os efeitos perceptivos da distorção, uma vez que não levam em conta aspectos perceptivos [6]. Idealmente, a avaliação de qualidade do sinal restaurado seria feita através de ouvintes humanos, entretanto, algoritmos para avaliação automática de qualidade, tais como o PEAQ (*Perceptual Evaluation of Audio Quality*) [7], são uma alternativa rápida e barata a testes subjetivos de audição.

Para fins de ilustração, consideremos o problema de restauração de sinais corrompidos por ruído impulsivo, caracterizados como distúrbios de curta duração e localização aleatória. A Figura 1.2 mostra um trecho de sinal de áudio corrompido por distúrbio impulsivo localizado entre as amostras 175 e 180. Podemos observar que o sinal de áudio subjacente possui características que o tornam até certo ponto preditível, isto é, conhecidas as propriedades estatísticas do sinal, é possível estimar amostras futuras com base nas amostras passadas. Sendo assim, pode-se estimar as amostras da região corrompida com base nas amostras vizinhas, tanto à esquerda quando à direita da degradação. Caso as propriedades estatísticas da degradação sejam conhecidas, a informação contida nas amostras degradadas também pode ser usada para que se obtenha uma estimativa mais confiável do sinal original.

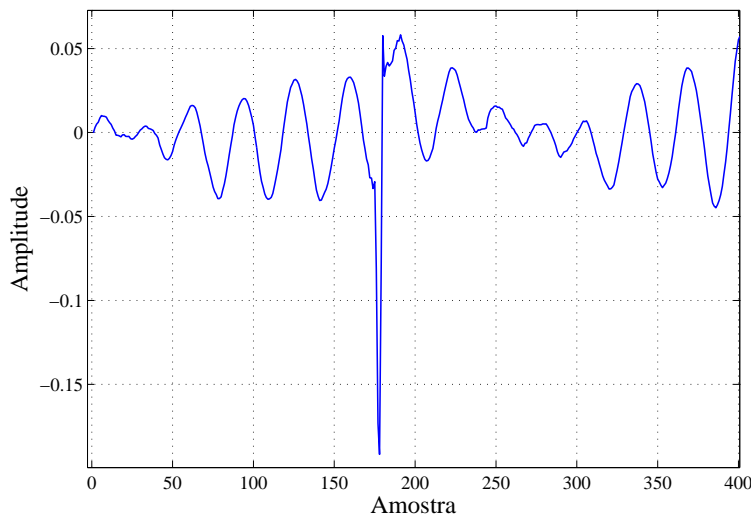


Figura 1.2: Sinal contaminado por ruído impulsivo.

1.2 Estrutura da tese

Os dois próximos capítulos destinam-se à apresentação de conceitos de inferência estatística e métodos numéricos de otimização que foram usados ao longo do trabalho. Devido à base teórica comum da presente tese com a dissertação de mestrado do autor [8], grande parte desses dois capítulos foi baseada naquele texto, com o cuidado de se manter apenas o essencial para o entendimento dos capítulos seguintes.

No Capítulo 4, vamos descrever de forma mais detalhada os modelos que usamos para sinais de áudio e distorções não-lineares. Começamos com a definição do modelo AR e mostramos como técnicas numéricas podem ser usadas para estimação de seus parâmetros no contexto bayesiano. Na segunda parte do capítulo, fazemos uma revisão sobre modelagem de sistemas não-lineares, enfatizando aqueles que serão usados nesta tese e destacando como eles se inserem no contexto geral.

Nos Capítulos 5, 6 e 7, encontram-se o desenvolvimento teórico dos algoritmos propostos na tese, a análise de seus resultados sobre sinais artificiais e reais e a comparação com métodos concorrentes, quando for o caso.

O Capítulo 5 se dedica ao problema de limitação de amplitude em sinais de áudio. Propomos algoritmos para obtenção das amostras desconhecidas com base nas amostras corretas, respeitando os limites de amplitude impostos pelo tipo de distorção. Também consideramos os casos em que ruído aditivo está presente, tanto antes quanto depois do sistema limitador de amplitude.

No Capítulo 6, procuramos resolver o problema de saturação gradual, representado por uma distorção não-linear invertível sem memória, através de algoritmos de graus variados de complexidade e desempenho. Adotamos um modelo que descreve a não-linearidade por sua representação em série de Taylor e propomos três algoritmos capazes de estimar os termos dessa série, permitindo assim obter a curva de distorção. Propomos também uma variação em um dos algoritmos que permite recuperar o sinal original quando ruído aditivo está presente na saída do sistema. Outra contribuição do capítulo é a proposta de um modelo linear por partes para modelagem da curva não-linear e de um método para estimação dos parâmetros de cada reta.

No Capítulo 7, mostramos como podemos resolver o problema de distorção não-linear com memória através de um modelo formado por uma cascata de uma não-linearidade sem memória com um filtro linear só-pólos, estrutura conhecida como de Hammerstein, que seria capaz de descrever satisfatoriamente um conjunto amplo de distorções não-lineares em sinais reais. Propomos dois algoritmos para estimação dos parâmetros desse modelo e avaliamos seus desempenhos em simulações.

No Capítulo 8, apresentamos as conclusões da tese, listamos suas contribuições e enunciamos os desdobramentos futuros desse trabalho.

Capítulo 2

Inferência Bayesiana

Frequentemente, enfrentamos situações de incerteza nas quais precisamos tomar decisões, sejam elas simples como se devemos sair de casa com o guarda-chuva ou complexas como qual profissão escolher. Ainda que inconscientemente, estimamos a probabilidade de eventos futuros, avaliamos riscos e fazemos previsões. Inferência estatística é a área da matemática que nos ajuda a realizar essas tarefas de forma rigorosa, dando suporte, mas não necessariamente substituindo, a intuição e o senso comum.

Existem duas escolas principais de inferência estatística que diferem essencialmente na forma como o conceito de probabilidade é definido, e conseqüentemente na aplicabilidade de seus métodos. A escola frequentista apenas admite a probabilidade aplicada a eventos replicáveis, enquanto a escola bayesiana [9, 10] se fundamenta no conceito de probabilidade como uma medida para o grau de incerteza sobre possíveis eventos, sendo portanto subjetiva.

O primeiro matemático a adotar a visão bayesiana foi o reverendo Thomas Bayes, primeiro a descobrir, em 1740, um caso particular do teorema que leva seu nome. O teorema fornece uma relação simples entre a probabilidade prévia de um determinado evento com a probabilidade desse mesmo evento após novas evidências serem adquiridas. Negligenciado durante a vida de Bayes, foi redescoberto independentemente três décadas mais tarde por Pierre Laplace, que lhe deu uma nova formulação e o aplicou em inúmeros problemas práticos da época. Não obstante o sucesso, o próprio Laplace se converteu ao frequentismo no final de sua vida.

Condenada por estatísticos do calibre de Ronald Fisher como anticientífica

devido a sua natureza assumidamente subjetiva, a visão bayesiana continuou sendo usada durante a maior parte do século XX, pois fornecia soluções de problemas para os quais a visão frequentista falhava. Avanços teóricos feitos por matemáticos como Jeffrey, De Finet e outros tornaram o bayesianismo mais respeitável, mas sua aceitação definitiva e popularização só vieram no fim da década de 80, quando o surgimento de métodos numéricos baseados em MCMC permitiram a aplicação da ferramenta em inúmeros problemas que antes eram intratáveis devido ao alto custo computacional. Atualmente, métodos bayesianos estão presentes em aplicações como aprendizado por máquina, genética, filtragem anti-*spam*, investigação forense e predição de resultados de eleições, entre muitas outras.

Para este trabalho, algumas características do paradigma bayesiano se mostraram importantes. A principal delas é a possibilidade de se incorporar a experiência subjetiva do projetista aos modelos, algo proibido pela ótica frequentista. Num sistema de restauração, o conhecimento sobre o comportamento típico de um defeito, usualmente resultado de experiências pessoais do projetista, é de grande valia para o desempenho do sistema. Além disso, como veremos na Seção 2.5, a abordagem bayesiana permite que parâmetros indesejados possam ser marginalizados (integrados) e retirados da análise, tornando desnecessário estimá-los previamente de forma sub-ótima.

Este capítulo se destina à apresentação de conceitos básicos de inferência bayesiana, com foco naqueles que consideramos importantes para o entendimento do restante do texto. Começamos com uma discussão sobre três importantes interpretações para o conceito de probabilidade: clássica, frequentista e subjetiva. Em seguida, apresentamos mais formalmente o conceito de inferência estatística, incluindo alguns exemplos simples de sua aplicação e diferenciando mais claramente as escolas bayesiana e frequentista. Na sequência, dedicamos seções para os conceitos de modelo hierárquico, distribuição *a priori* e eliminação de parâmetros. Concluímos o capítulo na Seção 2.6.

2.1 Algumas interpretações de probabilidade

Embora existam diversas interpretações para o conceito de probabilidade [11], três delas são mais relevantes em aplicações práticas: a clássica, a frequentista e a bayesiana (ou subjetiva). É importante ressaltar que todas as interpretações são compatíveis com a definição axiomática da teoria de probabilidade formalizada por Kolmogorov, sendo suas diferenças de caráter mais filosófico do que matemático.

Pela definição clássica de probabilidade, parte-se de um conjunto de eventos elementares igualmente prováveis dentro de um espaço amostral U , e associa-se a um evento qualquer $E \in U$ uma probabilidade calculada pela razão entre número de eventos elementares em E e o número de eventos elementares em U . Essa definição parece natural quando a simetria do experimento sugere que não há razão para que alguns eventos sejam mais prováveis que outros. Exemplos típicos desse tipo de experimento são o lançamento de um dado ou de uma moeda, a retirada de uma carta em um baralho ou de bolas numeradas em sorteios de loteria.

Outro conceito de probabilidade bastante usual é o frequentista, cujo nome se deve à noção de frequência relativa, na qual se baseia. Em vez de estabelecer previamente um conjunto de eventos possíveis, a visão frequentista considera o número de eventos efetivamente obtidos em experimentos supostamente repetidos em condições similares. Se o número de experimentos é suficientemente grande, a probabilidade de um evento é definida pela razão entre o número de vezes em que esse evento ocorreu e o número total de experimentos. Pela lei dos grandes números, essa razão deve se aproximar cada vez mais de um certo valor quanto maior for o número de experimentos.

Consideremos o experimento de lançamento de um dado. Pela simetria do experimento, um adepto da visão clássica diria que os seis resultados possíveis são igualmente prováveis, e que portanto a probabilidade de uma dada face ser obtida é de $1/6$. A probabilidade de o resultado ser um número par seria então igual à razão entre a quantidade de números pares entre 1 e 6 e a quantidade total resultados possíveis, ou seja $1/2$. Para um frequentista, é impossível determinar previamente as probabilidades envolvidas, pois estas só podem ser conhecidas após um grande número de experimentos. A probabilidade de sair um número par é obtida contando-se o número de vezes em que este resultado realmente foi obtido e dividindo-se o

resultado pelo número total de experimentos. Se o dado não for viciado, espera-se que essa razão coincida com obtido pela interpretação clássica. No entanto, para um frequentista, só se pode afirmar que o dado não é viciado após os experimentos terem sido realizados.

É comum, entretanto, associar probabilidades a eventos que não podem ser o resultado de experimento replicável e para os quais nem a noção clássica nem a frequentista fazem sentido. Exemplos seriam questões como a probabilidade de existência de vida fora da Terra, possíveis acontecimentos históricos e consequências de certas políticas públicas. Para permitir o tratamento matemático desse tipo de questão, a escola bayesiana de estatística adota a noção de probabilidade vista como “grau de confiança”, portanto subjetiva, sobre a veracidade de uma afirmação ou ocorrência de um evento.

A interpretação subjetiva tende a coincidir com uma das outras duas, dependendo do caso. Na ausência de dados experimentais, o estatístico pode usar considerações sobre a simetria do experimento para atribuir probabilidades subjetivas aos possíveis eventos, coincidindo assim com a visão clássica. No caso extremo oposto, quando o número de dados tende ao infinito, a probabilidade assumida inicialmente torna-se irrelevante, e a probabilidade bayesiana tenderá a coincidir com a frequentista.

2.2 Problema geral de inferência estatística e algumas soluções

Usamos inferência estatística quando desejamos obter informações sobre grandezas inobserváveis a partir de dados que podem ser descritos por modelos probabilísticos. Como um exemplo de Engenharia, podemos considerar o receptor de um sistema de comunicação que deve decidir os dados mais prováveis de terem sido emitidos no transmissor com base nos dados que chegam no receptor após sofrerem distorções pelo canal de comunicação. Similarmente, o assunto desta tese é o de estimar o sinal de áudio original com base em sua versão distorcida.

Para formalizar o conceito, seja x uma variável aleatória (possivelmente mul-

tidimensional) descrita pela distribuição¹ $p(\mathbf{x}; \boldsymbol{\theta})$, em que $\boldsymbol{\theta}$ é um vetor contendo parâmetros desconhecidos que definem certas características da distribuição, como por exemplo sua média e variância. O problema de inferência é extrair informações sobre os parâmetros $\boldsymbol{\theta}$ dispondo de um conjunto de dados \mathbf{x} .

A escola frequentista entende os parâmetros $\boldsymbol{\theta}$ como valores fixos porém desconhecidos, e toda inferência é feita com base na distribuição $p(\mathbf{x}; \boldsymbol{\theta})$. Por sua vez, a escola bayesiana enxerga os parâmetros como variáveis aleatórias com certas distribuições associadas. Antes de os dados serem observados, a distribuição associada aos parâmetros é chamada de distribuição *a priori* (ou simplesmente priori), e quantifica o grau de confiança do projetista sobre os possíveis valores dos parâmetros. Após os dados serem levados em consideração, a nova distribuição dos parâmetros é conhecida como distribuição *a posteriori* (ou simplesmente posteriori).

2.2.1 Teorema de Bayes

Através do teorema de Bayes, podemos obter a distribuição *a posteriori* a partir da distribuição *a priori*, levando em conta os dados observados. Seja $p(\boldsymbol{\theta}|H)$ a distribuição *a priori*, em que H (de história) representa todas as informações à disposição do usuário sobre os parâmetros de interesse antes de os dados estarem disponíveis. Pelo teorema de Bayes, podemos obter a distribuição *a posteriori* através de:

$$p(\boldsymbol{\theta}|\mathbf{x}, H) = \frac{p(\mathbf{x}|\boldsymbol{\theta}, H)p(\boldsymbol{\theta}|H)}{p(\mathbf{x}|H)}, \quad (2.1)$$

em que $p(\mathbf{x})$ é a densidade de probabilidade associada ao vetor de dados \mathbf{x} e na equação acima é um fator de proporcionalidade que garante que a integral do quociente sobre o conjunto dos números reais seja igual a 1, sendo dada por:

$$p(\mathbf{x}) = \int_{\boldsymbol{\theta}} p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (2.2)$$

Como os dados são supostos conhecidos, $p(\mathbf{x})$ é uma constante e não afeta

¹O nome “distribuição” de uma variável aleatória X se referirá tanto à densidade de probabilidade $p(x)$ (no caso de V.A. contínua) quanto à probabilidade $P(x)$ (no caso de V.A. discreta); quando for importante será feita a distinção clara disto. Mas haverá alguma liberdade de terminologia, por exemplo, ao dizer que a maximização da densidade de probabilidade maximiza a probabilidade.

a forma da função $p(\boldsymbol{\theta}|\mathbf{x}, H)$. Por essa razão é usual escrever o teorema de Bayes como uma relação de proporcionalidade, removendo $p(\mathbf{x})$:

$$p(\boldsymbol{\theta}|\mathbf{x}, H) \propto p(\mathbf{x}|\boldsymbol{\theta}, H)p(\boldsymbol{\theta}|H). \quad (2.3)$$

2.2.2 Alguns critérios de otimização

A *função de verossimilhança*, $l(\mathbf{x}; \boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})$ (omitiremos H daqui em diante) é interpretada como uma quantificação da probabilidade de os dados \mathbf{x} serem observados quando o vetor de parâmetros é $\boldsymbol{\theta}$. Na escola bayesiana, é o elemento de ligação entre a priori e a posteriori, enquanto na escola frequentista é usada isoladamente para inferência. O popular critério ML (*Maximum Likelihood*) [12, 13] estabelece que o valor estimado para $\boldsymbol{\theta}$ é aquele que maximiza a função de verossimilhança, ou seja, é o valor de $\boldsymbol{\theta}$ que faz os dados observados serem os mais prováveis.

A escola bayesiana também permite obter estimativas pontuais do parâmetro com base em sua distribuição posteriori, através de critérios de otimização como o BMSE (*Bayesian MSE*) e o MAP (*Maximum a Posteriori*) [13]. Pelo critério MAP o valor estimado para os parâmetros é aquele que maximiza a distribuição *a posteriori*, enquanto o BMSE resulta na média dessa distribuição. É importante ressaltar que podemos extrair muitas outras informações a partir da distribuição *a posteriori*, tais como valores esperados, distribuições marginais, intervalos de confiança, etc.

2.2.3 Exemplo

Para ilustrar algumas das soluções, consideremos o modelo linear geral, que aparece frequentemente em diversos problemas de processamento de sinais. Nesse modelo, o vetor de dados observados \mathbf{x} é obtido pela soma de um produto matricial e um ruído \mathbf{v} , ou seja, $\mathbf{x} = \mathbf{G}\boldsymbol{\theta} + \mathbf{v}$. Esse modelo pode descrever, por exemplo, um sinal formado por uma mistura de senoides com amplitudes desconhecidas imersas em ruído.

Usualmente \mathbf{v} pode ser descrito como ruído branco gaussiano de média zero e variância σ_v^2 , caso em que a verossimilhança é dada por [2]:

$$p(\mathbf{x}|\boldsymbol{\theta}) = p_v(\mathbf{x} - \mathbf{G}\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_v^2)^{N/2}} \exp \left\{ -\frac{1}{2\sigma_v^2} (\mathbf{x} - \mathbf{G}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{G}\boldsymbol{\theta}) \right\}. \quad (2.4)$$

A maximização dessa função em relação a $\boldsymbol{\theta}$ gera o estimador ML [2]:

$$\boldsymbol{\theta}^{\text{ML}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{x}. \quad (2.5)$$

A escola bayesiana resolveria esse mesmo problema tratando $\boldsymbol{\theta}$ como uma variável aleatória e lhe atribuindo uma distribuição *a priori*. Considerando, por exemplo, uma priori gaussiana com média \mathbf{m}_θ e matriz de covariância \mathbf{C}_θ , teríamos uma posteriori cuja maximização geraria o seguinte estimador para $\boldsymbol{\theta}$:

$$\mathbf{x}^{\text{MAP}} = (\mathbf{G}^T \mathbf{G} + \sigma_v^2 \mathbf{C}_\theta^{-1})^{-1} (\mathbf{G}^T \mathbf{x} + \sigma_v^2 \mathbf{C}_\theta^{-1} \mathbf{m}_\theta). \quad (2.6)$$

Podemos ver na equação acima que se os elementos de \mathbf{C}_θ forem altos, as parcelas introduzidas pela priori se tornam irrelevantes, e então a solução MAP se aproxima da ML. Isso é razoável, uma vez que valores altos para \mathbf{C}_θ indicam que a priori é bastante “espalhada” e todos os valores são quase igualmente prováveis.

2.3 Modelo bayesiano hierárquico

Em muitos problemas práticos, os dados observados dependem de diversos parâmetros, os quais podem apresentar uma relação hierárquica entre si. Por exemplo, num sinal contaminado com ruído impulsivo em discos de vinil, a taxa de amostras corrompidas p é uma variável desconhecida, podendo ser modelada probabilisticamente. Naturalmente, esperamos alguma dependência nos valores de p para discos com características parecidas. Podemos considerar, portanto, que os p para diferentes discos são gerados por uma distribuição comum. Os dados observados, no caso as amostras degradadas do sinal, dependeriam de uma instância particular dessa distribuição.

A Figura 2.1 ilustra esse conceito de forma mais geral. Nessa figura, diversos conjuntos de dados observados possíveis, x_1, x_2, \dots, x_n , dependem de parâmetros $\theta_1, \theta_2, \dots, \theta_n$, respectivamente, que por sua vez são instâncias de variáveis aleatórias descritas por parâmetros ϕ .

O teorema de Bayes permite obter a distribuição *a posteriori* dos parâmetros desconhecidos em função dos dados observados. No exemplo em questão, considere-

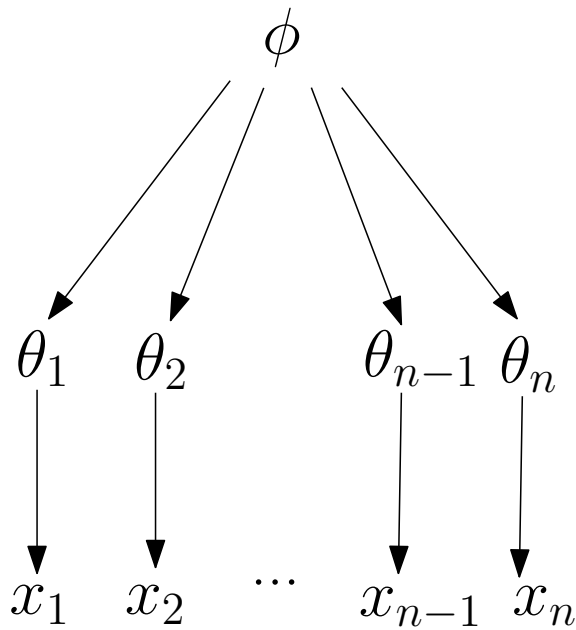


Figura 2.1: Exemplo de modelo bayesiano hierárquico. Os conjuntos de dados x_1, x_2, \dots, x_n , dependem de parâmetros $\theta_1, \theta_2, \dots, \theta_n$, respectivamente, que por sua vez são instâncias de variáveis aleatórias descritas por parâmetros ϕ .

mos a distribuição *a posteriori* $p(\boldsymbol{\theta}|\mathbf{x})$. A distribuição *a priori* de $\boldsymbol{\theta}$ pode depender de outros parâmetros, digamos ϕ , que são também desconhecidos e podem ter uma probabilidade *a priori* associada. Aplicando a regra de Bayes duas vezes, temos:

$$p(\boldsymbol{\theta}, \phi|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\theta}, \phi)p(\boldsymbol{\theta}|\phi)p(\phi). \quad (2.7)$$

2.4 Distribuição *a priori*

Como afirmamos anteriormente, a distribuição *a priori* é uma função que reflete o conhecimento do projetista sobre os possíveis valores dos parâmetros sem levar em consideração os dados. Embora seja possível utilizar critérios frequentistas para a escolha de $p(\boldsymbol{\theta})$, usualmente o conhecimento proveniente da experiência subjetiva do projetista é o fator preponderante para a escolha da probabilidade *a priori*. Diversas técnicas servem para guiar o projetista nessa tarefa, dentre as quais destacamos o método do histograma, o método da função de distribuição e o método de verossimilhança relativa [10]. Apesar de úteis, essas técnicas costumam gerar distribuições complicadas e de difícil tratamento analítico, o que prejudica o procedimento posterior de otimização.

2.4.1 Prioris conjugadas

Para garantir que a distribuição *a posteriori* seja analiticamente tratável, é comum o uso de prioris conjugadas, que são definidas como distribuições com estrutura algébrica similar à da verossimilhança. Isso garante que a posteriori seja do mesmo tipo da priori, mas com parâmetros modificados. Por exemplo, se a verossimilhança é gaussiana, a escolha de uma gaussiana como priori gera uma posteriori também gaussiana, que é uma distribuição com propriedades interessantes para otimização ou obtenção de amostras.

2.4.2 Priori não-informativa

Priori não-informativa, como o nome sugere, é uma distribuição que idealmente não contém qualquer informação sobre os possíveis valores dos parâmetros. A escolha aparentemente mais óbvia é uma distribuição uniforme, que associa igual probabilidade a todos os valores possíveis dos parâmetros, fazendo com que a posteriori coincida com a função de verossimilhança. No caso discreto, essa escolha não apresenta problemas; para variáveis contínuas, entretanto, uma distribuição uniforme é inadequada porque não é invariante em relação a uma transformação de variáveis um-para-um, isto é, se a distribuição de $\boldsymbol{\theta}$ é uniforme, a distribuição de $\boldsymbol{\phi} = f(\boldsymbol{\theta})$ será não-uniforme [10]. Esse resultado não é razoável, pois se não temos conhecimento sobre $\boldsymbol{\theta}$, o mesmo deve ocorrer com $\boldsymbol{\phi}$.

Para eliminar esse inconveniente, Jeffreys introduziu uma classe de distribuições invariantes a transformações um-para-um [10]. A priori de Jeffreys associada à variável $\boldsymbol{\theta}$ é dada por:

$$p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}, \quad (2.8)$$

onde $\mathbf{I}(\boldsymbol{\theta})$ é a informação de Fisher de $\boldsymbol{\theta}$, definida por:

$$I(\boldsymbol{\theta}) = E_{\mathbf{X}|\boldsymbol{\theta}} \left[-\frac{\partial^2 \log(p(\mathbf{X}|\boldsymbol{\theta}))}{\partial \boldsymbol{\theta}^2} \right], \quad (2.9)$$

que mostra que a priori depende apenas da verossimilhança.

No caso de modelos invariantes com relação à escala, como a distribuição do

desvio-padrão σ da gaussiana, por exemplo, obtém-se a seguinte *priori* de Jeffreys:

$$p(\sigma) \propto \frac{1}{\sigma}. \quad (2.10)$$

Há outras alternativas para a escolha de prioris não-informativas. Jaynes [14] propôs que o critério de máxima entropia fosse usado na especificação da distribuição *a priori*. Bernardo [15] também defende que a escolha da priori não-informativa deva ser feita através de critérios baseados na teoria da informação.

Na maioria dos casos, a distribuição gerada é imprópria, isto é, sua integral é infinita, como na Equação (2.10). Esse fato nem sempre é um problema, visto que estamos interessados, muitas vezes, na forma da distribuição, e não nos seus valores exatos.

2.5 Eliminação de parâmetros

É comum que o vetor de parâmetros contenha elementos que não são de interesse para estimação. Nesse caso, podemos integrar esses parâmetros indesejados (*nuisance parameters*) e então trabalhar apenas com os de interesse. Se θ for particionado entre os parâmetros indesejados e os de interesse, de tal forma que $\theta = (\phi, \psi)$, a posteriori para os parâmetros de interesse é:

$$p(\phi) = \int_{\psi} p(\phi, \psi | \mathbf{x}) d\psi. \quad (2.11)$$

Essa operação nem sempre pode ser feita analiticamente, especialmente nos casos em que a distribuição é multivariável e não possui uma forma conhecida. Nesses casos, os algoritmos que serão vistos no próximo capítulo são indicados para a computação numérica da integral acima.

2.6 Comentários conclusivos

Neste capítulo vimos os conceitos mais importantes da teoria de inferência bayesiana. Embora seja possível resolver problemas práticos apenas com esses conceitos, em boa

parte das aplicações as distribuições envolvidas são multivariáveis e multi-modais e os problemas de otimização que surgem naturalmente são difíceis de ser resolvidos pelas técnicas clássicas de otimização. No capítulo seguinte, veremos alguns métodos e algoritmos indicados para essa tarefa.

Capítulo 3

Técnicas Numéricas de Estimação

A modelagem realista de sistemas físicos muitas vezes exige a utilização de modelos hierárquicos sofisticados, com distribuições complexas e intratáveis analiticamente. Para lidar com esses problemas de estimação, foram propostas diversas técnicas numéricas, dentre as quais destacam-se o algoritmo *Expectation-Maximization* (EM), o amostrador de Gibbs e os algoritmos de Metropolis-Hastings (MH) e de Metropolis-Hastings com saltos reversíveis. Com exceção do primeiro, esses algoritmos se baseiam em técnicas de Monte Carlo via Cadeias de Markov (*Markov-Chain Monte Carlo*, MCMC). Em geral, essas técnicas consistem no projeto de uma cadeia de Markov para obtenção de amostras de uma certa distribuição, e posteriormente na utilização de técnicas de Monte Carlo para obter informações de interesse com base nas amostras geradas. Este capítulo dedica-se à descrição de alguns desses métodos.

Começamos apresentando a idéia geral por trás de técnicas de Monte Carlo. Em seguida, descrevemos o método da rejeição, uma técnica simples usada para obtenção de amostras de uma distribuição, e que permitirá uma compreensão intuitiva dos algoritmos mais sofisticados que serão descritos mais adiante. Prosseguimos com a apresentação da teoria de Cadeias de Markov (MC), base para os algoritmos que serão descritos em seguida. Nas Seções 3.4 e 3.5 apresentamos os algoritmos de amostragem de Gibbs e Metropolis-Hastings, dois dos mais populares algoritmos do tipo MCMC. O capítulo termina com a apresentação de algumas técnicas simples para diagnóstico de convergência dos algoritmos.

Optamos por descrever os algoritmos de forma sucinta, nos moldes do trabalho de Andrieu et. al [16] e de Neal [17]. Como referência mais completa sobre o

assunto, recomendamos os livros de Gamerman [18] e Casella [19].

3.1 Técnicas de Monte Carlo

Técnicas de Monte Carlo consistem em gerar amostras i.i.d. de uma certa distribuição e em seguida usá-las para obter uma aproximação de alguma característica da distribuição difícil de obter analiticamente. Mais precisamente, de posse de um conjunto de amostras $X = \{x^{(1)}, \dots, x^{(N)}\}$ de uma certa distribuição $p(x)$, podemos aproximar uma integral do tipo:

$$I(f) = \int_x f(x)p(x)dx \quad (3.1)$$

através de um somatório:

$$I_N(f) = \frac{1}{N} \sum_{i=1}^N f(x^{(i)}). \quad (3.2)$$

É possível mostrar que o estimador acima é não polarizado e, pela lei dos grandes números, converge para a integral da Equação (3.1) *quase certamente* (isto é, com probabilidade 1) quando N tende a infinito.

Obter amostras de uma distribuição diretamente nem sempre é trivial. As técnicas descritas nas próximas seções são formas de realizar essa tarefa indiretamente.

3.2 Método da Rejeição

O método da rejeição é usado quando se deseja obter amostras de uma distribuição complexa usando uma distribuição auxiliar supostamente fácil de se amostrar. Sejam $\pi(x)$ a distribuição de interesse e $q(x)$ uma distribuição que, para algum A , satisfaz $Aq(x) \geq \pi(x), \forall x$. O método da rejeição consiste em se gerar uma amostra de $q(x)$, digamos x^* , e aceitá-la com probabilidade $\pi(x^*)/(Aq(x^*))$. Na prática, obtém-se uma amostra u de uma distribuição uniforme entre 0 e 1 e aceita-se x^* se $u \leq \pi(x^*)/(Aq(x^*))$.

A eficiência do método depende do valor de A , uma medida da “proximidade” entre $q(x)$ e $\pi(x)$. Quanto maior o valor de A , maior é o percentual de amostras

rejeitadas e consequentemente maior o número de iterações necessárias para se obter o conjunto de amostras exigidas para a estimação de Monte Carlo.

A principal desvantagem desse método é a dificuldade, em certos casos, de escolher uma distribuição auxiliar que leve a uma probabilidade de aceitação não muito baixa. Para distribuições multivariáveis essa tarefa é particularmente difícil, e nesses casos os algoritmos baseados em cadeias de Markov são a escolha mais indicada.

3.3 Cadeias de Markov

A teoria de Cadeias de Markov é a base para os algoritmos MCMC que serão descritos mais adiante. Nesta seção, apresentaremos uma breve descrição da teoria e enunciaremos algumas propriedades importantes para o desenvolvimento e análise de algoritmos MCMC. Começamos considerando cadeias de Markov com espaço de estados discreto e em seguida estenderemos os resultados para o caso contínuo.

3.3.1 Conceituação

Uma Cadeia de Markov é um processo aleatório discreto no tempo que apresenta a *propriedade de Markov*, que estabelece que o estado atual da cadeia depende apenas do estado imediatamente anterior. Mais formalmente, seja $X^{(n)}$ a variável aleatória que representa o estado da cadeia no instante n e seja S o espaço de estados para as variáveis $X^{(n)}$. Então:

$$P(X^{(n)} \in A^{(n)} | X^{(n-1)} \in A^{(n-1)}, \dots, X^{(0)} \in A^{(0)}) = P(X^{(n)} \in A^{(n)} | X^{(n-1)} \in A^{(n-1)}), \quad (3.3)$$

para quaisquer $A^{(0)}, \dots, A^{(n)} \in S$.

No caso em que S é um conjunto contável (tipicamente, mas não necessariamente, finito), a cadeia de Markov é dita discreta. Consideremos o caso finito com espaço de estados $S = \{s_1, \dots, s_N\}$. As probabilidades de transição de um estado da cadeia para outro, no instante n , definem a *matriz de transição* \mathbf{T}_n , cujo elemento

da linha i e da coluna j é dado por:

$$T_n(i|j) = P(X^{(n)} = s_i | X^{(n-1)} = s_j). \quad (3.4)$$

Como cada coluna dessa matriz contém a probabilidade associada a cada um dos elementos do conjunto amostral, a soma dos elementos em cada coluna é igual a 1. Matrizes com essa propriedade são chamadas matrizes estocásticas [18] e possuem uma série de propriedades importantes para a análise de algoritmos baseados em MCMC. Uma delas é a existência de pelo menos um autovalor igual a 1. Além disso, no caso em que todos os elementos de \mathbf{T}_n são maiores que zero (caso em que é sempre possível passar de um estado qualquer para outro qualquer), pode-se mostrar que todos os demais autovalores são distintos e menores que 1.

A distribuição de probabilidade no instante n , denotada por $P_n(i)$, definida para $i = \{1, \dots, N\}$, pode ser obtida a partir da distribuição de probabilidade no instante $(n - 1)$, através de:

$$P_n(i) = \sum_{j=1}^N T_n(i|j)P_{n-1}(j), \quad (3.5)$$

que pode ser escrita vetorialmente como:

$$\mathbf{P}_n = \mathbf{T}_n \mathbf{P}_{n-1}, \quad (3.6)$$

em que $\mathbf{P}_n = [P_n(1) \dots P_n(N)]^T$.

Se \mathbf{T}_n é independente de n , dizemos que a cadeia de Markov é *homogênea* e a matriz de transição passa a ser denotada simplesmente por \mathbf{T} . Nesse caso, a aplicação da Equação (3.6) n vezes fornece:

$$\mathbf{P}_n = \mathbf{T}^n \mathbf{P}_0, \quad (3.7)$$

em que \mathbf{P}_0 é a distribuição do estado inicial da cadeia.

Para o desenvolvimento de algoritmos MCMC, a Cadeia de Markov deve possuir as duas propriedades abaixo:

- Irredutibilidade: Partindo de qualquer estado, existir uma probabilidade não-nula de a cadeia mover-se para qualquer outro estado em um número finito de passos. Isso equivale a termos $T^n(i|j) > 0$, para algum n , de forma que a probabilidade de transição em n passos do estado s_i para o estado s_j é diferente de zero.
- Aperiodicidade: A cadeia não ficar presa em ciclos.

3.3.2 Distribuição invariante

Uma distribuição é dita invariante (ou estacionária) se permanece fixa sob a aplicação da matriz de transição. No desenvolvimento de algoritmos MCMC, estamos interessados em construir cadeias de Markov que façam com que uma certa distribuição seja invariante.

Denotando a distribuição invariante por $\pi(i)$, devemos ter:

$$\pi(i) = \sum_{j=1}^N T(i|j)\pi(j). \quad (3.8)$$

Vetorialmente, podemos escrever $\boldsymbol{\pi} = \mathbf{T}\boldsymbol{\pi}$, donde vemos que $\boldsymbol{\pi}$ é um autovetor associado ao autovalor $\lambda = 1$. Para se determinar $\boldsymbol{\pi}$ unicamente, impõe-se a condição de que a soma de seus elementos seja igual a 1.

Para algoritmos MCMC, é usual impor a condição *detailed balance* para assegurar que uma dada distribuição seja invariante. Essa condição estabelece que a probabilidade de a cadeia passar de um estado s_i no instante $(n - 1)$ para o estado s_j no instante (n) é igual à probabilidade de a transição inversa ocorrer, isto é:

$$\pi(j)T(i|j) = \pi(i)T(j|i). \quad (3.9)$$

Para ver que $\pi(i)$ é distribuição invariante, basta somar os dois membros da igualdade acima para todos os valores possíveis de j , e perceber que o resultado é a Equação (3.8). Embora mais restritiva que a Equação (3.8), essa condição é mais simples de se impor para algoritmos MCMC.

3.3.3 Ergodicidade

Além de garantir que $\pi(i)$ seja a distribuição invariante desejada, devemos assegurar que $P_n(i)$ convirja para $\pi(i)$ quando n tender a infinito, qualquer que seja a distribuição inicial $P_0(i)$. Nesse caso, dizemos que $\pi(i)$ é a distribuição-limite da cadeia.

Essa propriedade é conhecida como *ergodicidade*. Para garantir que a cadeia de Markov seja ergódica, é necessário que a cadeia seja aperiódica e irredutível. No caso discreto, basta termos os autovalores de \mathbf{T} todos distintos, podendo assim escrever a distribuição inicial usando os autovetores de \mathbf{T} como base e em seguida calcular a expressão $\mathbf{T}^n \mathbf{P}_0$ [17]:

$$\mathbf{P}_0 = \boldsymbol{\pi} + c_2 \mathbf{v}_2 + \dots + c_N \mathbf{v}_N; \quad (3.10)$$

$$\mathbf{P}_n = \mathbf{T}^n \mathbf{P}_0 = \boldsymbol{\pi} + c_2 \lambda_2^n \mathbf{v}_2 + \dots + c_N \lambda_N^n \mathbf{v}_N. \quad (3.11)$$

Na primeira equação usamos o fato de que $\boldsymbol{\pi}$ é o autovetor associado ao autovalor 1. Como sabemos que todos os demais autovalores são menores que 1, concluímos que:

$$\lim_{n \rightarrow \infty} \mathbf{P}_n = \lim_{n \rightarrow \infty} \mathbf{T}^n \mathbf{P}_0 = \boldsymbol{\pi}. \quad (3.12)$$

3.3.4 Cadeias de Markov para espaço de estados contínuo

Para o caso em que o espaço de estados S é contínuo, as propriedades descritas nas seções anteriores são expressas através de densidades de probabilidade. A propriedade de Markov é definida por:

$$p(x^{(n)} | x^{(n-1)}, \dots, x^{(0)}) = p(x^{(n)} | x^{(n-1)}). \quad (3.13)$$

O *núcleo de transição* $K_n(x|y)$ no instante n é definido por:

$$K_n(x|y) = p_{X^{(n)}}(x | X^{(n-1)} = y), \quad (3.14)$$

em que $p_{X^{(n)}}(x)$ denota a densidade de probabilidade da variável aleatória $X^{(n)}$.

Assim, a distribuição do estado da cadeia no instante n é dada por:

$$p_n(x) = \int_{y \in S} K_n(x|y)p_{n-1}(y)dy. \quad (3.15)$$

No caso de a cadeia ser homogênea, K_n independe de n e a condição *detailed balance* se torna:

$$\int_A \int_B K(x|y)\pi(y)dydx = \int_B \int_A K(y|x)\pi(x)dx dy, \quad (3.16)$$

para quaisquer conjuntos A e B pertencentes a S .

Como no caso discreto, a condição *detailed balance* é suficiente para que $\pi(i)$ seja uma distribuição invariante da cadeia de Markov definida por $K(i|j)$. Para garantir que a distribuição invariante seja também a distribuição-limite, a cadeia deve ser aperiódica e irredutível.

3.4 Amostrador de Gibbs

O amostrador de Gibbs (*Gibbs Sampler*) [20] é um caso especial de MCMC, indicado para os casos em que a distribuição conjunta é mais difícil de amostrar do que as condicionais. A técnica consiste em particionar a variável conjunta em diversos componentes (possivelmente multivariáveis) e obter amostras das distribuições condicionais de cada componente, considerando os demais fixos. O processo é repetido usando os últimos valores amostrados de cada componente como condicionantes da distribuição dos demais componentes.

Seja $\pi(\boldsymbol{\theta})$ a distribuição conjunta da qual se deseja obter amostras. A variável $\boldsymbol{\theta}$ é então particionada em k componentes, de tal forma que $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k\}$. A i -ésima iteração do amostrador de Gibbs pode ser expressa como:

$$\boldsymbol{\theta}_1^{(i)} \sim \pi(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) \quad (3.17)$$

$$\boldsymbol{\theta}_2^{(i)} \sim \pi(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1^{(i)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) \quad (3.18)$$

⋮

$$\boldsymbol{\theta}_k^{(i)} \sim \pi(\boldsymbol{\theta}_k | \boldsymbol{\theta}_1^{(i)}, \dots, \boldsymbol{\theta}_{k-1}^{(i)}), \quad (3.19)$$

em que o símbolo \sim indica que a variável da esquerda é uma amostra da distribuição à direita.

Para ver que $\pi(\boldsymbol{\theta})$ é uma distribuição invariante em cada operação acima, vamos calcular a distribuição de $\boldsymbol{\theta}$ após a primeira operação, definida na Equação (3.17). Supondo que a distribuição da cadeia de Markov no final da iteração $(i-1)$ é $\pi(\boldsymbol{\theta})$, isto é: $p(\boldsymbol{\theta}^{(i-1)}) = \pi(\boldsymbol{\theta}^{(i-1)})$, obtemos:

$$\begin{aligned} p(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) &= p(\boldsymbol{\theta}_1^{(i)} | \boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) p(\boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) \\ &= \pi(\boldsymbol{\theta}_1^{(i)} | \boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) \pi(\boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}) \\ &= \pi(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i-1)}, \dots, \boldsymbol{\theta}_k^{(i-1)}). \end{aligned} \quad (3.20)$$

Logo, a transição gerada pelo amostrador de Gibbs preserva a distribuição de $\boldsymbol{\theta}$. O mesmo raciocínio aplicado às operações seguintes indica que, a cada amostragem, a distribuição resultante permanece igual a $\pi(\boldsymbol{\theta})$. Sob condições fáceis de se obter na prática, mostra-se que a cadeia é ergódica, isto é, após a convergência, as amostras geradas em cada iteração correspondem a amostras da distribuição conjunta $\pi(\boldsymbol{\theta})$.

3.5 Algoritmo de Metropolis-Hastings

Nem sempre as distribuições condicionais necessárias para o amostrador de Gibbs são fáceis de se obter. Nesses casos, o algoritmo de Metropolis-Hastings (MH) é mais indicado. A ideia do algoritmo é semelhante à do método da rejeição, descrito na Seção 3.2. Obtêm-se amostras de uma distribuição auxiliar (supostamente mais simples que a distribuição de interesse) e decide-se aceitar ou rejeitar essa amostra dependendo de algum critério.

Pelo algoritmo MH, uma amostra x^* é obtida a partir de uma distribuição proposta, designada por $q(x^* | x^{(i)})$, em que $x^{(i)}$ é o estado atual da cadeia de Markov. Essa amostra é aceita com uma probabilidade α dada por:

$$\alpha(x^{(i)}, x^*) = \min \left(1, \frac{\pi(x^*) q(x^{(i)} | x^*)}{\pi(x^{(i)}) q(x^* | x^{(i)})} \right). \quad (3.21)$$

Se a amostra gerada for aceita, o novo estado da cadeia é $x^{i+1} = x^*$; em caso contrário, a cadeia permanece no seu estado atual, isto é, $x^{(i+1)} = x^{(i)}$. O núcleo de transição é dado por:

$$K(x^{(i+1)}|x^{(i)}) = q(x^{(i+1)}|x^{(i)})\alpha(x^{(i)}, x^{(i+1)}) + \delta_{x^{(i)}}(x^{(i+1)})r(x^{(i)}), \quad (3.22)$$

em que

$$r(x^{(i)}) = \int_{x^* \in S} q(x^*|x^{(i)}) (1 - \alpha(x^{(i)}, x^*)) dx^* \quad (3.23)$$

é a probabilidade de a cadeia permanecer no estado atual. A expressão $\delta_{x^{(i)}}(x^{(i+1)})$ é a função impulso unitário (delta de Dirac) localizada em $x^{(i)}$ e aplicada em $x^{(i+1)}$, que indica uma distribuição “concentrada” em $x^{(i)}$.

Podemos verificar que esse núcleo de transição satisfaz a condição *detailed balance*, e portanto $\pi(x)$ é uma distribuição invariante da cadeia. Para garantir que $\pi(x)$ é também a distribuição-limite, temos que verificar a irreduzibilidade e aperiódicidade da cadeia. Como o algoritmo sempre permite a rejeição, segue que a cadeia é aperiódica. Para assegurar irreduzibilidade, é preciso que o suporte de $q(\cdot)$ inclua o suporte de $\pi(\cdot)$ [16].

A eficiência do algoritmo MH depende fundamentalmente da escolha da proposta, $q(\cdot)$. É usual escolher como proposta uma gaussiana centrada no estado atual, isto é, $q(x^*|x^{(i)}) = N(x^*|x^{(i)}, \sigma_q^2 I)$. A escolha da variância σ_q^2 é crucial. Se $q(\cdot)$ é muito estreita, apenas estados próximos ao máximo de $\pi(x)$ são visitados. Por outro lado, se $q(\cdot)$ é muito ampla, o percentual de amostras rejeitadas é muito alto e, conseqüentemente, as amostras geradas serão altamente correlacionadas entre si, invalidando a hipótese de independência das amostras, necessária para a estimação de Monte Carlo. Para resolver esse problema, recomenda-se descartar uma fração das amostras, reduzindo assim a correlação entre elas, mas, por outro lado, aumentando o tempo para convergência.

3.6 Análise de convergência

Os algoritmos apresentados até agora garantem que os estados da cadeia de Markov sejam amostras da distribuição-limite quando o número de iterações tende a infinito.

Como na prática precisamos usar um número finito de iterações, é importante saber quando a cadeia está suficientemente próxima da distribuição limite para que suas amostras possam ser usadas para estimação de Monte Carlo.

Muito trabalho tem sido desenvolvido em torno desse problema. Métodos teóricos para diagnóstico de convergência têm sido propostos, mas os resultados tiveram até o momento pouco impacto prático [18]. Na maioria das aplicações, usam-se mais frequentemente métodos informais de convergência, que consistem de análise estatística dos dados gerados pela cadeia. Embora simples, essas técnicas não permitem garantir a convergência de forma geral.

Dentre os métodos informais, três abordagens podem ser usadas. A primeira baseia-se na realização de n cadeias paralelas. Calcula-se o histograma de n amostras na m -ésima iteração e se o compara com o histograma obtido k iterações adiante. Se os histogramas forem bastante similares, considera-se que a convergência foi atingida. O valor de k deve ser grande o bastante para evitar que a correlação entre os estados sucessivos da cadeia causem uma falsa impressão de similaridade entre os histogramas. O critério para comparação dos histogramas é flexível, sendo uma possibilidade a divergência de Kullback-Leibler.

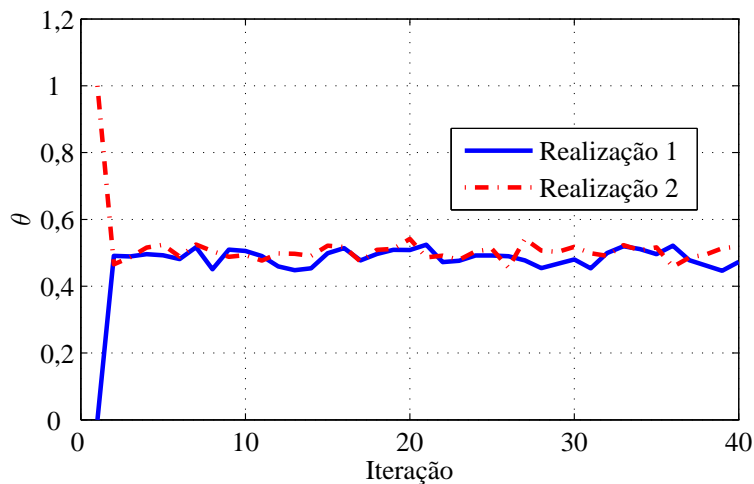


Figura 3.1: Evolução de um parâmetro de uma cadeia de Markov a partir de diferentes valores iniciais. Vemos que as amostras tendem a variar na mesma região a partir de certo ponto, independente da inicialização, evidenciando a convergência do algoritmo.

Outro método baseia-se na construção de uma única cadeia e no cálculo da média ergódica das amostras obtidas. Espera-se que após a convergência, a média se

aproxime de um certo valor constante. Portanto, através da análise visual do gráfico das médias, é possível diagnosticar a convergência.

A terceira técnica consiste na análise visual da evolução dos parâmetros da cadeia. Após a convergência, esses parâmetros devem exibir um padrão constante, que pode ser usado para diagnosticá-la. Analisando a Figura 3.1, que mostra a evolução de um parâmetro de uma cadeia de Markov partindo de pontos iniciais diferentes, podemos concluir que a convergência ocorre em torno da iteração 90.

3.7 Conclusões

Esse capítulo fez uma breve exposição teórica de técnicas numéricas indicadas para exploração de distribuições multidimensionais de formato complicado (por exemplo, multimodal). Em particular, o amostrador de Gibbs e os algoritmos de Metropolis-Hastings foram descritos com algum detalhe.

Capítulo 4

Modelagem de Sinais de Áudio com Distorções Não-Lineares

Sistemas lineares são aqueles que satisfazem as propriedades de escalamento e superposição, isto é, a resposta de um sistema linear F a uma combinação linear de entradas x_1 e x_2 é gerada da seguinte forma:

$$F\{a_1x_1 + a_2x_2\} = a_1F\{x_1\} + a_2F\{x_2\}. \quad (4.1)$$

Um sinal de áudio corrompido com distorções não-lineares pode ser caracterizado pelo diagrama de blocos da Figura 4.1, em que o bloco à esquerda representa o sistema linear variante no tempo que descreve o processo de geração do sinal de áudio a partir de uma excitação aleatória $e(n)$, e o bloco à direita modela a distorção não-linear, que é usualmente invariante no tempo.

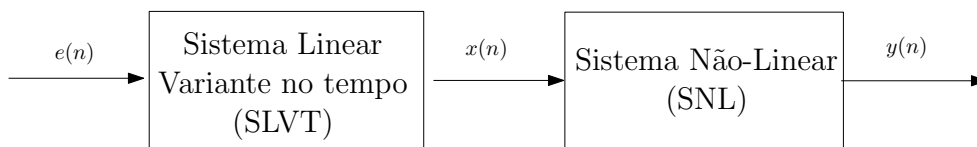


Figura 4.1: Modelo de sinal de áudio com distorções não-lineares.

Nas próximas seções, vamos discutir a modelagem dos dois componentes desse sistema do ponto-de-vista bayesiano. Descrevemos um método simples baseado no amostrador de Gibbs para estimação dos parâmetros do modelo AR que será usado extensivamente no restante desta tese. Além disso, mencionaremos os trabalhos

anteriores em modelagem de distorções não-lineares em áudio, e daremos uma visão geral das soluções propostas nos próximos capítulos.

4.1 Modelagem de sinais de áudio através de modelo autorregressivo

Trechos curtos de sinais de áudio podem ser satisfatoriamente descritos por um modelo autorregressivo, que interpreta a amostra atual $x(n)$ como uma combinação linear de amostras passadas somada a um erro de predição aleatório $e(n)$:

$$x(n) = \sum_{i=1}^P a_i x(n-i) + e(n). \quad (4.2)$$

No contexto de processamento de sinais, tal modelo representa um filtro só-pólos $A(z) = (1 - \sum_{i=1}^P a_i z^{-i})^{-1}$ excitado por ruído branco gaussiano de potência σ_e^2 . O vetor $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_P]^T$ definem os coeficientes do filtro só-pólos de ordem P .

Como, pela Equação (4.2), $e(n) = x(n) - \sum_{i=1}^P a_i x(n-i)$, a distribuição de $x(n)$, condicionada às P amostras anteriores de $x(n)$, é a distribuição de $e(n)$ com média deslocada de $\sum_{i=1}^P a_i x(n-i)$, o que resulta em:

$$p(x(n)|x(n-1), x(n-2), \dots, x(n-P)) = p_e \left(x(n) - \sum_{i=1}^P a_i x(n-i) \right). \quad (4.3)$$

Como a excitação $\mathbf{e} = [e(1) \ e(2) \ \dots \ e(N)]^T$ é, por hipótese, composta de amostras estatisticamente independentes, podemos obter a distribuição conjunta das amostras $\mathbf{x} = [x(P+1) \ x(P+2) \ \dots \ x(N)]^T$, sendo N o tamanho de um bloco do sinal, condicionada às P amostras iniciais $\mathbf{x}_0 = [x(1) \ x(2) \ \dots \ x(P)]^T$:

$$p(x(P+1), \dots, x(N)|x(1), \dots, x(P)) = \prod_{n=P+1}^N p_e \left(x(n) - \sum_{i=1}^P a_i x(n-i) \right). \quad (4.4)$$

Usualmente assume-se $e(n)$ branco, gaussiano, com variância σ_e^2 desconhe-

cida, ou seja, a distribuição de \mathbf{e} é dada por:

$$p(\mathbf{e}|\sigma_e^2) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}} \right)^N \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{n=1}^N \mathbf{e}^T \mathbf{e} \right\}. \quad (4.5)$$

Podemos reescrever a Equação (4.2) em forma vetorial:

$$\mathbf{x} = \mathbf{X}\mathbf{a} + \mathbf{e}, \quad (4.6)$$

em que a matriz \mathbf{X} é dada por:

$$\mathbf{X} = \begin{bmatrix} x(P) & x(P-1) & \dots & x(2) & x(1) \\ x(P+1) & x(P) & \dots & x(3) & x(2) \\ \vdots & & \ddots & & \vdots \\ x(N-2) & x(N-3) & \dots & x(N-P) & x(N-P-1) \\ x(N-1) & x(N-2) & \dots & x(N-P+1) & x(N-P) \end{bmatrix}. \quad (4.7)$$

A Equação (4.4) pode então ser reescrita como:

$$p(\mathbf{x}|\mathbf{x}_0, \mathbf{a}, \sigma_e^2) = p_e(\mathbf{x} - \mathbf{X}\mathbf{a}). \quad (4.8)$$

Substituindo a expressão acima em (4.5), obtemos:

$$p(\mathbf{x}|\mathbf{x}_0, \mathbf{a}, \sigma_e^2) = \frac{1}{(2\pi\sigma_e^2)^{\frac{N-P}{2}}} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{x} - \mathbf{X}\mathbf{a})^T (\mathbf{x} - \mathbf{X}\mathbf{a}) \right\}. \quad (4.9)$$

Se o vetor \mathbf{x} for conhecido, a maximização da expressão em (4.9) fornece o estimador de máxima verossimilhança (ML) para os coeficientes \mathbf{a} :

$$\mathbf{a}^{\text{ML}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}. \quad (4.10)$$

A Equação (4.8) pode ser escrita num formato equivalente, que será útil na derivação de alguns algoritmos desta tese. Definindo a matriz \mathbf{A} de $(N-P)$ linhas

e N colunas por:

$$\mathbf{A} = \begin{bmatrix} -a_P & \dots & a_1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -a_P & \dots & -a_1 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \dots & 0 & 0 & -a_P & \dots & -a_1 & 1 & 0 & 0 \\ 0 & \dots & 0 & 0 & -a_P & \dots & -a_1 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & -a_P & \dots & -a_1 & 1 \end{bmatrix}, \quad (4.11)$$

podemos escrever

$$\mathbf{e} = \mathbf{A} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x} \end{bmatrix}, \quad (4.12)$$

e a distribuição condicional de \mathbf{x} se torna:

$$p(\mathbf{x}|\mathbf{x}_0, \mathbf{a}, \sigma_e^2) = \frac{1}{(2\pi\sigma_e^2)^{\frac{(N-P)}{2}}} \exp \left\{ -\frac{1}{2\sigma_e^2} \begin{bmatrix} \mathbf{x}_0^T & \mathbf{x}^T \end{bmatrix} \mathbf{A}^T \mathbf{A} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x} \end{bmatrix} \right\}. \quad (4.13)$$

Num processamento feito bloco a bloco, pode-se considerar conhecidas as P primeiras amostras, obtidas através do processamento do bloco anterior, e usar a Equação (4.13) para obtenção da verossilhança. Se isso não puder ser feito, usa-se a aproximação:

$$p(\mathbf{x}, \mathbf{x}_0|\mathbf{a}) \approx p(\mathbf{x}|\mathbf{x}_0, \mathbf{a}), \quad (4.14)$$

que será tão mais precisa quanto maior for a diferença entre N e P . No restante desta tese, vamos assumir que \mathbf{x}_0 é conhecido, a menos que afirmado o contrário. Para simplificar a notação, vamos omitir essa variável no argumento das distribuições condicionais.

4.1.1 Estimação dos parâmetros do modelo AR

Diversos algoritmos podem ser usados para estimação dos coeficientes do modelo AR com base no sinal \mathbf{x} , sendo um deles o estimador de mínimos quadrados da Equação (4.10). Nesta seção apresentamos um algoritmo baseado no amostrador de Gibbs para estimação desses parâmetros.

O problema consiste na obtenção de amostras da distribuição *a posteriori*

conjunta de \mathbf{a} e σ_e^2 , dado \mathbf{x} , isto é $p(\mathbf{a}, \sigma_e^2 | \mathbf{x})$. Como vimos na Seção 3.4, pelo amostrador de Gibbs, tal tarefa pode ser feita através do procedimento a seguir:

1. Inicialização do parâmetro $\mathbf{a}^{(0)}$
2. Para j de 1 até N_{it} , repita:

$$\begin{aligned}\sigma_e^2(j) &\sim p(\sigma_e^2 | \mathbf{a}^{(j-1)}, \mathbf{x}), \\ \mathbf{a}^{(j)} &\sim p(\mathbf{a} | \sigma_e^2(j), \mathbf{x}).\end{aligned}$$

Em princípio o algoritmo converge qualquer que seja o valor inicial de \mathbf{a} , mas espera-se que a convergência seja mais rápida se o algoritmo iniciar com uma estimativa de boa qualidade, como por exemplo o estimador ML da Equação (4.10).

Para obtenção das distribuições condicionais acima, precisamos antes especificar as distribuições *a priori* para os parâmetros \mathbf{a} e σ_e^2 . Como feito em [2], assumimos esses parâmetros independentes *a priori* e escolhemos a distribuição imprópria $p(\mathbf{a}, \sigma_e^2) = p(\mathbf{a})p(\sigma_e^2) \propto \text{IG}(\sigma_e^2 | \alpha, \beta)$, com $\alpha = \beta = 10^{-10}$ (em que IG denota a distribuição gama-invertida [10]). Na prática, a quantidade de amostras em \mathbf{x} deverá ser grande o suficiente para que os parâmetros sejam bem determinados mesmo com essa *priori* bastante vaga.

Através da regra de Bayes, podemos obter a distribuição conjunta *a posteriori*:

$$p(\mathbf{a}, \sigma_e^2 | \mathbf{x}) \propto p(\mathbf{x} | \mathbf{a}, \sigma_e^2) p(\mathbf{a}, \sigma_e^2). \quad (4.15)$$

A substituição de (4.13) em (4.15) resulta em:

$$p(\mathbf{a}, \sigma_e^2 | \mathbf{x}) \propto \frac{1}{(2\pi\sigma_e^2)^{\frac{(N-P)}{2}}} \exp \left\{ -\frac{1}{2\sigma_e^2} \begin{bmatrix} \mathbf{x}_0^T & \mathbf{x}^T \end{bmatrix} \mathbf{A}^T \mathbf{A} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x} \end{bmatrix} \right\} \cdot \text{IG}(\sigma_e^2 | \alpha, \beta). \quad (4.16)$$

A distribuição condicional total para cada variável pode ser obtida a partir da distribuição conjunta olhando-se apenas para a variável que se deseja amostrar. Escrevendo a Equação (4.16) no formato de uma gaussiana padrão em função de \mathbf{a} através da Equação (4.6), obtemos [2]:

$$\mathbf{a}^{(j+1)} \sim \text{N}(\mathbf{a} | (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{x}, \sigma_e^{2(j)} (\mathbf{X}^T \mathbf{X})^{-1}), \quad (4.17)$$

enquanto para σ_e^2 , temos [2]

$$\sigma_e^{2(j+1)} \sim \text{IG}(\sigma_e^2 | \alpha + (N - P)/2, \beta_e + E(\mathbf{x}^{(j+1)}, \mathbf{a}^{(j+1)})/2), \quad (4.18)$$

em que \mathbf{X} é obtido a partir do sinal $\mathbf{x}^{(j+1)}$ de acordo com a Equação (4.7) (o índice foi omitido por clareza) e $E(\mathbf{x}^{(j)}, \mathbf{a}^{(j)})/2 = \mathbf{e}^T \mathbf{e}$, sendo $\mathbf{e} = \mathbf{A}^{(j+1)}[\mathbf{x}_0^T \ \mathbf{x}^{(j+1)T}]^T$ o erro de predição calculado usando-se os últimos dados amostrados de \mathbf{a} e \mathbf{x} .

Para fins de ilustração, aplicamos o algoritmo acima a um sinal com $N = 800$ amostras gerado a partir de um modelo AR com $\sigma_e^2 = 5 \times 10^{-6}$ e

$$\mathbf{a} = [5,1713 \quad -11,7727 \quad 15,1104 \quad -11,5384 \quad 4,9675 \quad -0,9415]^T. \quad (4.19)$$

A Figura 4.2 mostra esse sinal. Iniciando o algoritmo com todos os parâmetros iguais a zero e executando-o por 1000 iterações, obtemos os gráficos das Figuras 4.3 e 4.4, nos quais podemos ver que o algoritmo converge após poucas iterações. As Figuras 4.5 e 4.6 mostram os histogramas dos valores amostrados de σ_e^2 e dos elementos de \mathbf{a} , respectivamente. As 100 primeiras iterações foram descartadas para garantir que só fossem consideradas amostras após a convergência. Podemos ver que, em ambos os casos, o pico da distribuição está bem próximo do valor correto do parâmetro, conforme esperado. Nestes e nos demais gráficos mostrados no restante do texto, o quadrado indica o valor do parâmetro que foi usado para gerar os sinais, sendo portanto o ponto em torno do qual esperamos que o algoritmo convirja.

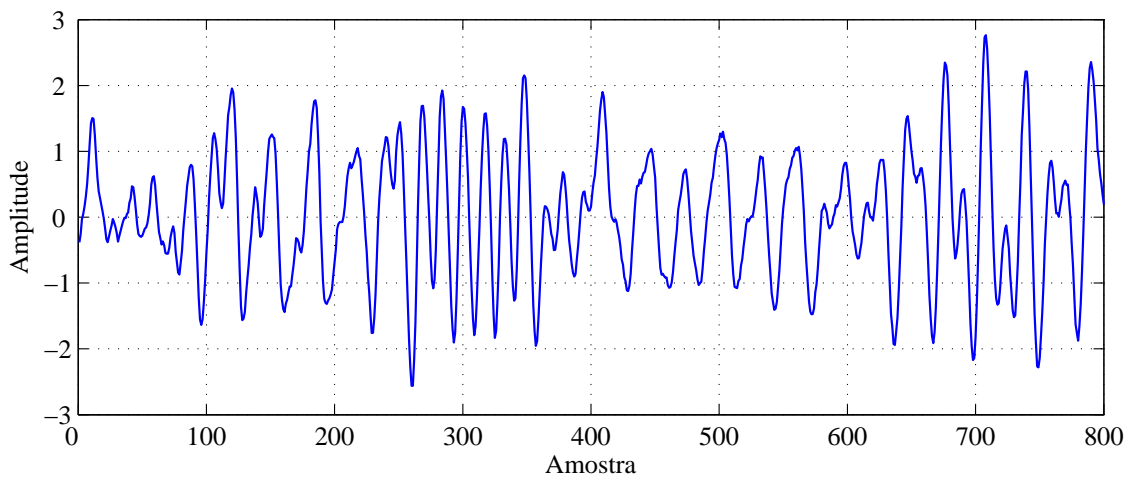


Figura 4.2: Trecho de sinal gerado por modelo AR.

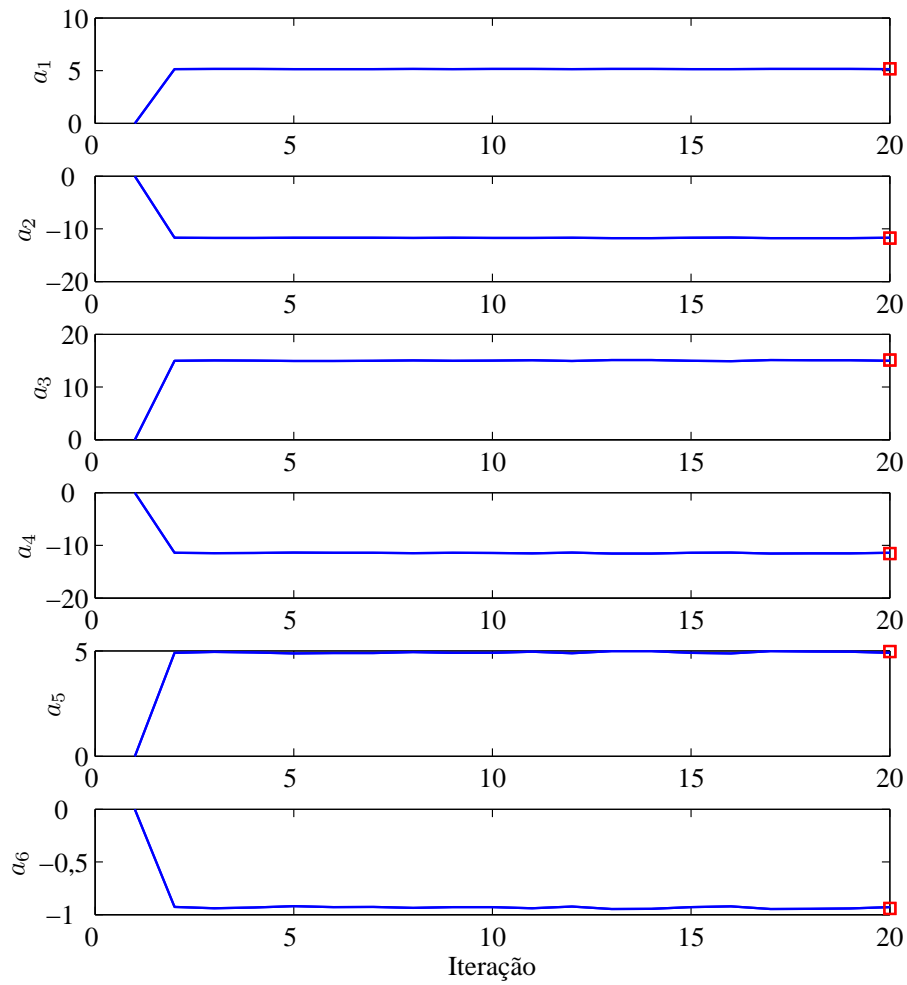


Figura 4.3: Evolução dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs.

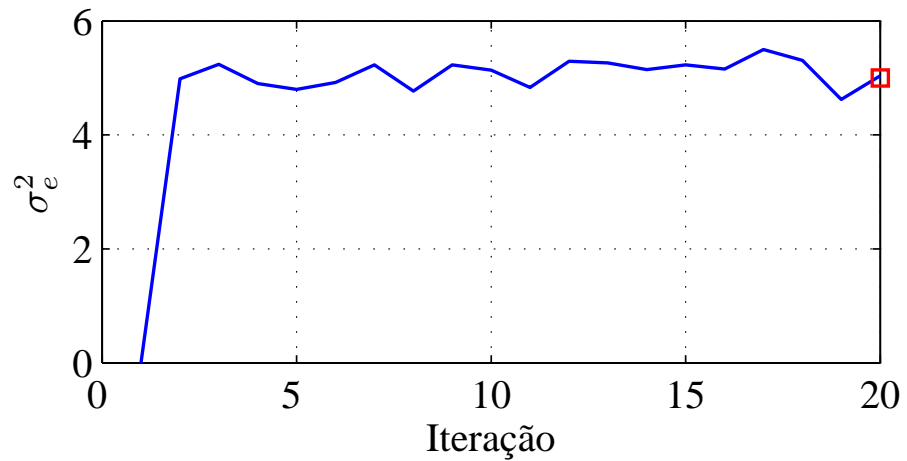


Figura 4.4: Evolução da variância de excitação do modelo AR obtida pelo amostrador de Gibbs.

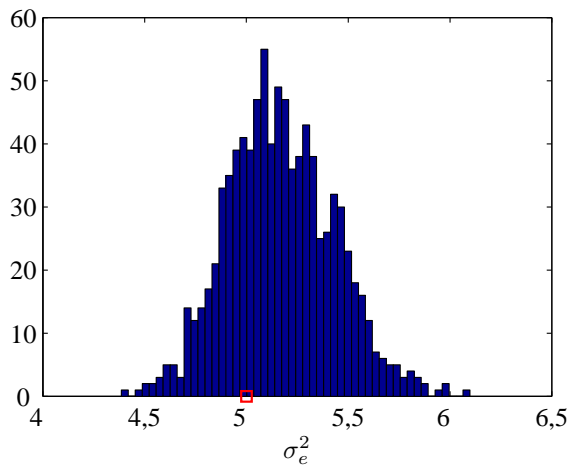


Figura 4.5: Histograma da variância de excitação do modelo AR, obtida pelo amostrador de Gibbs.

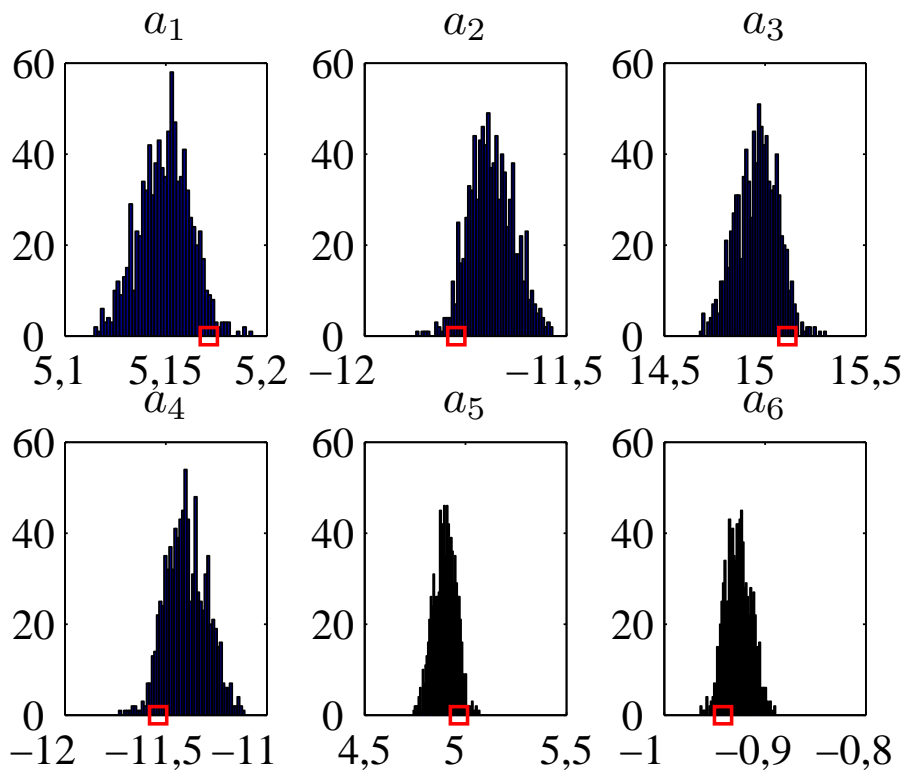


Figura 4.6: Histogramas de cada elemento dos coeficientes do modelo AR, obtidos a partir das amostras geradas pelo amostrador de Gibbs.

4.1.2 Incorporando um modelo para o ruído

Para alguns dos problemas tratados nesta tese, procuramos tornar os algoritmos robustos à presença de ruído aditivo. Nesta seção vamos mostrar como o ruído afeta a estimação dos parâmetros do modelo AR e como sua modelagem permite obter

melhores estimativas.

Consideremos que o sinal $x(n)$ está contaminado com ruído aditivo, sendo o sinal observado $y(n)$ a soma do sinal original com o ruído:

$$y(n) = x(n) + v(n). \quad (4.20)$$

Vamos assumir que $v(n)$ é branco e gaussiano com média zero e variância σ_v^2 , suposta conhecida. Como antes, assumimos que $x(n)$ é obtido a partir de um modelo AR com variância de excitação σ_e^2 e coeficientes $\mathbf{a} = [a_1 a_2 \dots a_P]$. Queremos estimar esses parâmetros com base em $y(n)$, a partir da distribuição *a posteriori*:

$$p(\mathbf{a}, \sigma_e^2 | \mathbf{y}) = \int_{\mathbf{x}} p(\mathbf{x}, \mathbf{a}, \sigma_e^2 | \mathbf{y}) d\mathbf{x} \quad (4.21)$$

Embora seja possível realizar a integral acima analiticamente, o resultado é de difícil tratamento para posterior otimização. Uma maneira de realizar essa tarefa numericamente é através do amostrador de Gibbs, que, neste contexto, além das operações descritas nas Equações (4.17) e (4.18), contém a amostragem da variável \mathbf{x} a partir de sua distribuição condicional, dada por:

$$p(\mathbf{x} | \mathbf{y}, \mathbf{a}, \sigma_e^2, \mathbf{x}_0) \propto p(\mathbf{y} | \mathbf{x}, \mathbf{a}, \sigma_e^2) p(\mathbf{x} | \mathbf{a}, \sigma_e^2) p(\mathbf{a}, \sigma_e^2). \quad (4.22)$$

Como $\mathbf{y} = \mathbf{x} + \mathbf{v}$, podemos dizer que a distribuição de \mathbf{y} dado \mathbf{x} , escrita em função de \mathbf{x} , é a distribuição de \mathbf{v} com média deslocada de \mathbf{y} , isto é:

$$p(\mathbf{y} | \mathbf{x}, \mathbf{a}, \sigma_e^2) = N(\mathbf{x} | \mathbf{y}, \mathbf{C}_v), \quad (4.23)$$

com $\mathbf{C}_v = \sigma_v^2 \mathbf{I}_{N-P}$. Usando a fórmula para o produto de gaussianas multivariáveis (ver Apêndice A.1), podemos mostrar que o produto dessa distribuição pela distribuição de $p(\mathbf{x} | \mathbf{a}, \sigma_e^2)$ resulta em:

$$p(\mathbf{x} | \mathbf{y}, \mathbf{a}, \sigma_e^2) = N(\mathbf{x} | \boldsymbol{\mu}_p, \mathbf{C}_p), \quad (4.24)$$

com $\boldsymbol{\mu}_p = (\boldsymbol{\mu}_x^T \mathbf{C}_x^{-1} + \mathbf{y}^T \mathbf{C}_v^{-1}) \mathbf{C}_p^{-1}$, $\mathbf{C}_p = (\mathbf{C}_x^{-1} + \mathbf{C}_v^{-1})^{-1}$, $\boldsymbol{\mu}_x = \mathbf{C}_x^{-1} \mathbf{A}_{P+1:N}^T \mathbf{A}_{1:P} \mathbf{x}_0$ e

$$\mathbf{C}_x = \sigma_e^2 (\mathbf{A}_{P+1:N}^T \mathbf{A}_{P+1:N})^{-1}. \quad ^1$$

Como ilustração, consideremos um sinal obtido com o mesmo modelo da seção anterior, agora corrompido por ruído branco aditivo a 30 dB de SNR, mostrado na Figura 4.7. Rodamos o amostrador de Gibbs por 10000 iterações, partindo de estimativas iniciais obtidas a partir do sinal ruidoso. Devido à má qualidade da estimativa inicial, o algoritmo demorou a convergir, como podemos ver nas Figuras 4.8 e 4.9. Os histogramas das Figuras 4.10 e 4.11, obtidos com as amostras a partir da iteração 5000, mostram que cada um dos parâmetros converge ao seu valor correto, embora a distribuição resultante tenha uma variância maior do que no caso sem ruído.

Para fins de comparação, mostramos nas Figuras 4.12 e 4.13 a evolução dos parâmetros do modelo AR obtida para esse mesmo sinal com o amostrador de Gibbs da seção anterior, que não modela o ruído. Podemos ver que a estimativa se torna bastante inacurada, mesmo para uma SNR relativamente alta (30 dB). Espera-se que para SNR menores que esta, a polarização seja ainda maior.

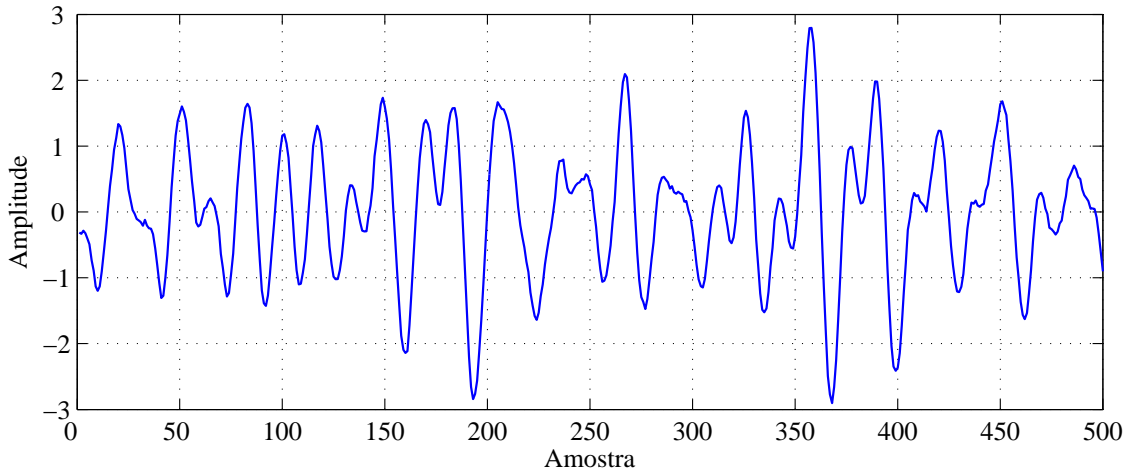


Figura 4.7: Sinal gerado por modelo AR, contaminado com ruído.

4.2 Modelagem da Distorção Não-linear

O sistema não-linear da Figura 4.1 pode ser classificado em quatro categorias: (1) sem memória, invertível; (2) sem memória, não-invertível; (3) com memória, inver-

¹A notação $\mathbf{A}_{i,j}$ indica a matriz formada pelas linhas de i a $(j - P + 1)$ de \mathbf{A} e colunas de i a j de \mathbf{A} .

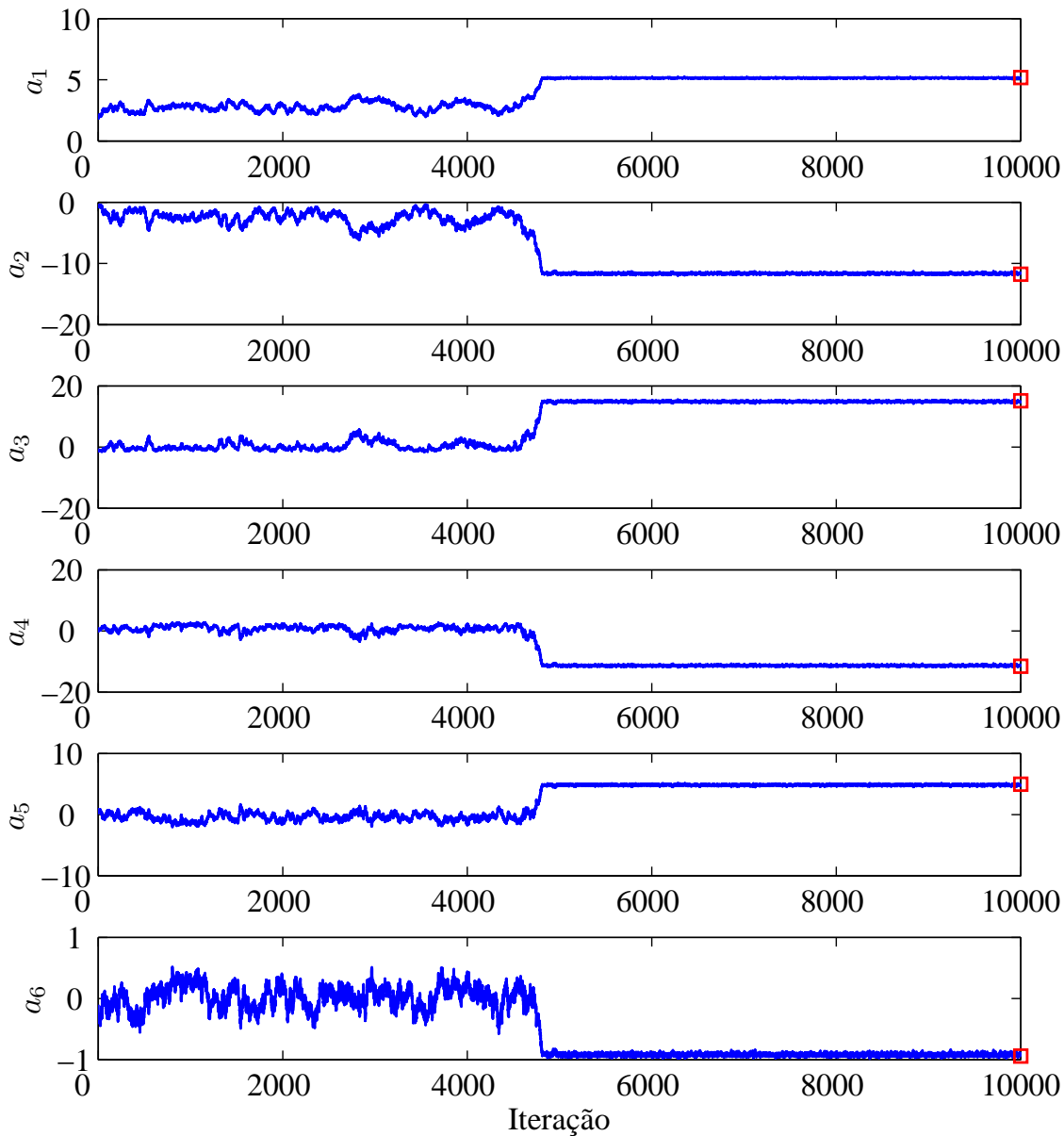


Figura 4.8: Evolução dos coeficientes do modelo AR para sinal com ruído.

tível; (4) com memória, não-invertível.²

O problema de restauração consiste em se estimar o sinal $x(n)$ com base no sinal degradado $y(n)$. Quando a não-linearidade é conhecida e invertível, sua inversa pode ser obtida de forma eficiente através de métodos iterativos [21]. Nosso problema é mais complicado porque, via de regra, a distorção é desconhecida ou, quando conhecida, não é invertível.

²Sistemas sem memória são aqueles em que a saída no instante n depende apenas da entrada nesse mesmo instante. Invertibilidade implica que, conhecidos o sistema e o sinal degradado, pode-se recuperar o sinal original exatamente.

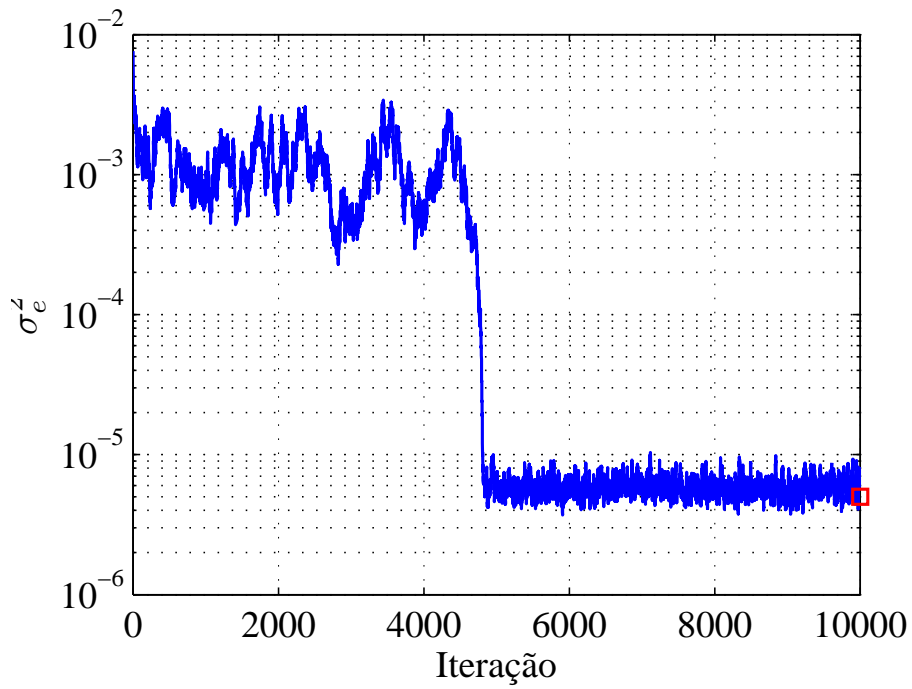


Figura 4.9: Evolução da variância de excitação do modelo AR, obtidos pelo amostrador de Gibbs, para sinal com ruído.

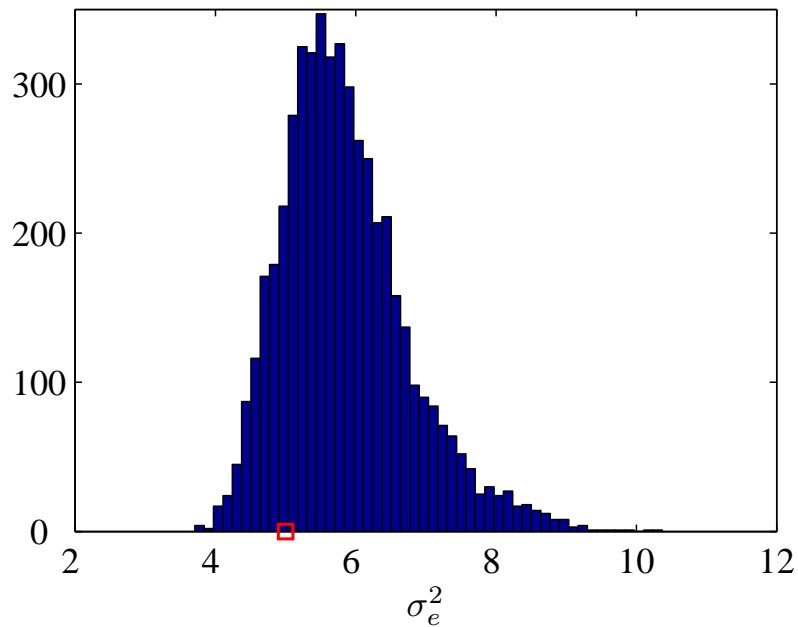


Figura 4.10: Histograma da variância de excitação do modelo AR, obtidos pelo amostrador de Gibbs, para sinal com ruído.

Em geral, a solução desse problema no domínio do tempo envolve, além da modelagem do sinal como na seção anterior, a escolha da estrutura algébrica do sistema não-linear. Devido à multiplicidade de distorções possíveis, essa tarefa

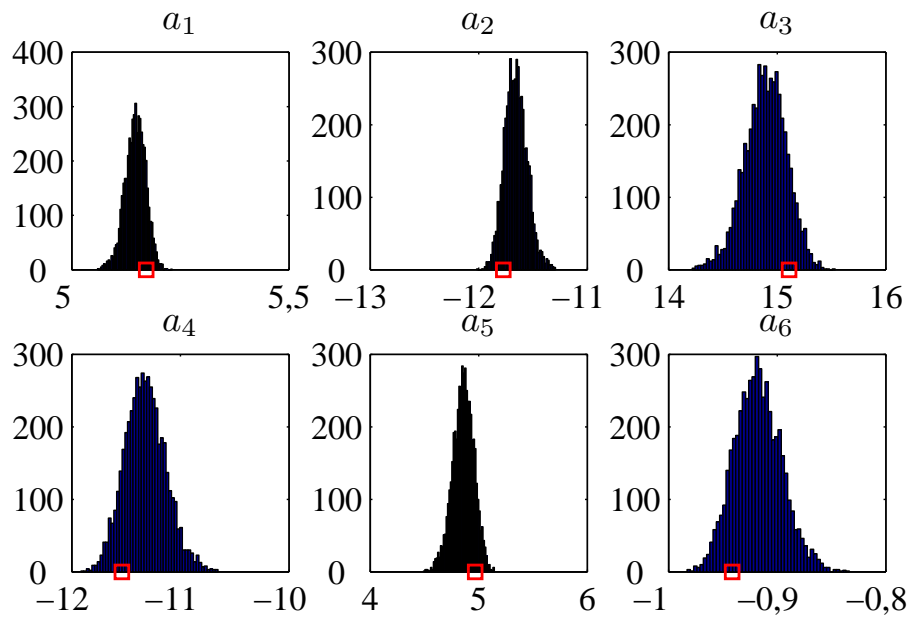


Figura 4.11: Histogramas de cada elemento dos coeficientes do modelo AR, obtidos pelo amostrador de Gibbs, para sinal com ruído.

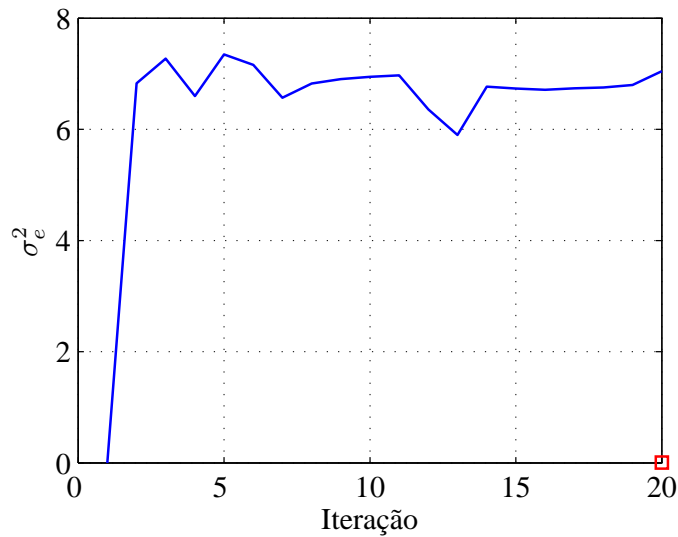


Figura 4.12: Evolução da variância de excitação do modelo AR obtidos pelo amostrador de Gibbs a partir de um sinal ruidoso, mas sem modelar o ruído.

é o maior desafio: estruturas muito genéricas como a série de Volterra gerariam uma combinação impraticavelmente alta de termos da série, tornando a tarefa de identificação praticamente impossível em casos reais [22]. Nesta tese, atacamos separadamente alguns casos particulares, de relevância prática, em que a estrutura do sistema não-linear pudesse ser identificada de forma relativamente simples.

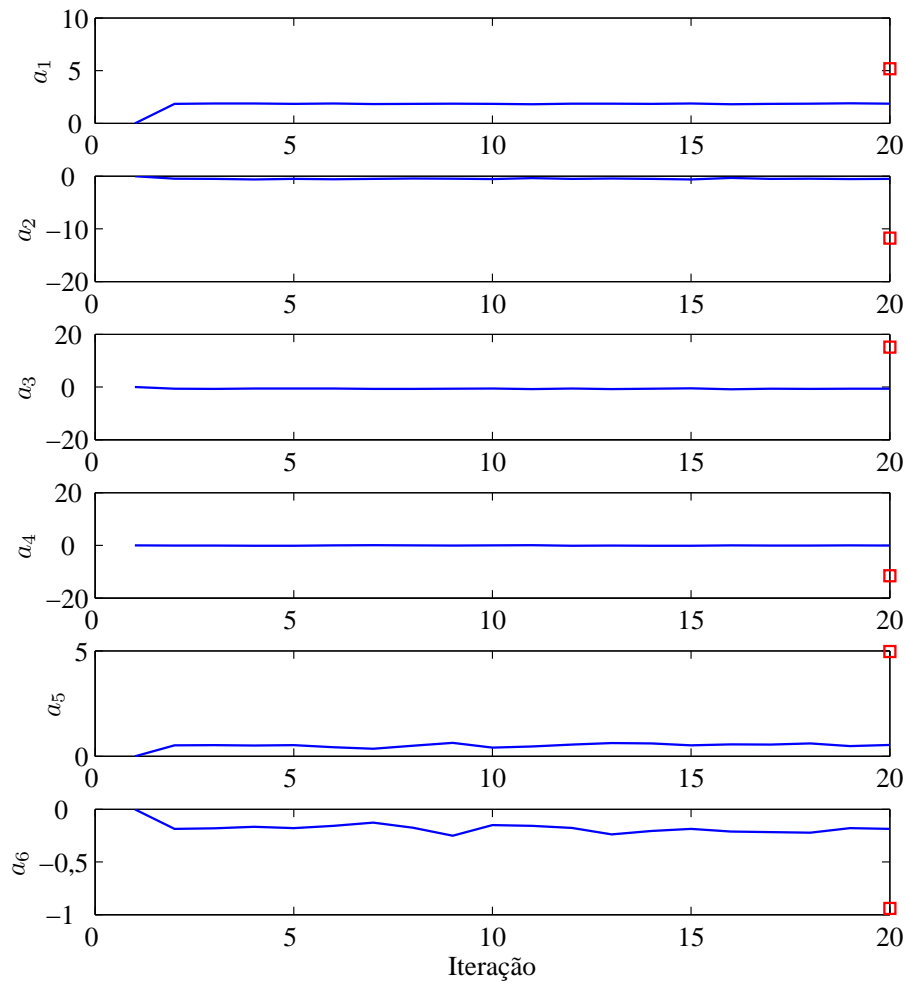


Figura 4.13: Evolução dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs a partir de um sinal ruidoso, mas sem modelar o ruído.

4.2.1 Resumo das soluções propostas

Em vez de procurarmos estruturas gerais o suficiente para modelar todos os casos de não-linearidade, nesta tese nos voltamos para o estudo de casos particulares importantes em problemas de áudio. Em cada um dos capítulos associados, faremos a respectiva revisão bibliográfica.

Como um caso de não-linearidade não-invertível sem memória (NISM), consideramos o problema de saturação radical, que se caracteriza pela limitação da amplitude do sinal em certos valores máximos e mínimos. Nesse caso, a identificação da curva de distorção é trivial, e o problema se resume ao de interpolação de amostras perdidas com restrições. Esse é o assunto do Capítulo 5.

O problema de saturação gradual, representativo de não-linearidade invertível

sem memória (ISM), é o assunto do Capítulo 6. Assume-se que a curva não-linear é invertível, podendo ser parametrizada por coeficientes da série de Taylor em torno da origem, ou modelada por uma curva não-linear por partes. Estimamos os coeficientes da curva por duas formas: (1) maximização da verossimilhança através do método de Gauss-Newton ou (2) maximização da distribuição *a posteriori* através de MCMC.

No Capítulo 7, consideramos um modelo invertível com memória (ICM), representável pelo modelo de Hammerstein tendo como entrada o sinal original, não observável, e como saída o sinal degradado, observável. Assumindo-se o modelo invariante no tempo, propõe-se um método bayesiano capaz de identificá-lo.

Na maioria dos casos estudados, procuramos tornar os algoritmos de restauração robustos à presença de ruído aditivo. Por conveniência matemática, assumimos ruído branco e gaussiano, com variância conhecida e fixa, podendo ser estimada a partir de períodos de silêncio do sinal.

Capítulo 5

Restauração de Sinais com Limitação em Amplitude

No contexto de áudio, o problema de limitação de amplitude ocorre, por exemplo, quando a amplitude do sinal excede a faixa dinâmica permitida pelo sistema de transmissão ou armazenamento. Esse fenômeno, popularmente chamada de saturação e aqui definida como “saturação radical” (em contraste com “saturação gradual”, assunto do próximo capítulo), prejudica de forma substancial a qualidade da experiência auditiva, o que justifica o uso de técnicas para atenuação do problema.

A maneira mais simples de estimar o sinal original é tratar as amostras saturadas como perdidas e interpolar a região defeituosa usando informação das amostras das regiões não-corrompidas. Um método que pode ser usado para resolver esse problema foi proposto em [23], no qual se assume que o sinal é descrito por um modelo AR e o número de amostras desconhecidas é pequeno em relação às restantes. Embora sub-ótimo, verificamos que esse método gera bons resultados para sinais com saturação moderada.

Uma abordagem mais específica ao problema foi proposta [24], no qual se assume que o sinal foi superamostrado (isto é, amostrado acima da taxa de Nyquist). O método gera amostras que minimizam a energia do sinal na região saturada, sujeito a restrições nas amplitudes e na faixa espectral do sinal. Uma ideia similar é apresentada em [25], no qual se explora a esparsidade na representação do sinal em bases senoidais e também se levam em consideração restrições na amplitude. Ambos os métodos funcionam bem quando suas suposições são válidas, mas tendem a gerar

resultados insatisfatórios quando ruído ou componentes do sinal fora da faixa estão presentes.

Uma solução para o problema de deconvolução de sinais ultrassônicos saturados é proposta em [26]. Em princípio, esse método poderia ser aplicado a sinais de áudio, mas os parâmetros do modelo, tais como os coeficientes do modelo AR, precisariam ser estimados inicialmente de forma sub-ótima. Outra solução estatística, baseada em filtragem de partículas, foi proposta em [27], mas com pouca discussão sobre os resultados obtidos.

Nesse capítulo, propomos uma solução bayesiana baseada em modelos para o problema de restauração de sinais com saturação radical, que permite incluir de forma natural a modelagem de ruído aditivo. Um modelo autorregressivo é associado ao sinal de áudio e o ruído é assumido branco e gaussiano, com variância conhecida. É proposto um algoritmo formado pela combinação do algoritmo de Metropolis-Hastings com o amostrador de Gibbs, o qual permite obter amostras da distribuição *a posteriori* do sinal original na região corrompida. Em todos os casos abordados, vamos assumir que o sinal é mono, isto é, gravado em um único canal.

O capítulo está organizado da seguinte forma. Na Seção 5.1, apresentamos a formulação matemática do problema e alguns comentários sobre as possíveis soluções. Em seguida, na Seção 5.2, propomos uma solução bayesiana que permite obter amostras da distribuição *a posteriori* do sinal original através de MCMC. Variações do problema em que ruído aditivo está presente são apresentadas na Seção 5.3, para as quais também propomos uma solução bayesiana. Ao longo dessas seções, apresentaremos testes dos métodos propostos com sinais artificiais e reais, comparando seus desempenhos com soluções existentes, quando for o caso. Por fim, na Seção 5.4, apresentamos as conclusões deste capítulo.

5.1 Formulação do Problema

O sinal $y(n)$, saturado entre os níveis c_+ e c_- , é obtido a partir do sinal original $x(n)$ de acordo com:

$$y(n) = f[x(n)] = \begin{cases} x(n), & c_- \leq x(n) \leq c_+ \\ c_-, & x(n) < c_- \\ c_+, & x(n) > c_+. \end{cases} \quad (5.1)$$

Para identificar as diferentes regiões do sinal, definimos o sinal auxiliar $i(n)$:

$$i(n) = \begin{cases} 0, & c_- \leq x(n) \leq c_+ \\ -1, & x(n) < c_- \\ 1, & x(n) > c_+. \end{cases} \quad (5.2)$$

Se $y(n)$ for observado sem ruído, o sinal $i(n)$ pode ser obtido trivialmente olhando-se os limites superiores e inferiores do sinal. Na presença de ruído, a estimação de $i(n)$ é mais complicada e será discutida na Seção 5.3.

Definimos os vetores \mathbf{x}_c , \mathbf{x}_+ e \mathbf{x}_- como sendo formados pelas amostras de \mathbf{x} correspondentes às regiões em que $i(n) = 0$, $i(n) = 1$ e $i(n) = -1$, respectivamente. O objetivo da restauração é estimar as partes saturadas do sinal, isto é \mathbf{x}_+ e \mathbf{x}_- , com base em \mathbf{x}_c , idealmente respeitando os limites de amplitude definidos na Equação (5.1).

O sinal original \mathbf{x} pode ser escrito como $\mathbf{x} = \mathbf{I}_c \mathbf{x}_c + \mathbf{I}_s \mathbf{x}_s$, em que \mathbf{x}_s reúne os elementos de \mathbf{x} na região saturada, isto é $\mathbf{x}_s = [\mathbf{x}_+^T \ \mathbf{x}_-^T]^T$. \mathbf{I}_c e \mathbf{I}_s são matrizes compostas por zeros e uns que, ao serem multiplicadas por \mathbf{x}_c e \mathbf{x}_s , respectivamente, colocam os elementos desses vetores nas suas posições corretas no vetor \mathbf{x} .

Como visto na Seção 4.1, $\mathbf{e} = \mathbf{A}[\mathbf{x}_0^T \ \mathbf{x}^T]^T$. Com base nessa expressão, pode-se obter uma estimativa para os valores saturados de \mathbf{x} derivando-se a norma-2 de \mathbf{e} em relação a \mathbf{x}_s , igualando o resultado a zero e resolvendo para \mathbf{x}_s . Pode-se mostrar que esse procedimento resulta em [2]:

$$\mathbf{x}_s^{\text{LS}} = -(\mathbf{A}_s^T \mathbf{A}_s)^{-1} \mathbf{A}_s^T \mathbf{A}_c \mathbf{x}_c, \quad (5.3)$$

em que \mathbf{A}_s e \mathbf{A}_c correspondem a partições da matriz \mathbf{A} contendo as linhas e colunas

correspondentes a amostras saturadas e corretas, respectivamente.

Nessa expressão, ignoramos as restrições de amplitude definidas na Equação (5.1). A formulação do problema como o de otimização com restrições geraria uma solução que, embora ótima do ponto de vista de erro médio quadrático, poderia não ser agradável perceptivamente, pois algumas propriedades do sinal poderiam não ser respeitadas. Por exemplo, a solução ótima poderia ser todos os valores constantes e iguais ao limite da restrição (isto é, o valor mínimo de amplitude permitido), o que certamente geraria artefatos audíveis.

Na prática, os parâmetros do modelo AR não são conhecidos e precisam ser estimados a partir das amostras observadas. Exemplos de solução para a estimativa conjunta dos dados faltantes e dos parâmetros podem ser encontradas em [28] e [23]. Em particular, nesse último consegue-se obter estimativas de boa qualidade através de um procedimento iterativo simples, em que partindo-se de estimativas iniciais para \mathbf{a} e \mathbf{x}_s , geram-se estimativas mais refinadas para essas variáveis através da minimização do erro quadrático para cada variável condicionada ao conhecimento da outra.

Embora essa solução permita obter estimativas de boa qualidade para saturação moderada, ela é sub-ótima por dois motivos: (1) pode convergir para ótimos locais dependendo da inicialização e (2) as restrições quanto à amplitude do sinal na região saturada não são levadas em consideração. Na próxima seção apresentamos um procedimento bayesiano com solução baseada em MCMC que permite superar essas duas limitações.

5.2 Solução Bayesiana

Para a obtenção da solução bayesiana, começamos calculando a distribuição *a posteriori* das variáveis desconhecidas \mathbf{x}_+ , \mathbf{x}_- , \mathbf{a} e σ_e^2 dadas as amostras conhecidas \mathbf{x}_c e assumindo-se que $\mathbf{x}_+ > c_+$ e $\mathbf{x}_- < c_-$, isto é:

$$\begin{aligned}
 p(\mathbf{x}_s, \mathbf{a}, \sigma_e^2 | \mathbf{x}_c, \mathbf{x}_+ > c_+, \mathbf{x}_- < c_-) &\propto p(\mathbf{x}, \mathbf{x}_+ > c_+, \mathbf{x}_- < c_-, \mathbf{a}, \sigma_e^2) \\
 &\propto p(\mathbf{x} | \mathbf{a}, \sigma_e^2) p(\mathbf{a}, \sigma_e^2) \times \\
 &\quad u(\mathbf{x}_+ - c_+) u(-\mathbf{x}_+ c_-). \tag{5.4}
 \end{aligned}$$

5.2.1 Algoritmo proposto

Propomos um algoritmo bayesiano para a solução do problema que se baseia na amostragem da distribuição conjunta da Equação (5.4) através do amostrador de Gibbs, descrito a seguir:

1. Gerar valores iniciais $\mathbf{x}_s^{(0)}$, $\mathbf{a}^{(0)}$ e $\sigma_e^{2(0)}$.
2. Para j de 1 até N_{it}
 - (a) $\mathbf{x}_s^{(j)} \sim p(\mathbf{x}_s | \mathbf{x}_c, \mathbf{x}_+ > c_+, \mathbf{x}_- < c_-, \mathbf{a}^{(j-1)}, \sigma_e^{2(j-1)})$;
 - (b) $\mathbf{a}^{(j)} \sim p(\mathbf{a} | \mathbf{x}^{(j)}, \sigma_e^{2(j-1)})$;
 - (c) $\sigma_e^{2(j)} \sim p(\sigma_e^2 | \mathbf{x}^{(j)}, \mathbf{a}^{(j)})$.

A inicialização pode ser feita usando-se os resultados obtidos com o procedimento descrito na seção anterior. A distribuição condicional total para cada variável é obtida a partir da distribuição conjunta escrita na Equação (5.4) olhando-se apenas para a variável que se deseja amostrar. Para os parâmetros \mathbf{a} e σ_e^2 , condicionados a \mathbf{x} , o procedimento é idêntico ao apresentado na Seção 4.1.1.

Para obtenção da distribuição \mathbf{x}_s condicionada às demais variáveis, basta substituir \mathbf{x}_c nas posições correspondentes em \mathbf{x} e escrever a expressão resultante no formato de uma gaussiana padrão em \mathbf{x}_s . Pode-se mostrar que o resultado é dado por [2]:

$$p(\mathbf{x}_s | \mathbf{x}_c, \mathbf{x}_+ > c_+, \mathbf{x}_- < c_-, \mathbf{a}, \sigma_e^2) \propto N(\mathbf{x}_s | \boldsymbol{\mu}_{xp}, \mathbf{C}_{xp}) u(\mathbf{x}_+ - c_+) u(-\mathbf{x}_- + c_-), \quad (5.5)$$

com $\boldsymbol{\mu}_{xp} = \mathbf{C}_{xp}^{-1} \mathbf{A}_s \mathbf{A}_c^T \mathbf{x}_c$ e $\mathbf{C}_{xp} = \sigma_e^2 (\mathbf{A}_s^T \mathbf{A}_s)^{-1}$.

A distribuição acima representa uma gaussiana truncada na região definida por $\mathbf{x}_+ \geq c_+$ e $\mathbf{x}_- \leq c_-$. Para a amostragem dessa variável, diversas estratégias podem ser adotadas. A mais simples seria obter amostras a partir da gaussiana multivariável e só aceitar o vetor de amostras se todas caíssem na região permitida. Dependendo da dimensão da gaussiana a ser amostrada e da região permitida, essa estratégia pode ser impraticável se, em um percentual alto das iterações, as amostras geradas caírem na região não permitida. Para mitigar esse problema, pode-se subdividir o conjunto de variáveis em blocos menores e amostrar cada bloco

separadamente, condicionado aos demais, produzindo assim uma probabilidade de aceitação maior (ao custo de aumentar o número de iterações para convergência). Verificamos em testes com sinais reais que para saturação moderada, a primeira estratégia funciona melhor, pois a probabilidade de aceitação é quase sempre bastante alta. Para níveis de saturação mais severos, a estratégia de amostrar cada “semi-ciclo” (maior número possível de amostras adjacentes com o mesmo sinal, positivo ou negativo) por vez se mostrou bastante eficiente.

5.2.2 Resultados

Para ilustração do algoritmo, vamos considerar um bloco de sinal gerado pelo mesmo modelo da Subseção 4.1.1, artificialmente limitado num ponto máximo e num ponto mínimo, como mostrado na Figura 5.1. Iniciando os parâmetros a partir do método descrito em [23], rodamos o amostrador de Gibbs por 1000 iterações, sendo as 100 primeiras mostradas na Figura 5.2 e 5.3. Podemos ver que os coeficientes do modelo AR já foram inicializados em valores bem próximos dos corretos, e por isso variam pouco ao longo das iterações. Já a variância da excitação se inicia num valor bem abaixo do correto e converge para uma região próxima a este após algumas dezenas de iterações.

Usando as amostras das 900 últimas iterações do algoritmo, geramos os histogramas da excitação σ_e^2 , dos elementos de \mathbf{a} e da amostra 340 (escolhida por estar na região saturada) do sinal original, mostrados respectivamente nas Figuras 5.4, 5.5 e 5.6. Com base nesses histogramas podemos verificar a acurácia do algoritmo proposto.

A Figura 5.7 mostra a comparação em termos de SNR entre o algoritmo proposto e o método descrito em [23], para diferentes níveis de saturação. Nesse gráfico, a abscissa representa a fração da máxima amplitude do sinal original em que o sinal degradado foi limitado. Portanto, quanto menor a abscissa, maior é o nível de degradação. Podemos concluir que a diferença entre os métodos é pequena quando o nível de saturação é baixo, mas tende a crescer quando este aumenta.

Aplicando nosso método a sinais de música reais, com limitação de amplitude artificialmente introduzida, concluímos que este gera melhorias substanciais na qualidade do sinal. Entretanto, para a maioria dos níveis de limitação encontrados

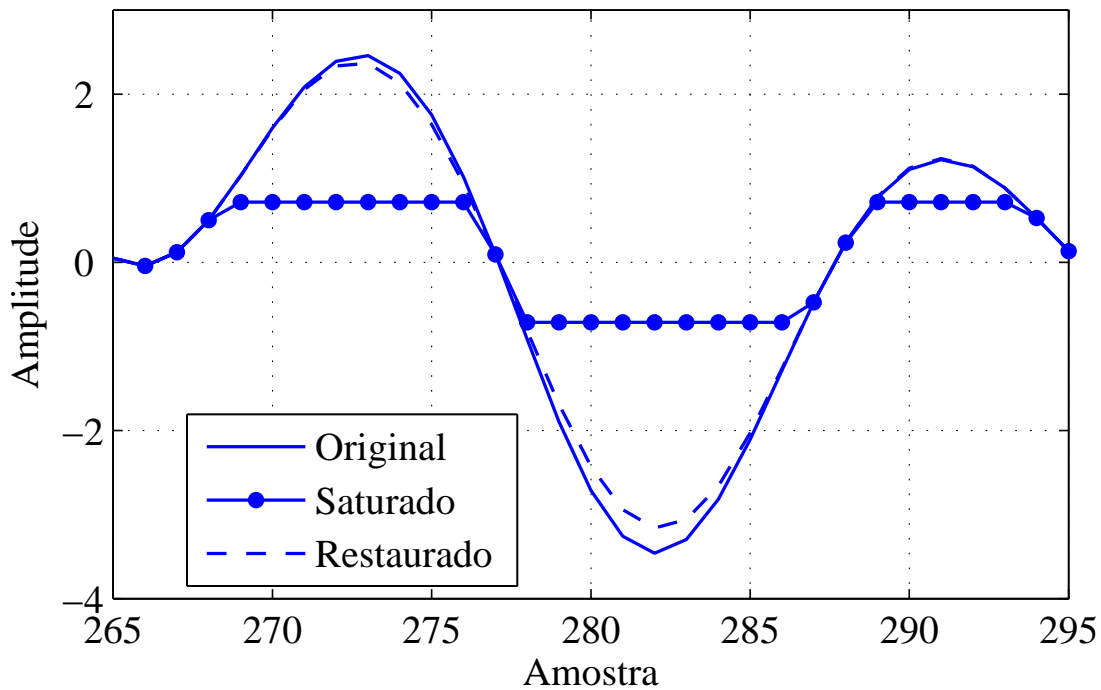


Figura 5.1: Comparação entre trechos de sinal original, saturado e restaurado pelo método proposto. O sinal saturado foi limitado a uma amplitude de 25 % da amplitude máxima do sinal original.

na prática, o ganho de nosso método em relação ao método de referência é apenas modesto, sendo muitas vezes auditivamente imperceptível. Para caso de limitação severa, a diferença perceptiva é maior, porém este nível de degradação não é muito comum na prática.

5.3 Modelagem de ruído aditivo

Esta seção propõe uma abordagem bayesiana para o problema de restauração de sinais saturados na presença de ruído. Propomos um algoritmo baseado no amostrador de Gibbs, que permite a obtenção de amostras da distribuição *a posteriori* conjunta do sinal original \mathbf{x} , e dos parâmetros \mathbf{a} e σ_e^2 .

Dois casos possíveis para a incidência do ruído são considerados, como mostra a Figura 5.8. No caso A, o ruído incide no sinal antes da saturação, enquanto no caso B o oposto ocorre. No primeiro caso, os limites de saturação podem ser obtidos de forma trivial, enquanto no segundo é preciso tratá-los como variáveis e incorporá-los ao modelo. Nos dois casos, assumimos ruído branco gaussiano com variância

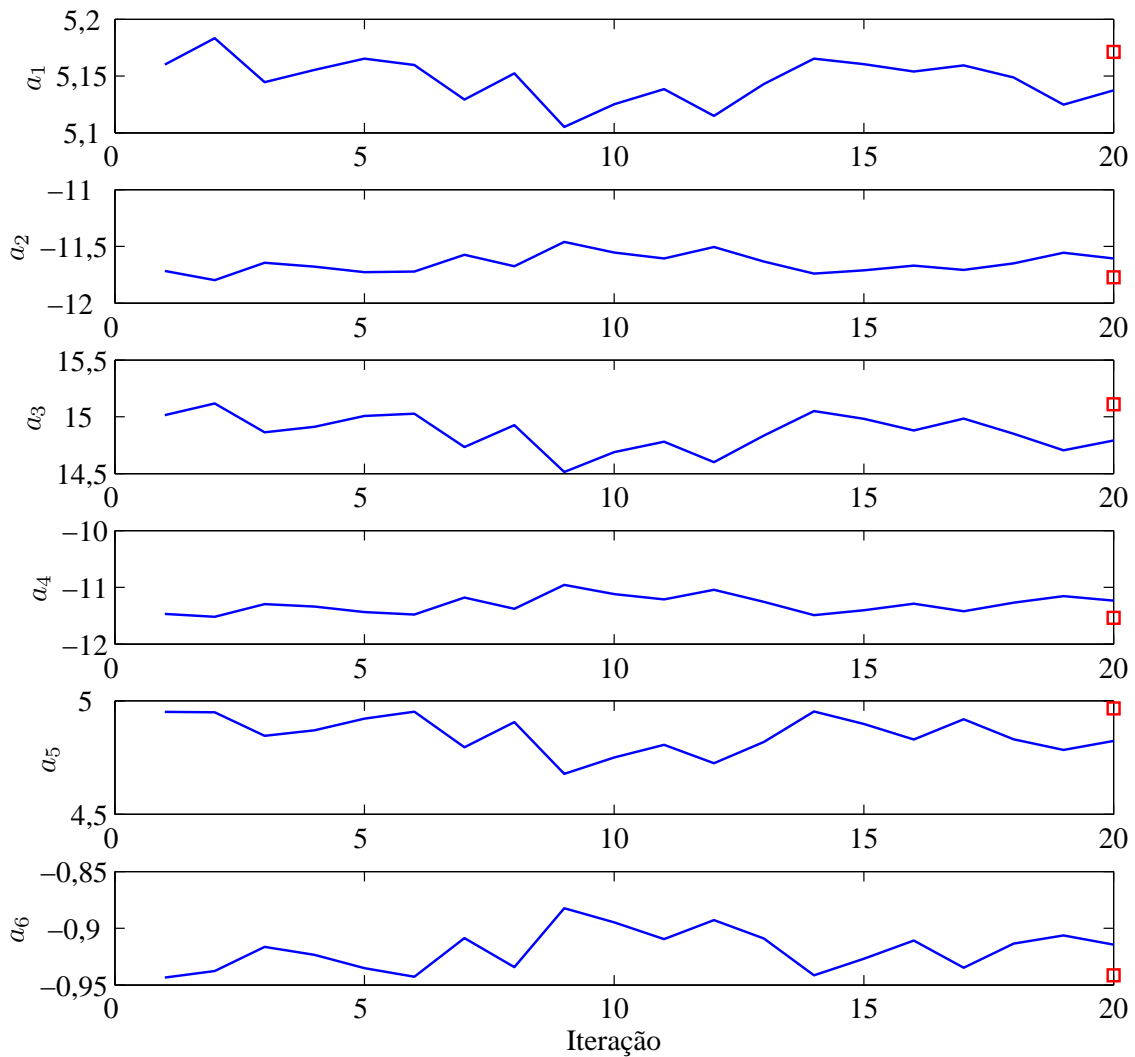


Figura 5.2: Evolução dos coeficientes do modelo AR, obtidos com o amostrador de Gibbs a partir de um sinal com saturação. A inicialização com base no método de mínimos quadrados já está bem próxima dos valores corretos. O amostrador de Gibbs corrige essa polarização.

conhecida. Nas próximas seções, vamos analisar cada caso isoladamente.

5.3.1 Ruído seguido de não-linearidade

O sinal degradado é dado por:

$$y(n) = \begin{cases} x(n) + v(n), & c_- \leq x(n) + v(n) \leq c_+ \\ c_+, & x(n) + v(n) > c_+ \\ c_-, & x(n) + v(n) < c_-. \end{cases} \quad (5.6)$$

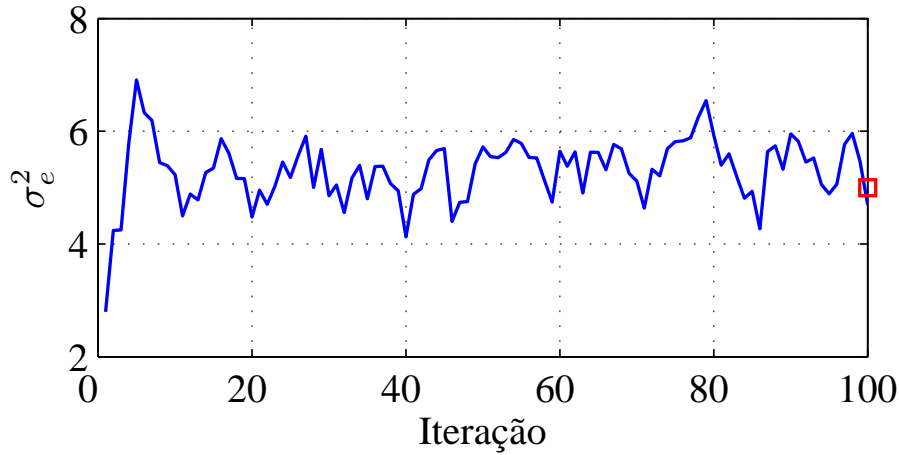


Figura 5.3: Evolução da variância de excitação do modelo AR, obtidos com o amostrador de Gibbs a partir de um sinal com saturação. Como a inicialização com base no método dos mínimos quadrados minimiza a variância do ruído, essa variável tende a ser subestimada.

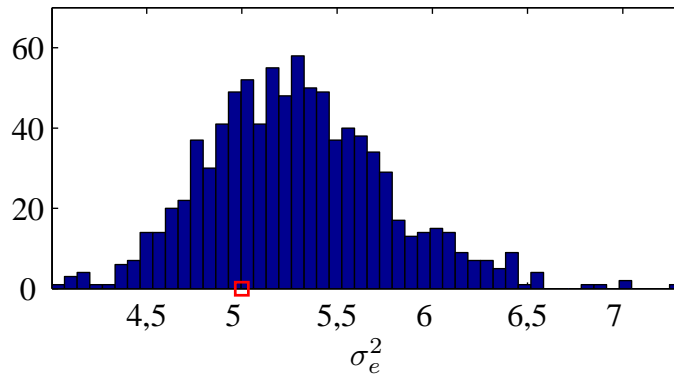


Figura 5.4: Histograma da variância excitação do modelo AR obtido a partir das amostras geradas pelo amostrador de Gibbs a partir de um sinal com saturação.

Como antes, o objetivo é estimar o sinal original $x(n)$ com base no sinal degradado $y(n)$. Em uma abordagem bayesiana baseada em modelos, precisamos especificar modelos estatísticos para o sinal original, como feito anteriormente, e para o ruído, aqui modelado como branco e gaussiano, com média zero, isto é $p_v(\mathbf{v}) = N(\mathbf{v}|\mathbf{0}, \sigma_v^2 \mathbf{I}_N)$, em que a variância σ_v^2 é assumida constante e conhecida dentro de cada bloco, podendo ser estimadas a partir de períodos de silêncio no sinal.

Os limites de saturação c_+ e c_- podem ser facilmente obtidos calculando-se os valores máximo e mínimo do sinal degradado. Vamos então nos concentrar na

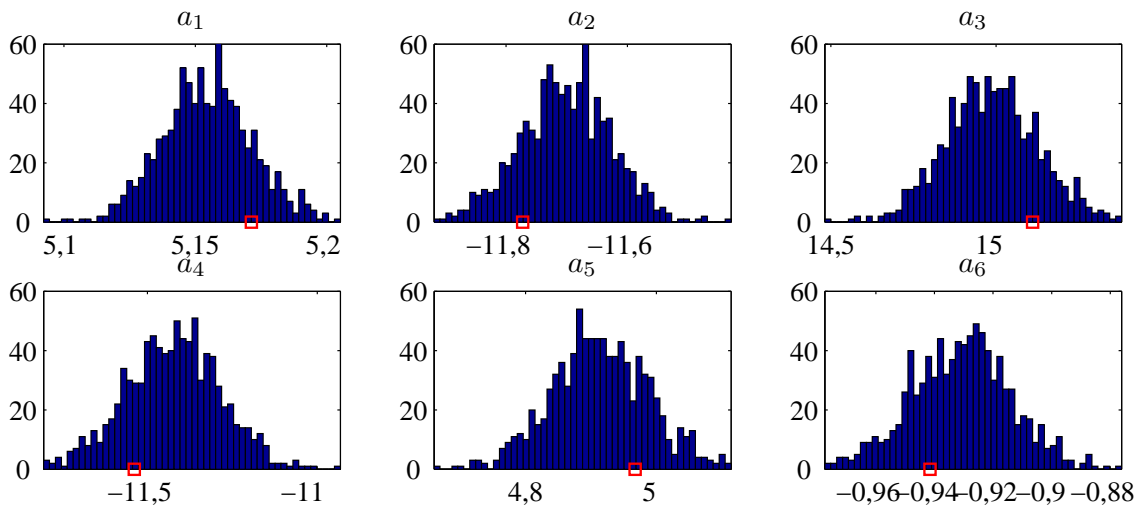


Figura 5.5: Histogramas de cada elemento dos coeficientes do modelo AR, obtidos a partir das amostras geradas pelo amostrador de Gibbs a partir de um sinal com saturação.

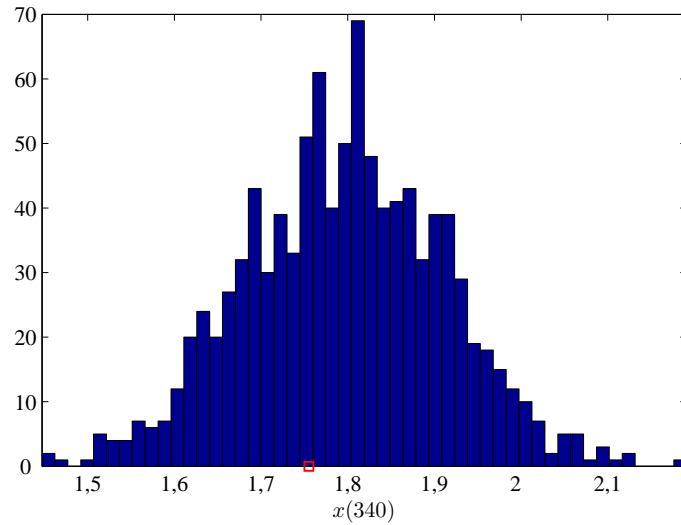


Figura 5.6: Histograma da amostra 340 (numa região corrompida) do sinal restaurado obtido a partir do amostrador de Gibbs a partir de um sinal com saturação.

estimação do sinal original \mathbf{x} em conjunto com os parâmetros \mathbf{a} e σ_e^2 .

5.3.1.1 Cálculo da distribuição *a posteriori*

Nas equações seguintes, os conjuntos C , C_+ e C_- contêm os índices das amostras corretas, saturadas superiormente e saturadas inferiormente, respectivamente. Como as amostras de $y(n)$ são independentes entre si dado \mathbf{x} , a verossimilhança é dada

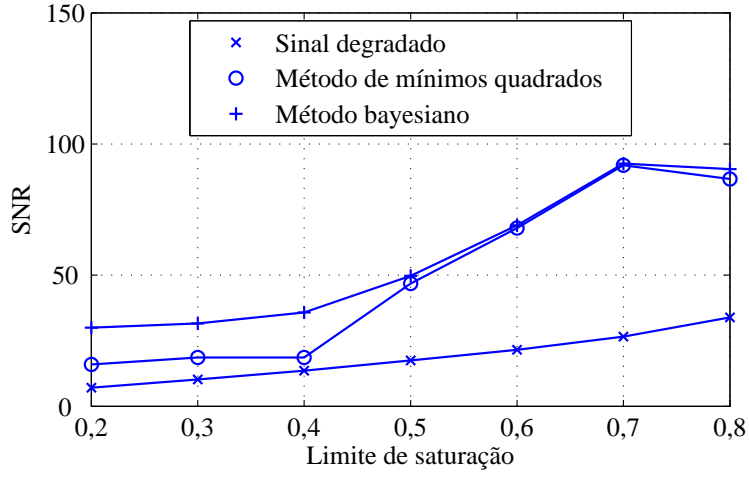


Figura 5.7: Comparação entre o método proposto (bayesiano) e o método baseado em mínimos quadrados, usado como referência, em termos de SNR. Ambos os métodos aumentam a SNR em relação ao sinal saturado. Para saturação muito severa, o método proposto oferece aumentos maiores de SNRs, mas os resultados são muito próximos para saturação moderada ou leve.

por:

$$\begin{aligned}
 p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) &= p_v(\mathbf{x}_c - \mathbf{y}_c) \times \\
 &P\{x(i) + v(i) > c_+, i \in C_+\} \times \\
 &P\{x(i) + v(i) < c_-, i \in C_-\}. \tag{5.7}
 \end{aligned}$$

Manipulando os dois últimos termos e usando a simetria da gaussiana, obtemos:

$$\begin{aligned}
 p(\mathbf{y}|\mathbf{x}, \mathbf{a}, \sigma_e^2) &= p_v(\mathbf{x}_c - \mathbf{y}_c) \times \\
 &\prod_{i \in C_+} \phi\left(\frac{x(i) - c_+}{\sigma_v^2}\right) \times \\
 &\prod_{i \in C_-} \phi\left(\frac{c_- - x(i)}{\sigma_v^2}\right). \tag{5.8}
 \end{aligned}$$

onde $\phi(\cdot)$ é a função distribuição de probabilidade cumulativa (CDF) da variável aleatória gaussiana de média zero e variância 1.

Pela regra de Bayes, a distribuição *a posteriori* conjunta é dada por:

$$p(\mathbf{x}, \mathbf{a}, \sigma_e^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \mathbf{a}, \sigma_e^2)p(\mathbf{x}|\mathbf{a}, \sigma_e^2)p(\mathbf{a}, \sigma_e^2), \tag{5.9}$$

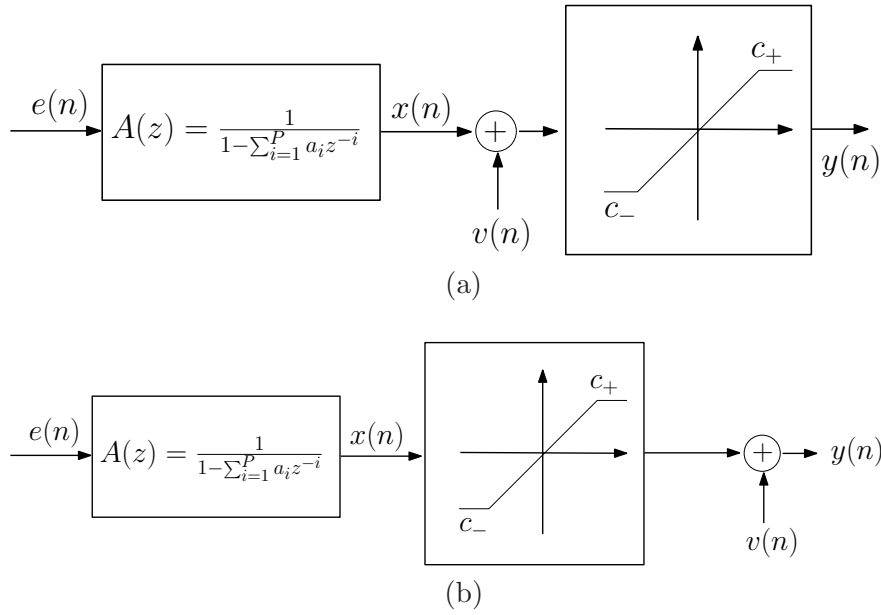


Figura 5.8: Diagrama de blocos descrevendo o modelo para sinais com saturação contaminado com ruído. (a) Ruído antes da saturação. (b) Ruído depois da saturação.

onde $p(\mathbf{a}, \sigma_e^2)$ é a distribuição *a priori* de \mathbf{a} e σ_e^2 , já discutida no capítulo anterior.

Combinando a priori dos parâmetros com a verossimilhança, a distribuição *a posteriori* pode ser escrita como:

$$\begin{aligned}
 p(\mathbf{x}, \mathbf{a}, \sigma_e^2 | \mathbf{y}) &\propto p_v(\mathbf{x}_c - \mathbf{y}_c) \times \\
 &\prod_{i \in C_+} \phi\left(\frac{x(i) - c_+}{\sigma_v^2}\right) \prod_{i \in C_-} \phi\left(\frac{c_- - x(i)}{\sigma_v^2}\right) \times \\
 &p(\mathbf{x} | \mathbf{a}, \sigma_e^2) p(\mathbf{a}, \sigma_e^2).
 \end{aligned} \tag{5.10}$$

5.3.1.2 Algoritmo proposto

A ideia do algoritmo proposto é estimar o sinal original através do cálculo da média das amostras da distribuição *a posteriori* marginalizada $p(\mathbf{x} | \mathbf{y})$. Uma combinação do amostrador de Gibbs com o algoritmo de Metropolis-Hastings foi implementada, permitindo gerar amostras das variáveis \mathbf{a} e σ_e^2 juntamente com amostras do sinal original \mathbf{x} . A estrutura do algoritmo é mostrada a seguir:

- Geração de valores iniciais para os parâmetros $\mathbf{x}^{(0)}$, $\mathbf{a}^{(0)}$ e $\sigma_e^{2(0)}$.
- Para j de 1 até N_{it}

1. $\mathbf{x}^{(j)} \sim p(\mathbf{x} | \mathbf{y}, \mathbf{a}^{(j-1)}, \sigma_e^{2(j-1)})$

2. $\mathbf{a}^{(j)} \sim p(\mathbf{a}|\mathbf{x}^{(j)}, \sigma_e^{2(j-1)})$
3. $\sigma_e^{2(j)} \sim p(\sigma_e^2|\mathbf{x}^{(j)}, \mathbf{a}^{(j)})$

Para obter cada distribuição condicional, precisamos apenas olhar para a distribuição conjunta da Equação (5.10) e descartar os termos não relacionados com a variável que queremos amostrar. Para os parâmetros do modelo AR, a amostragem é feita como na Subseção 4.1.1.

Devido à presença de termos de \mathbf{x} na CDF da Equação (5.10), a distribuição condicional resultante para o sinal original não é gaussiana. Por esse motivo, o algoritmo de Metropolis-Hastings foi a escolha natural para se extrair amostras dessa distribuição. Como distribuição proposta, escolhemos a distribuição de interesse retirando os termos correspondentes a amostras de \mathbf{x} na região saturada, o que resulta em:

$$\begin{aligned} q(\mathbf{x}) &= p_v(\mathbf{x}_c - \mathbf{y}_c)p(\mathbf{x}|\mathbf{a}, \sigma_v^2), \\ &= N(\mathbf{x}_c|\mathbf{y}_c, \sigma_v^2\mathbf{I}_{N_s})N(\mathbf{x}|\boldsymbol{\mu}_x, \mathbf{C}_x), \end{aligned} \quad (5.11)$$

em que N_s é o número de amostras não saturadas.

É possível mostrar, usando o mesmo argumento do Apêndice A.1, que o produto acima é uma gaussiana $q(\mathbf{x}^*) = N(\mathbf{x}^*|\boldsymbol{\mu}_p, \mathbf{C}_p)$, com matriz de covariância

$$\mathbf{C}_p = \left(\mathbf{C}_x^{-1} + \frac{1}{\sigma_v^2} \text{diag}\{1 - |i(1)|, \dots, 1 - |i(N)|\} \right), \quad (5.12)$$

e média

$$\boldsymbol{\mu}_p^T = \left(\boldsymbol{\mu}_x^T \mathbf{C}_x^{-1} + \mathbf{y}^T \frac{1}{\sigma_v^2} \text{diag}\{1 - |i(1)|, \dots, 1 - |i(N)|\} \right) \mathbf{C}_p^{-1}, \quad (5.13)$$

em que o elemento D_{nn} da matriz diagonal $D = \text{diag}\{1 - |i(1)|, \dots, 1 - |i(N)|\}$ é igual a zero quando n corresponde a uma amostra da região saturada ($i(n) = 1$ ou $i(n) = -1$) e igual a 1 quando n corresponde a uma amostra da região sem saturação ($i(n) = 0$).

Pela Equação (3.21), a probabilidade de aceitação para a transição de um estado $\mathbf{x}^{(i)}$ ao estado \mathbf{x}^* é dada por:

$$\alpha = \min \left(1, \frac{\prod_{k \in C_+} \phi \left(\frac{x^*(k) - c_+}{\sigma_v^2} \right) \prod_{k \in C_-} \phi \left(\frac{c_- - x^*(k)}{\sigma_v^2} \right)}{\prod_{k \in C_+} \phi \left(\frac{x^{(i)}(k) - c_+}{\sigma_v^2} \right) \prod_{k \in C_-} \phi \left(\frac{c_- - x^{(i)}(k)}{\sigma_v^2} \right)} \right). \quad (5.14)$$

5.3.2 Não-linearidade seguida de ruído

O sinal $y(n)$ é obtido a partir do sinal $x(n)$ da seguinte forma:

$$y(n) = \begin{cases} x(n) + v(n), & c_- \leq x(n) \leq c_+, \\ c_+ + v(n), & x(n) > c_+, \\ c_- + v(n), & x(n) < c_-. \end{cases} \quad (5.15)$$

5.3.2.1 Cálculo da distribuição *a posteriori*

Como o sinal saturado é contaminado com ruído, os limites de saturação c_+ e c_- são desconhecidos, e no contexto bayesiano, são tratados como variáveis aleatórias. Para o cálculo da verossimilhança, vamos usar a variável auxiliar $i(n)$, definida na Equação (5.2). Como os elementos de \mathbf{y} são independentes entre si se \mathbf{x} for conhecido, podemos concluir que a verossimilhança é dada por:

$$p(\mathbf{y}|\mathbf{x}, \mathbf{i}, c_+, c_-, \mathbf{a}, \sigma_e^2) = p_v(\mathbf{y}_c - \mathbf{x}_c) p_v(\mathbf{y}_+ - \mathbf{c}_+) p_v(\mathbf{y}_- - \mathbf{c}_-). \quad (5.16)$$

Pela regra de Bayes, calculamos a distribuição *a posteriori*:

$$p(\mathbf{x}, \mathbf{i}, c_+, c_-, \mathbf{a}, \sigma_e^2 | \mathbf{y}) \propto p_v(\mathbf{y}_c - \mathbf{x}_c) p_v(\mathbf{y}_+ - \mathbf{c}_+) p_v(\mathbf{y}_- - \mathbf{c}_-) p(\mathbf{i}) p(c_+) p(c_-) p(\mathbf{a}, \sigma_e^2). \quad (5.17)$$

5.3.2.2 Algoritmo proposto

Para obtenção de amostras dessa distribuição adotamos novamente o amostrador de Gibbs:

- Geração de valores iniciais para os parâmetros $\mathbf{x}^{(0)}$, $\mathbf{i}^{(0)}$, $c_+^{(0)}$, $c_-^{(0)}$, $\mathbf{a}^{(0)}$ e $\sigma_e^{2(0)}$.
- Para j de 1 até N_{it}

1. $(\mathbf{x}^{(j)}, \mathbf{i}^{(j)}) \sim p(\mathbf{x}, \mathbf{i} | \mathbf{y}, c_+^{(j-1)}, c_-^{(j-1)}, \mathbf{a}^{(j-1)}, \sigma_e^{2(j-1)})$
2. $(c_+^{(j)}, c_-^{(j)}) \sim p(c_+, c_- | \mathbf{y}, \mathbf{i}^{(j)})$
3. $\mathbf{a}^{(j)} \sim p(\mathbf{a} | \mathbf{x}^{(j)}, \sigma_e^{2(j-1)})$
4. $\sigma_e^{2(j)} \sim p(\sigma_e^2 | \mathbf{x}^{(j)}, \mathbf{a}^{(j)})$

Amostragem de \mathbf{x} e \mathbf{i} : Para essa tarefa, adotamos uma estratégia similar à de [2], em que, no contexto de localização de ruído impulsivo, essas variáveis são amostradas conjuntamente, de forma sequencial, o que provou ser uma forma eficiente de se acelerar a convergência do algoritmo.

No nosso caso, para evitar trabalhar com integrais multivariáveis, decidimos amostrar $x(n)$ e $i(n)$ para cada $n \in \{1, \dots, N\}$, em vez de amostrar os vetores completos. A distribuição condicional conjunta para essas variáveis é dada por:

$$p(i(n), x(n) | \mathbf{y}, \mathbf{i}_{-n}, \mathbf{x}_{-n}) \propto p(x(n) | \mathbf{y}, \mathbf{i}, \mathbf{x}_{-n}) p(i(n)), \quad (5.18)$$

com

$$\begin{aligned} p(x(n) | \mathbf{y}, \mathbf{i}, \mathbf{x}_{-n}) &\propto p_v(y(n) - c[x(n), i(n)]) p(x(n) | \mathbf{x}_{-n}) \\ &= \begin{cases} p_v(x(n) - y(n)) p(x(n) | \mathbf{x}_{-n}), & i(n) = 0 \text{ e } c_- \leq x(n) \leq c_+ \\ p_v(y(n) - c_+) p(x(n) | \mathbf{x}_{-n}), & i(n) = 1 \text{ e } x(n) > c_+ \\ p_v(y(n) - c_-) p(x(n) | \mathbf{x}_{-n}), & i(n) = -1 \text{ e } x(n) < c_- \\ 0, & \text{nas demais combinações.} \end{cases} \quad (5.19) \end{aligned}$$

em que $c[x(n)]$ indica a função não-linear que gera o sinal saturado a partir do sinal original.

Esta expressão indica que a distribuição resultante será uma gaussiana na região em que $i(n)$ e $x(n)$ forem compatíveis, e zero em caso contrário. A distribuição de $x(n)$ condicionada às demais amostras de \mathbf{x} pode ser obtida como um caso particular da Equação (5.5), resultando em:

$$p(x(n) | \mathbf{x}_{-n}) = \text{N} \left(x(n) \left| \frac{\mathbf{A}_{-n}^T \mathbf{A}_{-n} \mathbf{x}_{-n}}{1 + \mathbf{a}^T \mathbf{a}}, \frac{\sigma_e^2}{1 + \mathbf{a}^T \mathbf{a}} \right. \right), \quad (5.20)$$

Usando a fórmula do produto de gaussianas do Apêndice(A.1), obtemos:

$$p_v(x(n) - y(n))p(x(n)|\mathbf{x}_{-n}) = N(x(n)|\mu_x(n), \sigma_x^2(n)), \quad (5.21)$$

com

$$\mu_x(n) = \sigma_x^2(n) \left(\frac{y(n)}{\sigma_v^2} + \frac{\mathbf{A}_{-n}^T \mathbf{A}_{-n} \mathbf{x}_{-n}}{\sigma_e^2} \right), \quad (5.22)$$

$$\sigma_x^2(n) = \frac{1}{\frac{1}{\sigma_v^2} + \frac{1 + \mathbf{a}^T \mathbf{a}}{\sigma_e^2}}. \quad (5.23)$$

Para amostragem da distribuição conjunta de $i(n)$ e $x(n)$, basta obter sequencialmente amostras de $p(i(n)|\mathbf{y}, \mathbf{i}_{-n}, \mathbf{x})$ e $p(x(n)|\mathbf{y}, \mathbf{i}, \mathbf{x}_{-n})$. A primeira distribuição pode ser obtida a partir da Equação (5.18), integrando-se $x(n)$, isto é:

$$\begin{aligned} P(i(n)|\mathbf{y}, \mathbf{i}_{-n}, \mathbf{x}) &= kP(i(n)) \int_{x(n)} p_v(y(n) - c[x(n)])p(x(n)|\mathbf{x}_{-n})dx(n) \\ &= \begin{cases} kP(i(n))p_v(y(n) - c_+) \int_{c_+}^{\infty} p(x(n)|\mathbf{x}_{-n}), & i(n) = 1 \\ kP(i(n)) \int_{c_-}^{c_+} p_v(x(n) - y(n))p(x(n)|\mathbf{x}_{-n}), & i(n) = 0 \\ kP(i(n))p_v(y(n) - c_-) \int_{-\infty}^{c_-} p(x(n)|\mathbf{x}_{-n}), & i(n) = -1, \end{cases} \end{aligned}$$

em que k é uma constante de proporcionalidade que garante que o somatório das probabilidades *a posteriori* de $i(n)$ seja 1.

Seja $P(i) = p(i(n)|\mathbf{y}, \mathbf{i}_{-n}, \mathbf{x})$. Para se obter uma amostra de $i(n)$, basta calcular $P(i(n))$ para os três valores possíveis de $i(n)$. Em seguida, gera-se uma variável aleatória r uniforme entre 0 e 1. Por fim, o valor de $i(n)$ é obtido de acordo com o procedimento abaixo:

1. Se $r \leq P(-1)$, $i(n) = -1$;
2. Se $P(-1) < r < P(-1) + P(0)$, $i(n) = 0$;
3. Se $r \geq P(-1) + P(0)$, $i(n) = 1$.

Uma vez obtido $i(n)$, tiramos uma amostra de $x(n)$, cuja distribuição será uma gaussiana truncada, com parâmetros e região de suporte variando conforme o valor de $i(n)$, conforme descrito na Equação (5.19).

Amostragem de c_+ e c_- : Para amostragem de c_+ , basta olharmos para a distribuição *a posteriori* conjunta e perceber que os termos dependentes de c_+

resultam em:

$$\begin{aligned}
p(c_+|\mathbf{y}) &\propto \prod_{n \in C_+} p_v(y(n) - c_+)p(c_+), \\
&\propto \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{n \in C_+} (y(n) - c_+)^2 \right\} p(c_+). \tag{5.24}
\end{aligned}$$

Embora sabendo que c_+ deva ser positivo, decidimos escolher, por conveniência matemática, uma priori conjugada gaussiana de média $\mu_+ = 0$ e variância $\sigma_+^2 = 100$. Essa priori bastante vaga não deverá prejudicar o resultado, porque espera-se ter uma quantidade de dados bastante alta, implicando numa verossimilhança bastante informativa. Substituindo em (5.24), podemos mostrar que o resultado é:

$$p(c_+|\mathbf{y}) = \text{N} \left(c_+ \left| \frac{\sum_{n \in C_+} y(n)}{\sigma_v^2} \left(\frac{N_{C_+}}{\sigma_v^2} + \frac{1}{\sigma_+^2} \right)^{-1}, \left(\frac{N_{C_+}}{\sigma_v^2} + \frac{1}{\sigma_+^2} \right)^{-1} \right. \right). \tag{5.25}$$

$$p(c_-|\mathbf{y}) = \text{N} \left(c_- \left| \frac{\sum_{n \in C_-} y(n)}{\sigma_v^2} \left(\frac{N_{C_-}}{\sigma_v^2} + \frac{1}{\sigma_-^2} \right)^{-1}, \left(\frac{N_{C_-}}{\sigma_v^2} + \frac{1}{\sigma_-^2} \right)^{-1} \right. \right). \tag{5.26}$$

5.3.3 Resultados

Para ilustrar o funcionamento dos dois algoritmos, aplicamo-los a sinais gerados pelo mesmo modelo da Subseção 4.1.1. Para o caso A, inserimos ruído branco aditivo a 30 dB SNR e limitamos o resultado em pontos máximo e mínimo, como mostrado na Figura 5.9. Iniciando os parâmetros a partir do método descrito em [23], rodamos o amostrador de Gibbs por 10000 iterações, sendo as amostras das 3000 primeiras mostradas na Figura 5.10 e 5.11. Devido à estimativa inicial ruim, o algoritmo demora a convergir para a região esperada. Com base nas 8000 últimas iterações, geramos os histogramas da excitação σ_e^2 , dos elementos de \mathbf{a} e da amostra 192 do sinal original, mostrados respectivamente nas Figuras 5.12, 5.13 e 5.14. Comparando esses histogramas com os correspondentes na seção anterior, podemos ver que o ruído provoca aumento na variância das amostras, como esperado, mas as distribuições obtidas se mantêm compatíveis com os parâmetros corretos.

As Figuras de 5.15 a 5.19 mostram gráficos correspondentes para o caso A,

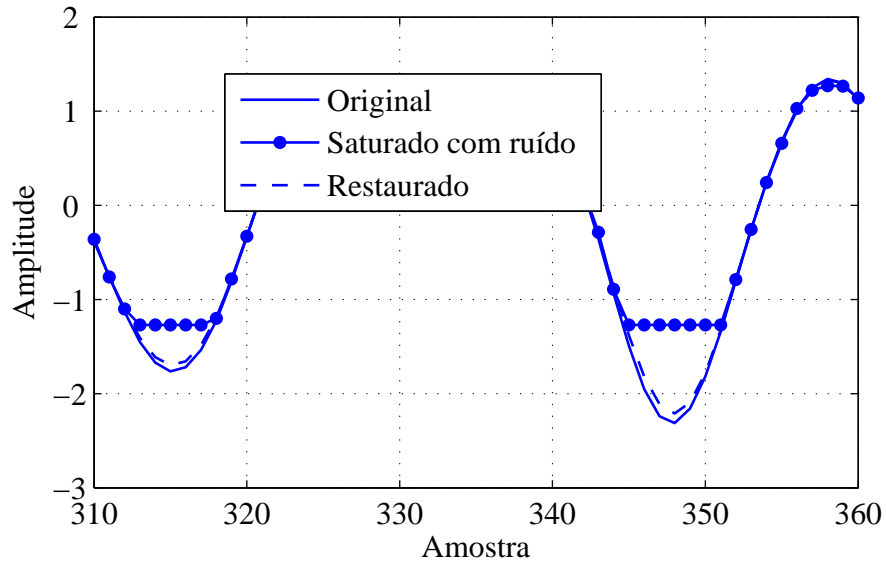


Figura 5.9: Comparação entre sinais original, saturado com ruído (antes da saturação) e restaurado pelo algoritmo proposto.

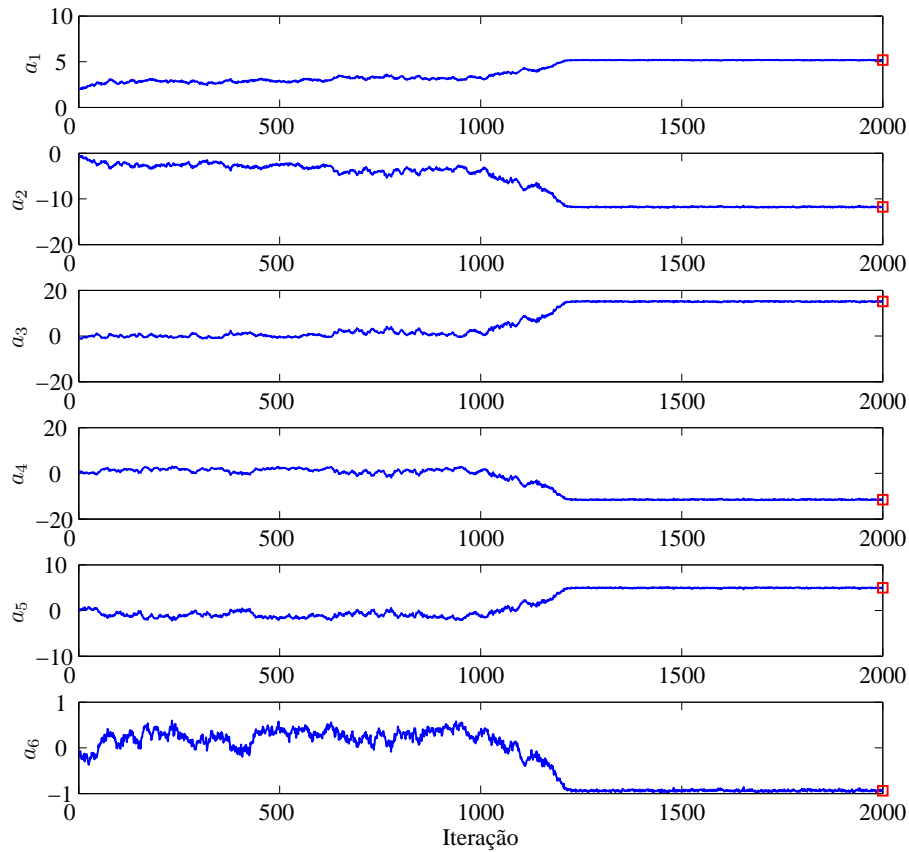


Figura 5.10: Evolução dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).

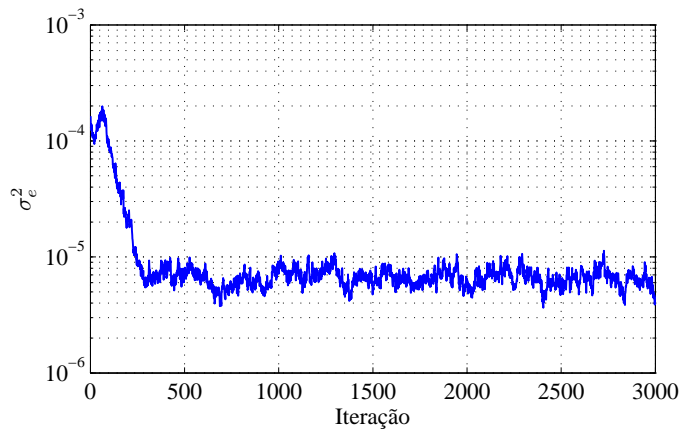


Figura 5.11: Evolução da variância de excitação do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).

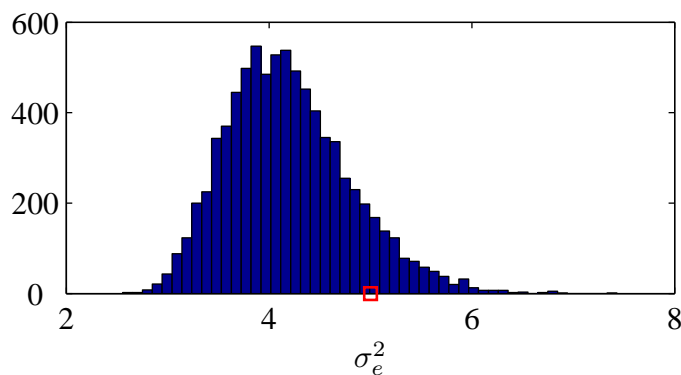


Figura 5.12: Histograma da variância de excitação do modelo AR obtida através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).

resultantes do algoritmo proposto aplicado a sinais com propriedades semelhantes às do caso A. Nas Figuras 5.20 e 5.21, mostramos a evolução das amostras das variáveis c_+ e c_- , respectivamente, e nas Figuras 5.22 e 5.23 seus histogramas. A Figura 5.24 mostra o histograma da amostra 120 do sinal restaurado.

Ambos os algoritmos funcionaram de acordo com esperado, sendo capazes de aumentar a SNR do sinal em até 10 dB, dentre os sinais artificiais e reais em que os testamos. A complexidade dos algoritmos é elevada, pois em cada iteração, realiza a amostragem de uma gaussiana com número de elementos igual ao tamanho do bloco (800 amostras em nossa implementação). Além disso, com a inicialização que adotamos, o número de iterações necessárias para convergência é da ordem dos milhares.

Adotamos duas estratégias para reduzir a complexidade. A primeira foi di-

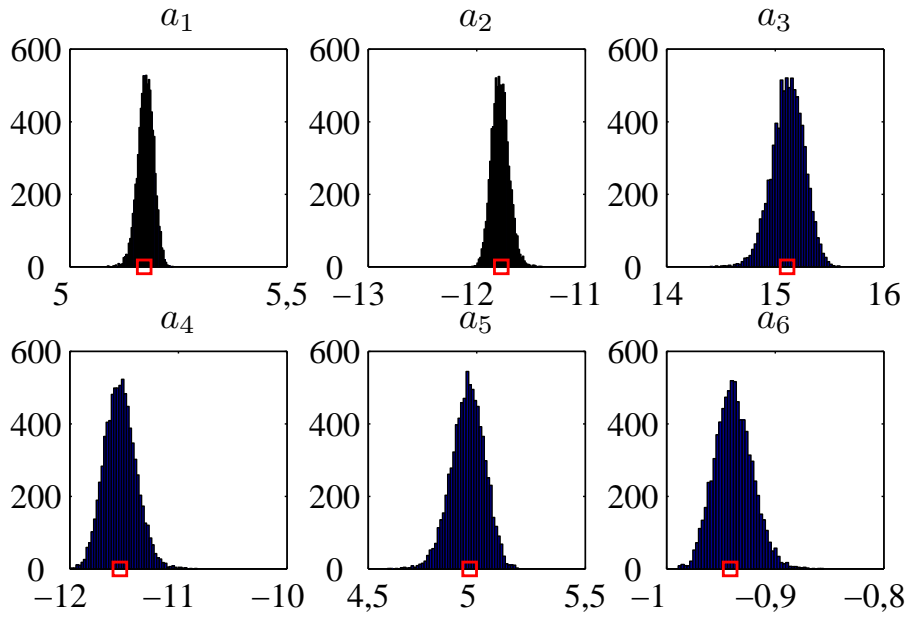


Figura 5.13: Histogramas de cada elemento dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).

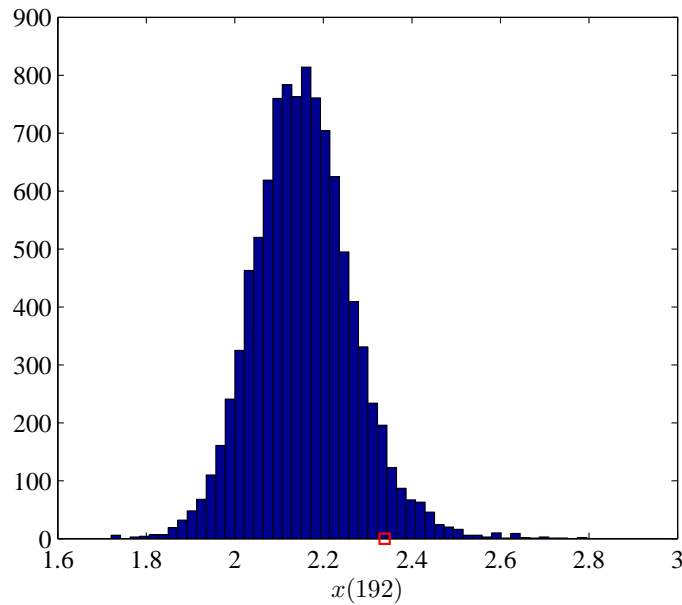


Figura 5.14: Histograma da amostra 192 do sinal restaurado obtido através do algoritmo proposto com base em um sinal saturado com ruído (antes da saturação).

vidir as amostras de cada bloco do sinal em sub-blocos e realizar a amostragem de cada sub-bloco sequencialmente. Assim, a complexidade que era de $O(N^3)$ passa a ser $\sum_{i=1}^Q O(P_i^3)$, onde Q é o número de sub-blocos em cada bloco e N_i é o seu número de amostras. O efeito dessa mudança é aumentar a correlação na cadeia de

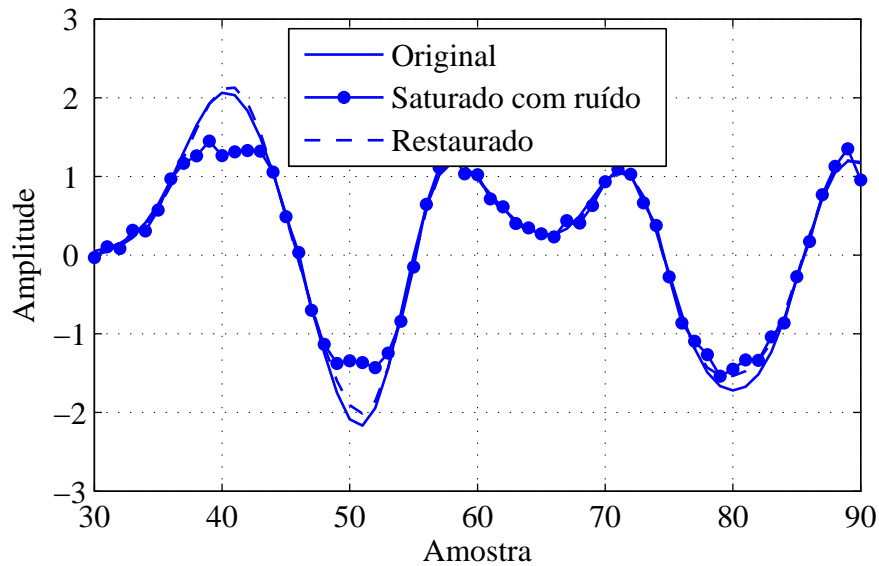


Figura 5.15: Comparação entre sinais original, saturado com ruído (depois da saturação) e restaurado.

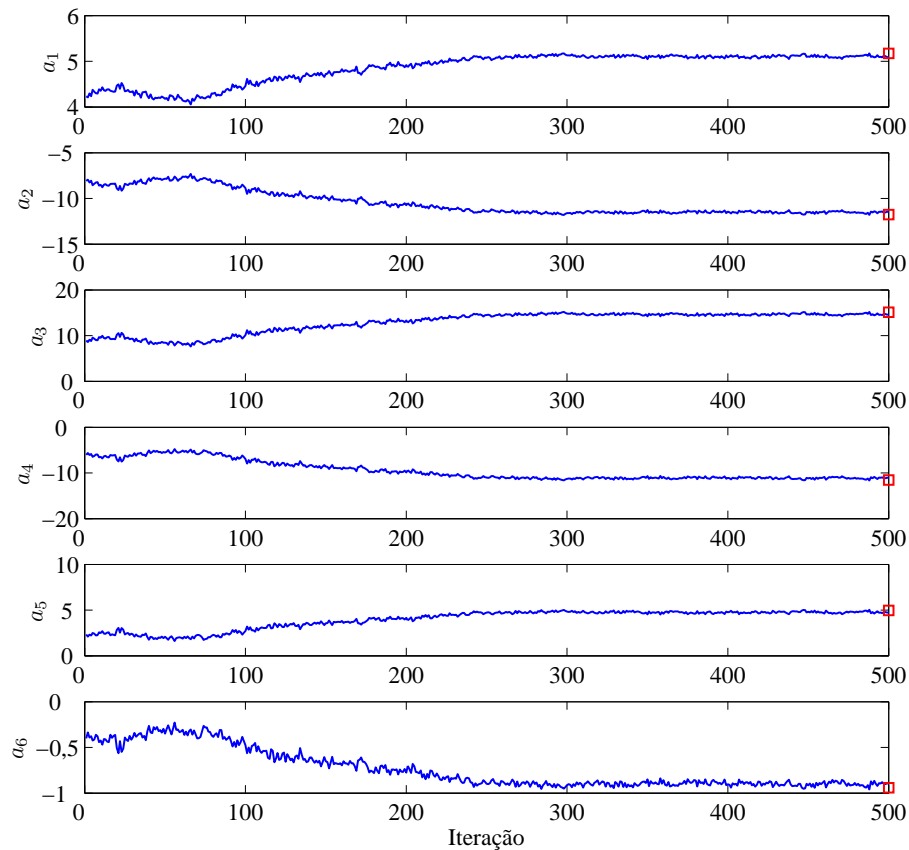


Figura 5.16: Evolução dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).

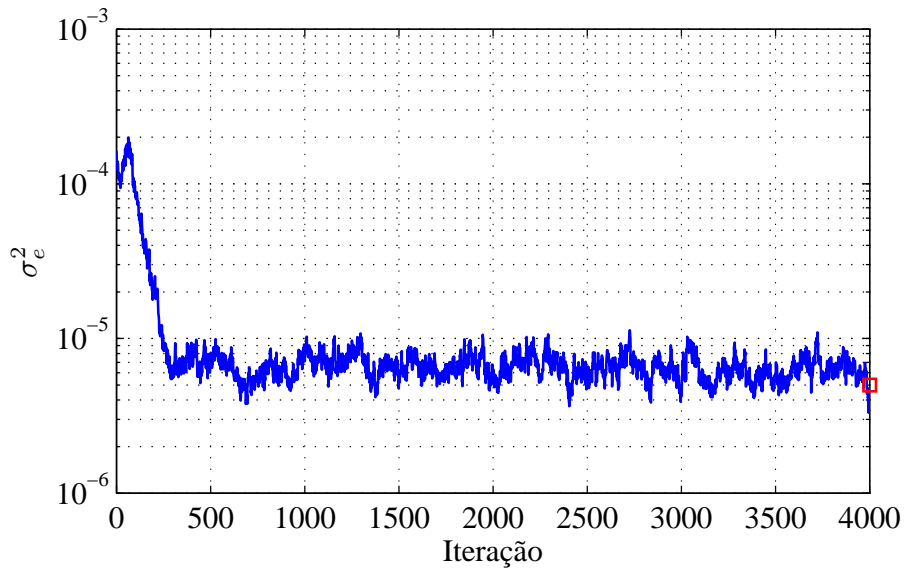


Figura 5.17: Evolução da variância de excitação do modelo AR obtido através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).

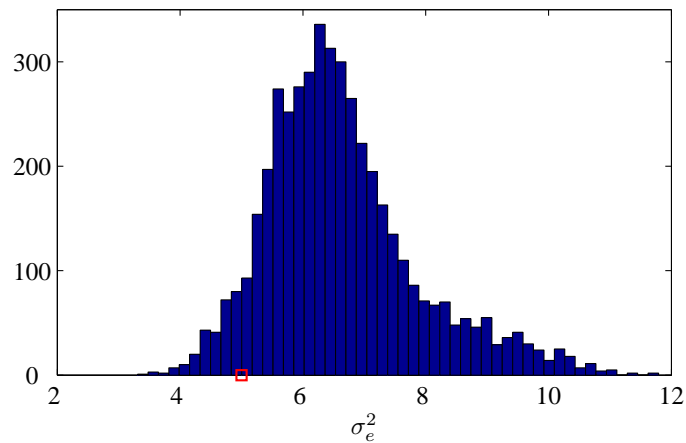


Figura 5.18: Histograma da variância excitação do modelo AR obtida através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).

Markov e assim aumentar o número de iterações necessárias para a convergência. Verificamos em simulações que adotar cada semiciclo do sinal como sub-bloco é um bom compromisso, além de ser a escolha natural para sinais com saturação.

A segunda estratégia é inicializar os parâmetros do modelo AR para um certo bloco com os valores estimados no bloco anterior, explorando o fato de que esses parâmetros variam pouco de bloco para bloco. Para partir de estimativas de boa qualidade, usamos um número alto de iterações nos primeiros blocos e o diminuimos gradualmente.

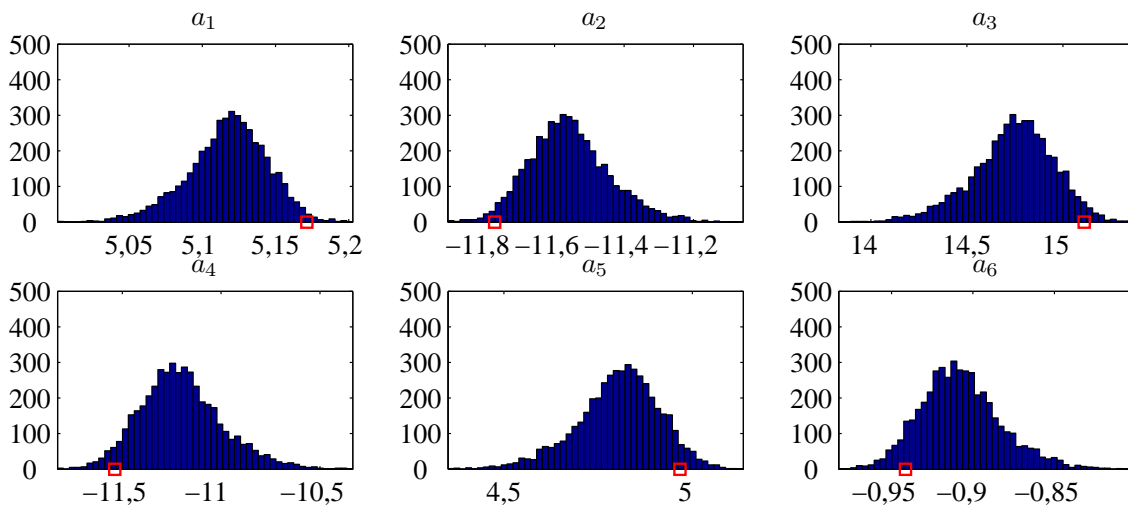


Figura 5.19: Histogramas de cada elemento dos coeficientes do modelo AR obtidos através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).

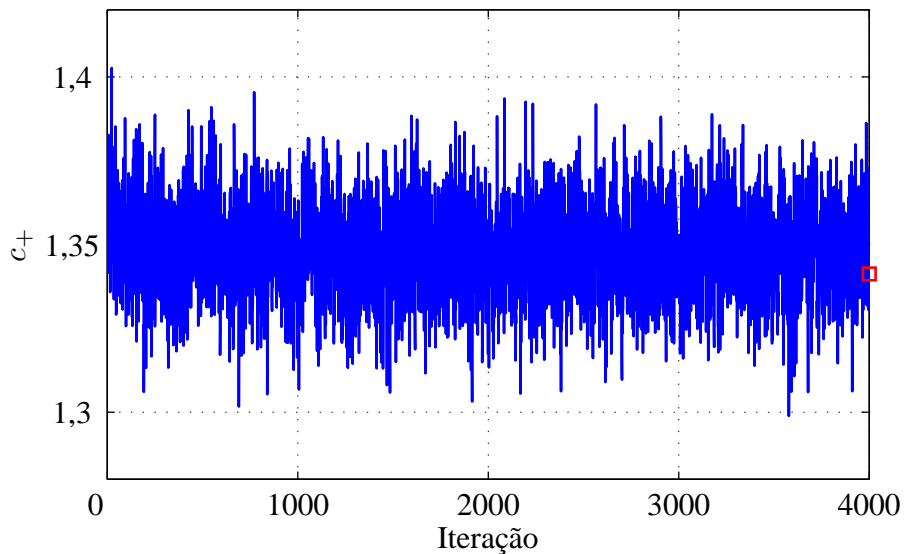


Figura 5.20: Evolução do limite superior de amplitude c_+ para um sinal com saturação e ruído (após a saturação).

5.4 Conclusões

Neste capítulo propusemos algoritmos para lidar com o problema de limitação de amplitude presentes em sinais de áudio modelados como processos AR, para três casos distintos, sinal limitado sem ruído, sinal com ruído inserido antes da limitação e sinal com ruído inserido após a limitação. No primeiro caso, o algoritmo proposto compara-se favoravelmente em relação à solução de referência baseada na

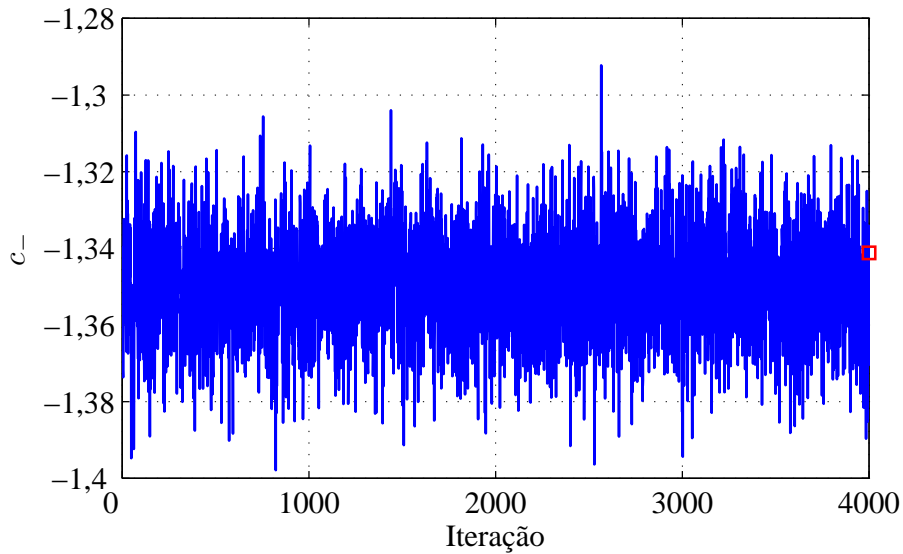


Figura 5.21: Evolução de amostras do limite inferior de amplitude c_- obtidas a partir do algoritmo proposto com base em um sinal com saturação e ruído (após a saturação).

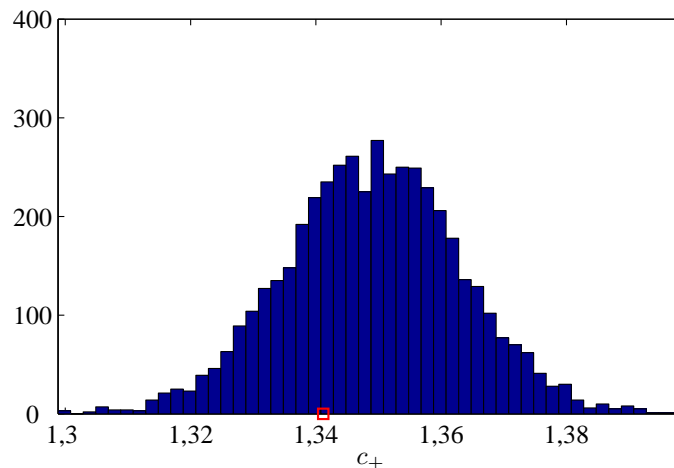


Figura 5.22: Histograma do limite superior c_+ obtida com o método proposto com base em um sinal com saturação e ruído (após a saturação).

minimização do erro médio quadrático, porém o ganho tende a ser modesto para as condições encontradas na prática. Nos casos com ruído, propusemos métodos bayesianos capazes de aumentar significativamente a SNR dos sinais. Propusemos ainda variações dos algoritmos para reduzirmos sua complexidade, tornando-os viáveis em aplicações práticas. Exemplos da aplicação dos algoritmos deste capítulo a sinais reais são apresentados no Apêndice B.

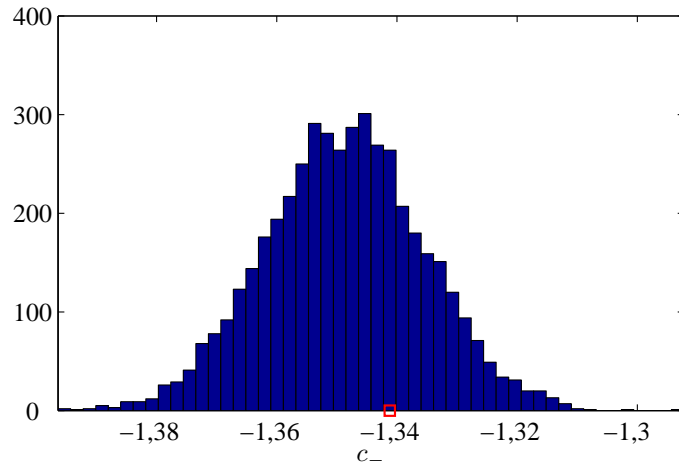


Figura 5.23: Histograma do limite superior c_- , obtida com o método proposto com base em um sinal com saturação e ruído (após a saturação).

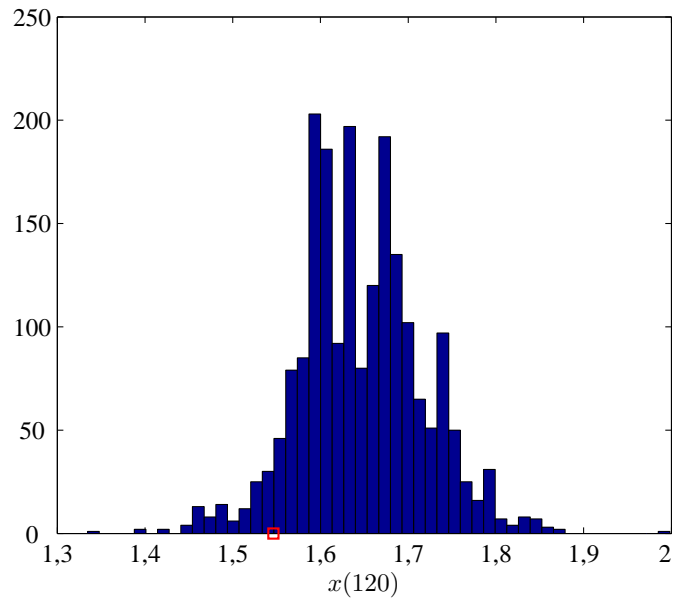


Figura 5.24: Histograma da amostra 120 do sinal restaurado obtidas através do algoritmo proposto com base em um sinal saturado com ruído (depois da saturação).

Capítulo 6

Distorção Não-Linear Invertível

Sem Memória

Neste capítulo trataremos do problema de restauração de sinais com distorção não-linear invertível sem memória, caracterizada pelo fato de que o sinal distorcido num certo instante n depende apenas do sinal original neste mesmo instante. No contexto de áudio, esse modelo descreve a saturação gradual, representado na Figura 6.1. O sinal original $x(n)$ é gerado por um filtro só-pólos $A(z)$ suposto invariante no tempo e excitado por ruído branco gaussiano $e(n)$, suposto estacionário no sentido amplo em trechos curtos do sinal. O sinal saturado é obtido após o sinal original passar pela distorção não-linear sem memória representada pela função $f[\cdot]$.

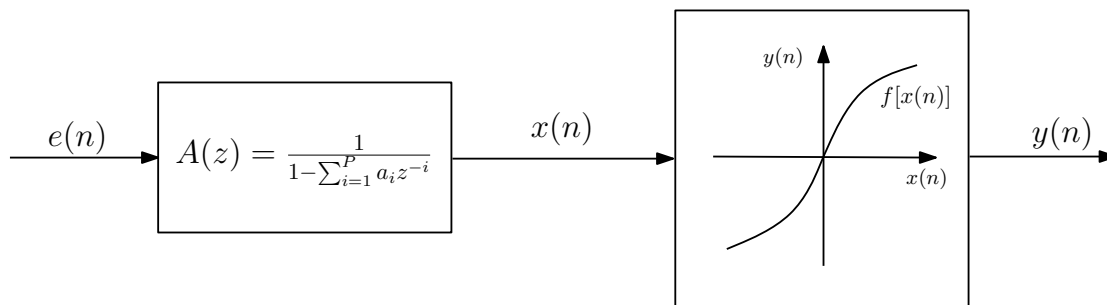


Figura 6.1: Diagrama de blocos de sinal distorcido com saturação gradual, representando uma distorção não-linear sem memória.

O formato da curva $f[\cdot]$ determina o grau em que o sinal é distorcido. Nesse trabalho, será assumido que a curva é antissimétrica, sua derivada é sempre não-negativa e sua derivada segunda é sempre negativa ou nula para $x(n) > 0$ e positiva

ou nula para $x(n) < 0$. Além disso, para x em torno da origem $f(x)$ pode ser aproximada por uma reta de inclinação unitária. Como veremos mais adiante, é conveniente trabalhar com a inversa da função $f[.]$, a qual denominamos $g(y) = f^{-1}(y)$. As propriedades descritas acima podem ser equivalentemente escritas em termos da função $g(\cdot)$, da seguinte forma:

$$g(y) = -g(-y), \forall y \in \mathbb{R} \quad (6.1)$$

$$g'(y) \geq 0, \forall y \in \mathbb{R} \quad (6.2)$$

$$g'(0) = 0 \quad (6.3)$$

$$g''(y) \geq 0, \forall y \in \mathbb{R}^+ \quad (6.4)$$

$$g''(y) \leq 0, \forall y \in \mathbb{R}^- . \quad (6.5)$$

Quando o ruído de medição é desprezível, o problema de restauração se resume ao da estimação da curva não-linear $g(y)$, a partir da qual o sinal original pode ser obtido de forma trivial. Como veremos mais adiante, a presença de ruído provocaria polarização na estimação de $g(y)$ e a aplicação de $g(y)$ ao sinal ruidoso causaria amplificação do ruído. Para evitar esses efeitos indesejáveis, propomos um tratamento separado para o caso em que o ruído está presente.

A idéia principal das soluções propostas consiste em escrever a curva não-linear através de sua expansão em bases polinomiais e formular o problema de restauração como o de estimação dos coeficientes dessa expansão. Uma escolha particular é expandir $g(\cdot)$ em série de Taylor em torno da origem, como feito, por exemplo em [29–31], com seus coeficientes desconhecidos:

$$x(n) = y(n) + m_1 y(n)^3 + m_2 y(n)^5 + \dots + m_M y(n)^{2M+1}, \quad (6.6)$$

em que os coeficientes pares foram feitos iguais a zero para atender $g(y) = -g(-y)$ e o coeficiente do termo linear foi feito igual a 1 para que $g'(0) = 1$. A ordem do modelo M é desconhecida, podendo ser estimada a partir dos dados ou escolhida alta o suficiente para garantir uma boa aproximação da curva.

Vetorialmente, podemos reescrever a equação acima como:

$$\mathbf{x} = \mathbf{y} + \mathbf{Y}\mathbf{m}, \quad (6.7)$$

em que

$$\mathbf{Y} = \begin{bmatrix} y(1)^3 & y(1)^5 & \dots & y(1)^M \\ y(2)^3 & y(2)^5 & \dots & y(2)^M \\ \vdots & \vdots & \ddots & \vdots \\ y(N)^3 & y(N)^5 & \dots & y(N)^M \end{bmatrix}, \quad (6.8)$$

$$\mathbf{m} = [m_1, \dots, m_M]^T. \quad (6.9)$$

Nas próximas seções vamos apresentar algoritmos para estimação dos coeficientes dessa expansão, com diferentes graus de complexidade e acurácia. Propomos também variações desses algoritmos para lidar com ruído aditivo incidente no sinal saturado, além de uma formulação alternativa do problema em que a curva não-linear é aproximada por uma sequência de retas com inclinações desconhecidas.

O capítulo está organizado da seguinte forma. Na Seção 6.1, descrevemos brevemente alguns dos trabalhos existentes na literatura relacionados a distorções não-lineares sem memória, comparando-as com as soluções propostas nesta tese. Na Seção 6.2, apresentamos procedimentos simples para estimativa inicial dos parâmetros do modelo, necessários para os algoritmos mais sofisticados do restante do capítulo. Na Seção 6.3, apresentamos o procedimento para o cálculo da verossimilhança dos parâmetros do modelo dado o sinal degradado, assumido sem ruído. Em seguida, na Seção 6.4, mostramos como os métodos do gradiente e de Gauss-Newton podem ser usados para maximização da verossimilhança. Na Seção 6.5, descrevemos como o problema pode ser tratado do ponto de vista de inferência bayesiana, apresentando um algoritmo baseado em MCMC para estimação da distribuição *a posteriori* dos parâmetros da curva não-linear. Na seção seguinte, apresentamos uma abordagem alternativa do problema, baseada na modelagem da curva não-linear por uma aproximação linear por partes, que tem vantagens do ponto de vista de complexidade computacional. O tratamento de ruído aditivo é feito na Seção 6.7, na qual propõem-se modificações nos algoritmos das seções anteriores, aumentando sua robustez em casos práticos. Por fim, na Seção 6.8, apresentamos as conclusões deste capítulo.

6.1 Trabalhos anteriores

O tratamento de distorções não-lineares invertível sem memória em sinais de áudio aparece na literatura em diversos contextos. Em [32], é proposta uma abordagem simples para remoção de saturação baseada em equalização por histograma. Um histograma é obtido para o sinal distorcido e comparado com o histograma estimado para o sinal original a partir de uma base para sinais sem degradação. Comparando os dois histogramas, a curva de saturação pode ser estimada, e sua inversa usada para se obter uma aproximação do sinal original. Esse método depende de que a base usada para se obter o histograma para o sinal original seja representativa do tipo de sinal que está sendo tratado. Essa hipótese não se aplica a muitos sinais de música, devido à sua alta variabilidade.

A idéia de comparar histogramas é a base do método proposto na Seção 6.2. A vantagem de nossa abordagem é a generalidade: sua única hipótese é que o sinal é aproximadamente gaussiano em trechos curtos. Conforme discutido no início do Capítulo anterior, sabe-se que essa é uma hipótese bastante robusta para muitos sinais de áudio e voz.

Com foco em distorções não-lineares geradas em alto-falantes do tipo corneta, em [33] é proposto um método de correção da não-linearidade que leva em conta as características físicas do alto-falante. Os métodos propostos neste capítulo, em princípio, se aplicam a este tipo particular de defeito, mas são mais gerais pois não assumem um formato específico para a curva não-linear.

6.2 Estimativa inicial

Nesta seção apresentamos algumas soluções simples para a obtenção de estimativas iniciais para os parâmetros da curva não-linear e do modelo AR, que servirão como ponto de partida para os algoritmos mais sofisticados das seções seguintes.

Propomos dois métodos que se baseiam na gaussianidade da excitação $e(n)$ e no fato de que a curva é aproximadamente a identidade em torno da origem. A ideia em ambos os métodos é determinar a curva não-linear estática que restaura a gaussianidade do sinal. Isso nos permitirá ter uma estimativa não-paramétrica da função não-linear $g(\cdot)$, definida nos pontos correspondentes a amostras de $y(n)$.

Com base nesses pontos da curva, estimam-se seus parâmetros através do método dos mínimos quadrados.

6.2.1 Método A

A idéia dessa solução é estimar os coeficientes do modelo AR com base nas amostras de mais baixa amplitude de $y(n)$, assumidas não distorcidas, e em seguida obter uma estimativa $\hat{\mathbf{x}}$ para o sinal original \mathbf{x} usando a solução proposta no capítulo anterior para o problema de saturação radical (caso sem ruído).

Podemos particionar um sinal \mathbf{y} corrompido por saturação gradual em duas partes: \mathbf{y}_l e \mathbf{y}_n , sendo a primeira correspondente a regiões aproximadamente lineares e a segunda a regiões com maior nível de distorção. Assumimos que $\mathbf{x}_l \approx \mathbf{y}_l$ e $\mathbf{x}_n \leq \mathbf{y}_n$.

Na prática, escolhemos um limiar de amplitude tal que as amostras menores (em módulo) que esse limiar são consideradas corretas, isto é, formam \mathbf{y}_l . Se o limiar é baixo demais, amostras corretas deixam de ser usadas para estimação das amostras distorcidas; por outro lado, um limiar alto demais faz com que amostras distorcidas sejam consideradas corretas, o que prejudica a estimação das amostras consideradas distorcidas. Verificamos na prática que um limiar de $0,3 \times \max |\mathbf{y}|$ é um bom compromisso.

Uma vez obtida a estimativa $\hat{\mathbf{x}}$ de acordo com o procedimento acima, podemos estimar a curva de saturação através das densidades de probabilidade acumuladas dos sinais $\hat{\mathbf{x}}$ e \mathbf{y} , denominadas $F_{\hat{X}}(\hat{x})$ e $F_Y(y)$, respectivamente, que podem ser aproximadamente calculadas pelas fórmulas abaixo:

$$\hat{F}_Y(y_i) = \frac{\#\{n, y(n) < y_i\}}{N}, \quad i = \{1, \dots, N\}, \quad (6.10)$$

$$F_{\hat{X}}(\hat{x}_i) = \frac{\#\{n, \hat{x}(n) < \hat{x}_i\}}{N}, \quad i = \{1, \dots, N\}. \quad (6.11)$$

Definindo $\hat{g}(y)$ como a função que mapeia $y(n)$ em $\hat{x}(n)$, podemos escrever:

$$\hat{x} = \hat{g}(y) = F_{\hat{X}}^{-1}(F_Y(y)). \quad (6.12)$$

Com base nessa curva não-paramétrica, podem-se estimar os coeficientes da expansão de Taylor, como explicado na Seção 6.2.3.

6.2.2 Método B

Como a excitação $e(n)$ é, por hipótese, gaussiana, o sinal $x(n)$, obtido a partir de $e(n)$ por transformação linear, também o é. A idéia dessa solução é estimar a curva não-linear $g(\cdot)$ como a transformação em \mathbf{y} que restaura a gaussianidade de \mathbf{x} . Como a variância de \mathbf{x} é desconhecida, ela é estimada assumindo-se, como no Método A que numa certa região em torno da origem $g(y)$ é aproximadamente a identidade. O algoritmo é resumido pelas seguintes etapas:

1. Estimamos a distribuição cumulativa de \mathbf{y} com base em:

$$\hat{F}(y_i) = \frac{\#\{n, y(n) < y_i\}}{N}, \quad i = \{1, \dots, N\}, \quad (6.13)$$

em que o símbolo $\#\{A\}$ indica a cardinalidade do conjunto A .

2. Calculamos a inclinação θ da curva $\hat{F}(y_i)$ na origem, que aproxima a densidade de probabilidade $f_X(0)$.
3. Estimamos a variância do sinal original pela seguinte expressão:

$$\sigma_x^2 = \frac{1}{2\pi\theta^2}. \quad (6.14)$$

4. Obtemos estimativas de pontos da curva não-linear através de:

$$\hat{x} = \hat{g}(y_i) = \Phi^{-1} \left(\hat{F}(y_i), \sigma_x^2 \right), \quad (6.15)$$

em que $\Phi(x, \sigma^2)$ representa a densidade de probabilidade cumulativa da gaussiana de média zero e variância σ^2 , calculada no ponto x .

O passo 3 pode ser entendido através da expressão da densidade de probabilidade gaussiana na origem, que é dada por:

$$\hat{f}_X(0) = \frac{1}{\sqrt{2\pi\sigma_x^2}} = \theta. \quad (6.16)$$

Resolvendo essa equação para σ_x^2 , obtemos a fórmula do passo 3.

6.2.3 Estimativa dos coeficientes

Uma vez obtidos pontos da curva $\hat{g}(y)$ por um dos dois métodos propostos, os coeficientes $\mathbf{m} = [m_1 \ m_2 \ \dots \ m_M]^T$ podem ser estimados através da minimização do erro médio quadrático, de acordo com o procedimento a seguir:

1. Escolhe-se um conjunto de K pares $(y_i, \hat{g}(y_i))$;
2. Forma-se a matriz \mathbf{Y} definida na Equação (6.8);
3. Determina-se o vetor de parâmetros \mathbf{m} através de:

$$\mathbf{m} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T (\hat{\mathbf{x}} - \mathbf{y}), \quad (6.17)$$

em que $\hat{\mathbf{x}} = [\hat{g}(y_1) \ \hat{g}(y_2) \ \dots \ \hat{g}(y_K)]^T$ e $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_K]^T$.

6.2.4 Simulações

Apresentamos simulações para ilustrar os métodos propostos. O sinal obtido com o modelo AR da Subseção 4.1.1 é distorcido com uma curva determinada pelos coeficientes $m_1 = 5$ e $m_2 = 30$. A Figura 6.2 mostra curvas do sinal distorcido, do sinal restaurado pelo método A e do sinal restaurado pelo método B, todas contra o sinal original, além da curva ideal (identidade), que corresponde à ausência de distorção. Observamos que ambos os métodos reduzem sensivelmente a não-linearidade do sinal, sendo que o método B gera um resultado melhor.

6.3 Cálculo da verossimilhança

Nesta seção, calculamos a verossimilhança dos parâmetros desconhecidos do modelo, isto é \mathbf{m} , \mathbf{a} e σ_e^2 . Começamos pela distribuição de $e(n)$ que, como afirmado no Capítulo 4, é dada por uma gaussiana de média zero e variância σ_e^2 :

$$p(\mathbf{e}|\sigma_e^2) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}} \right)^N \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{n=1}^N e^2(n) \right\}. \quad (6.18)$$

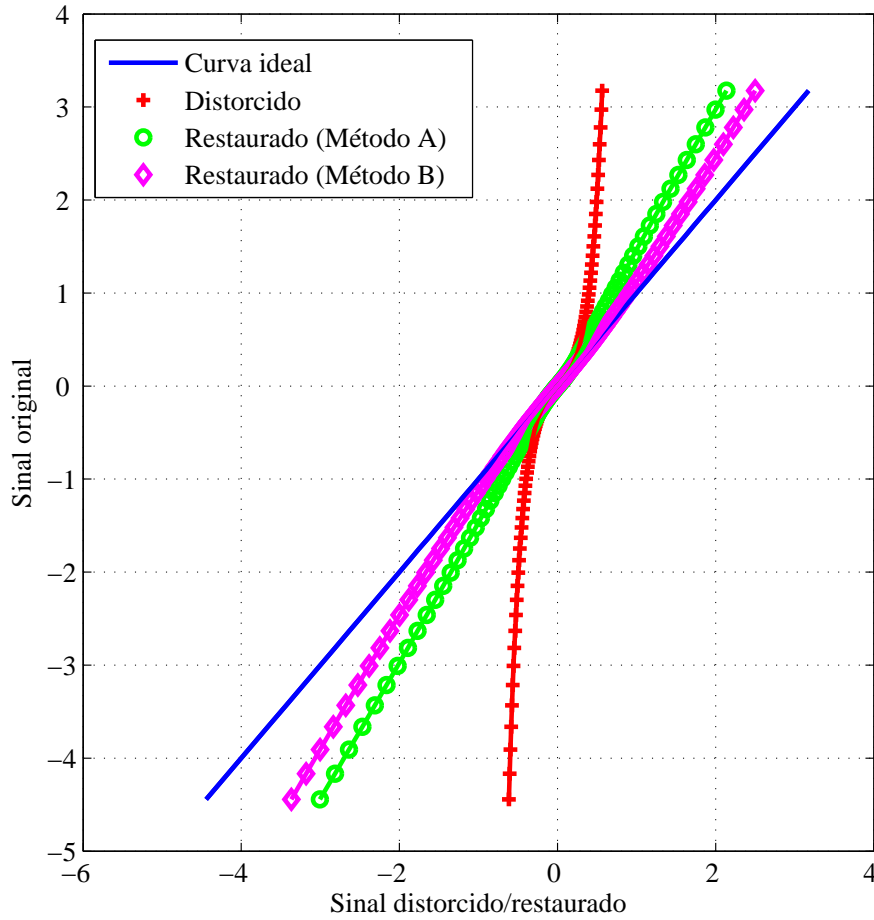


Figura 6.2: Sinal original contra sinais restaurados obtidos com métodos A e B, comparados com curva ideal (identidade) e curva. Ambos os métodos reduzem sensivelmente a não-linearidade do sinal, mas deixam distorções significativas em altas amplitudes.

Como, de acordo com o modelo AR, $e(n) = x(n) - \sum_{i=1}^P a_i x(n-i)$, a distribuição de $\mathbf{x} = [x(P+1)x(P+2) \dots x(N)]^T$ condicionada a $\mathbf{x}_0 = [x(1)x(2) \dots x(P)]^T$ é dada por:

$$p(\mathbf{x}|\mathbf{x}_0, \mathbf{a}, \sigma_e^2) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}} \right)^{N-P} \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{n=P}^N \left(x(n) - \sum_{i=1}^P a_i x(n-i) \right)^2 \right\}. \quad (6.19)$$

Por fim, levando em conta que $x(n)$ é obtido pela Equação (6.6) e, usando a fórmula de transformação monótona de variáveis aleatórias, a verossimilhança de $\mathbf{y} = [y(P+1)y(P+2) \dots y(N)]^T$ condicionada a $\mathbf{y}_0 = [y(1)y(2) \dots y(P)]^T$ é dada

por:

$$p(\mathbf{y}|\mathbf{y}_0, \mathbf{a}, \mathbf{m}, \sigma_e^2) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}} \right)^{N-P} \prod_{n=P+1}^N \left| 1 + \sum_{j=1}^M (2j+1)m_j y^{2j}(n) \right| \quad (6.20)$$

$$\times \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{n=P+1}^N \left(x(n) - \sum_{i=1}^P a_i x(n-i) \right) \right\}.$$

Lembrando que $g(\cdot)$ é sempre crescente ($|g'(\cdot)| = g'(\cdot)$) e usando $\mathbf{x} = \mathbf{y} + \mathbf{Y}\mathbf{m}$, a expressão acima pode ser reescrita como:

$$p(\mathbf{y}|\mathbf{y}_0, \mathbf{a}, \mathbf{m}, \sigma_e^2) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}} \right)^{N-P} \prod_{n=P+1}^N \{1 + \mathbf{h}_n \mathbf{m}\} \quad (6.21)$$

$$\times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} + \mathbf{Y}\mathbf{m})^T \mathbf{A}^T \mathbf{A} (\mathbf{y} + \mathbf{Y}\mathbf{m}) \right\}, \quad (6.22)$$

em que:

$$\mathbf{h}_n = \left[3y^2(n) \quad 5y^4(n) \quad \dots \quad (2M+1)y^{2M}(n) \right] \quad (6.23)$$

Para simplificar alguns cálculos subsequentes, definimos:

$$c(\mathbf{a}, \mathbf{m}) = (\mathbf{y} + \mathbf{Y}\mathbf{m})^T \mathbf{A}^T \mathbf{A} (\mathbf{y} + \mathbf{Y}\mathbf{m}) = (\mathbf{x} - \mathbf{X}\mathbf{a})^T (\mathbf{x} - \mathbf{X}\mathbf{a}), \quad (6.24)$$

em que os elementos de \mathbf{x} e \mathbf{X} podem ser obtidos através de $x(n) = g[y(n)]$.

6.4 Maximização da verossimilhança

Nesta seção apresentamos um procedimento para a maximização da verossimilhança definida pela Equação (6.20). A abordagem é inspirada em [30], no qual se resolve o problema de identificação cega de um sistema de Wiener (definido pela concatenação de um filtro linear com uma não-linearidade estática).

Começamos definindo o negativo do logaritmo da verossimilhança:

$$L(\mathbf{a}, \mathbf{m}, \sigma_e^2) = -\log [p(\mathbf{y}|\mathbf{a}, \mathbf{m}, \sigma_e^2)] = \quad (6.25)$$

$$\frac{(N-P)}{2} \log(2\pi) + \frac{(N-P)}{2} \log(\sigma_e^2) - \sum_{n=P+1}^N \log |1 + \mathbf{h}_n \mathbf{m}| + \frac{c(\mathbf{a}, \mathbf{m})}{2\sigma_e^2}$$

Derivando em relação a σ_e^2 e igualando a zero, obtemos:

$$\sigma_e^2 = \frac{c(\mathbf{a}, \mathbf{m})}{N - P}. \quad (6.26)$$

Substituindo σ_e^2 em (6.25), podemos mostrar que o resultado é proporcional à seguinte função-objetivo:

$$V(\mathbf{a}, \mathbf{m}) = \frac{(N - P)}{2} \log(c(\mathbf{a}, \mathbf{m})) - \sum_{n=P+1}^N \log(1 + \mathbf{h}_n \mathbf{m}). \quad (6.27)$$

6.4.1 Solução pelo método do Gradiente

Para minimização da função acima, usamos o algoritmo do gradiente. Agrupando os parâmetros em $\phi = \{\mathbf{a}, \mathbf{m}\}$, o gradiente de $V(\mathbf{a}, \mathbf{m})$ é dado por:

$$\frac{\partial V(\phi)}{\partial \phi} = \begin{bmatrix} \frac{\partial V(\phi)}{\partial \mathbf{a}} \\ \frac{\partial V(\phi)}{\partial \mathbf{m}} \end{bmatrix} \quad (6.28)$$

O gradiente em relação a \mathbf{a} é dado por:

$$\frac{\partial V(\phi)}{\partial \mathbf{a}} = \frac{(N - P)}{2c(\mathbf{a}, \mathbf{m})} \frac{\partial c(\mathbf{a}, \mathbf{m})}{\partial \mathbf{a}}, \quad (6.29)$$

em que, olhando para a Equação (6.24),

$$\frac{\partial c(\mathbf{a}, \mathbf{m})}{\partial \mathbf{a}} = 2\mathbf{X}^T \mathbf{X} \mathbf{a} - 2\mathbf{X}^T \mathbf{x}. \quad (6.30)$$

O gradiente em relação a \mathbf{m} é dado por:

$$\frac{\partial V(\phi)}{\partial \mathbf{m}} = \frac{(N - P)}{2c(\mathbf{a}, \mathbf{m})} \frac{\partial c(\mathbf{a}, \mathbf{m})}{\partial \mathbf{m}} - \sum_{n=P+1}^N \frac{\mathbf{h}_n^T}{1 + \mathbf{h}_n \mathbf{m}}, \quad (6.31)$$

em que, olhando para a Equação (6.24),

$$\frac{\partial c(\mathbf{a}, \mathbf{m})}{\partial \mathbf{m}} = 2\mathbf{Y}^T \mathbf{A}^T \mathbf{A} \mathbf{Y} \mathbf{m} + 2\mathbf{Y}^T \mathbf{A}^T \mathbf{A} \mathbf{y}. \quad (6.32)$$

Com base no gradiente calculado acima e usando como ponto inicial o resultado do procedimento da Seção 6.2, aplica-se o método do gradiente para obtenção

do mínimo global da função-objetivo da Equação (6.27). Testando o método em sinais reais, observamos que este algoritmo tem convergência mais lenta do que o método de Gauss-Newton (próxima seção), porém com complexidade computacional por iteração menor. Outro problema é que o método exige o ajuste de parâmetro μ , o que pode levar o algoritmo a divergir. Diante disso, decidimos não implementar esse algoritmo nos demais tópicos desta tese.

6.4.2 Solução pelo método de Gauss-Newton

O método de Gauss-Newton é obtido através de uma simplificação do método de Newton para o caso em que a função-objetivo pode ser escrita como uma soma de quadrados. Se

$$F : \mathbb{R}^D \rightarrow \mathbb{R}, F(\theta) = \sum_{n=1}^N r_n^2(\theta), \quad (6.33)$$

partindo de uma estimativa inicial θ_0 , o método de Gauss-Newton atualiza os parâmetros através de:

$$\theta^{(k+1)} = \theta^{(k)} + \Delta^k, \quad (6.34)$$

em que $\Delta^{(k)}$ é dado por:

$$\Delta^{(k)} = - (\mathbf{J}(\theta^{(k)})^T \mathbf{J}(\theta^{(k)}))^{-1} \mathbf{J}(\theta^{(k)})^T \mathbf{r}(\theta^{(k)}), \quad (6.35)$$

cada elemento do Jacobiano é definido por:

$$\mathbf{J}(\theta)_{ij} = \frac{\partial r_i(\theta)}{\partial \theta_j}, i = \{1, 2, \dots, N\}, j = \{1, 2, \dots, D\}, \quad (6.36)$$

e $\mathbf{r}(\theta)$ é o vetor formado pelos n componentes de $r_n(\theta)$, $n = \{1, 2, \dots, N\}$.

Manipulando a função-objetivo, podemos mostrar que ela é equivalente a:

$$V(\mathbf{m}, \mathbf{a}) = \sum_{n=P+1}^N r_n^2, \quad (6.37)$$

$$r_n = g(\mathbf{m})e(n, \mathbf{a}, \mathbf{m}) \quad (6.38)$$

$$g(\mathbf{m}) = \exp \left\{ -\frac{1}{N-P} \sum_{n=P+1}^N \log(1 + h_n \mathbf{m}) \right\} \quad (6.39)$$

$$e(n, \mathbf{a}, \mathbf{m}) = g[y(n)] - \sum_{i=1}^P a_i g[y(n-i)]. \quad (6.40)$$

No Apêndice A.2, mostramos os detalhes do cálculo do Jacobiano para a função-objetivo em questão. Aplicando o algoritmo ao mesmo sinal artificial descrito na Seção 6.2, obtemos os gráficos das Figuras 6.3, 6.4, 6.5 e 6.6, nos quais podemos ver que o algoritmo converge para valores muito próximos dos corretos.

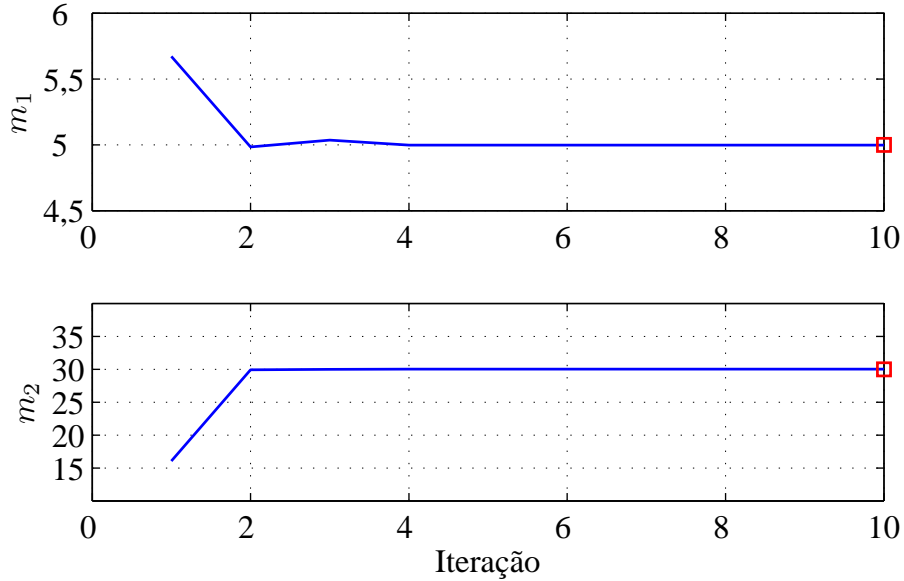


Figura 6.3: Evolução de coeficientes da expansão de Taylor pelo método de Gauss-Newton.

6.5 Solução bayesiana

A solução bayesiana permite a obtenção da distribuição a *posteriori* para os parâmetros \mathbf{a} , \mathbf{m} e σ_e^2 , que é dada por:

$$p(\mathbf{a}, \mathbf{m}, \sigma_e^2 | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{a}, \mathbf{m}, \sigma_e^2) p(\mathbf{a}, \sigma_e^2) p(\mathbf{m}). \quad (6.41)$$

A obtenção de estimativas dos parâmetros com base nessa distribuição é uma tarefa intratável analiticamente. Para este fim, adotamos a solução numérica baseada no amostrador de Gibbs, que consiste em obter amostras das distribuições condicionais de forma iterativa, de acordo com o seguinte procedimento:

1. Gerar valores iniciais $\mathbf{m}^{(0)}$, $\mathbf{a}^{(0)}$ e $\sigma_e^{2(0)}$.

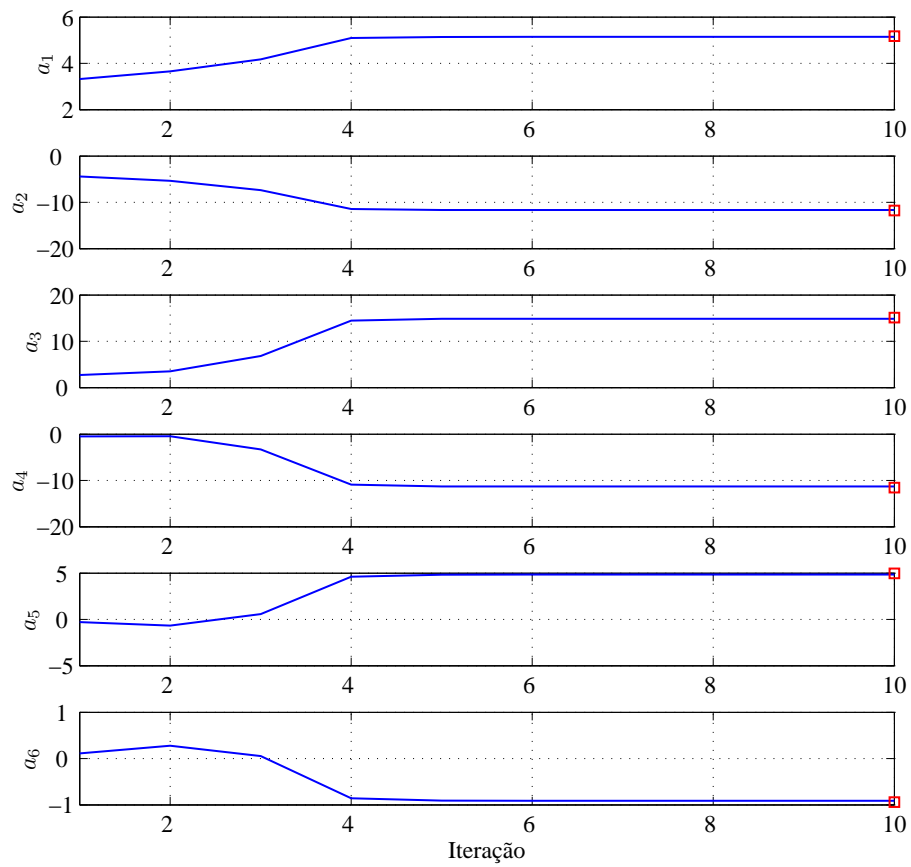


Figura 6.4: Evolução dos coeficientes do modelo AR pelo método de Gauss-Newton.

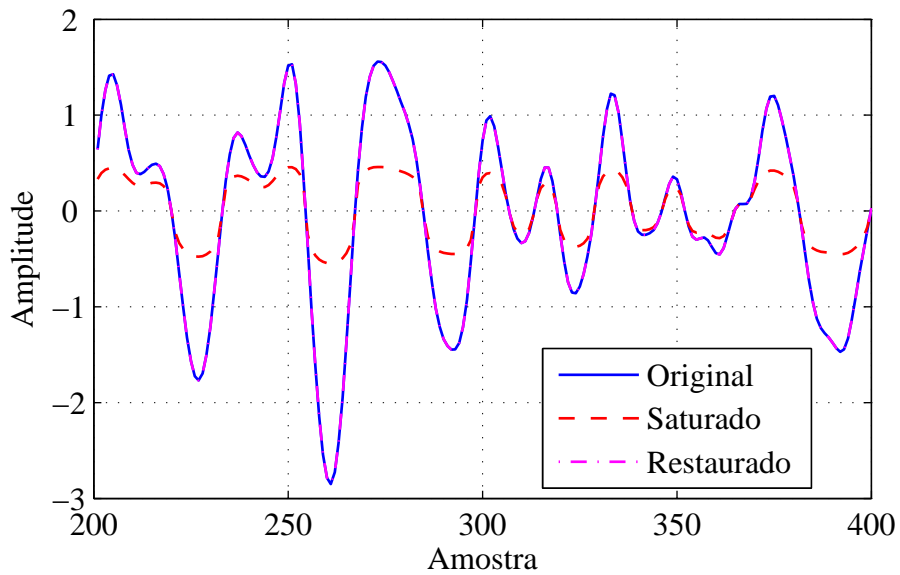


Figura 6.5: Comparação entre sinais original, saturado e restaurado pelo método de Gauss-Newton.

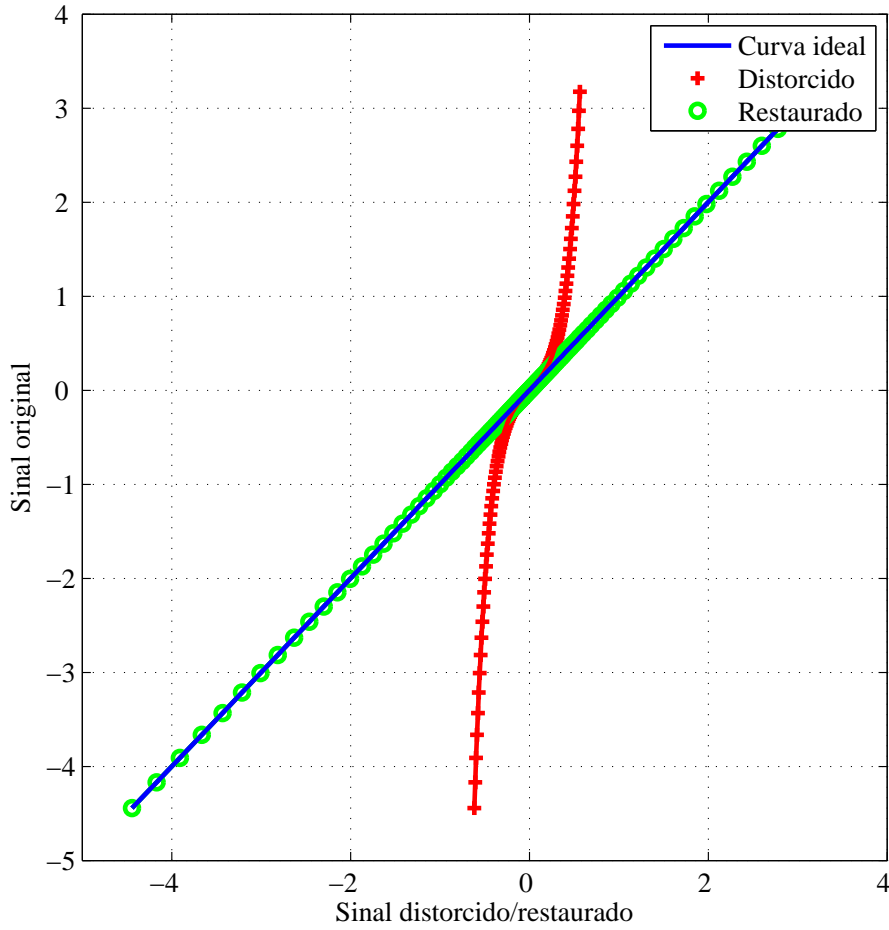


Figura 6.6: Sinal original contra sinal restaurado obtido com método de Gauss-Newton, comparados com curva ideal (identidade).

2. Para j de 1 até N_{it}

- (a) Amostrar $\mathbf{m}^{(j+1)}$ a partir da distribuição $p(\mathbf{m}|\mathbf{a}^{(j)}, \sigma_e^{2(j)}, \mathbf{y})$;
- (b) Amostrar $\mathbf{a}^{(j+1)}$ a partir da distribuição $p(\mathbf{a}|\mathbf{m}^{(j+1)}, \sigma_e^{2(j)}, \mathbf{y})$;
- (c) Amostrar $\sigma_e^{2(j+1)}$ a partir da distribuição $p(\sigma_e^2|\mathbf{m}^{(j+1)}, \mathbf{a}^{(j+1)}, \mathbf{y})$.

O objetivo é estimar os parâmetros da não-linearidade \mathbf{m} , tratando os demais como *nuisances*. Para a amostragem dos parâmetros \mathbf{a} e σ_e^2 o procedimento é idêntico ao que foi apresentado na Subseção 4.1.1, sendo que agora o vetor \mathbf{x} , necessário para a amostragem desses parâmetros é determinado a partir de \mathbf{y} através da função $g(\cdot)$.

Vamos então nos concentrar na distribuição *a posteriori* de \mathbf{m} , que é dada

por:

$$p(\mathbf{m}|\mathbf{y}, \mathbf{a}, \sigma_e^2) \propto \left(\frac{1}{\sqrt{2\pi}\sigma_e^2} \right)^{N-P} \left| \prod_{n=P+1}^N \{1 + \mathbf{h}_n \mathbf{m}\} \right| \quad (6.42)$$

$$\times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} + \mathbf{Y}\mathbf{m})^T \mathbf{A}^T \mathbf{A} (\mathbf{y} + \mathbf{Y}\mathbf{m}) \right\} p(\mathbf{m}), \quad (6.43)$$

em que se assume que os coeficientes \mathbf{m} são i.i.d. *a priori* e distribuídos de acordo com uma gaussiana de média 0 e variância σ_m^2 , ou seja:

$$p(\mathbf{m}) = N(\mathbf{m}|\mathbf{0}, \sigma_m^2 \mathbf{I}_M), \quad (6.44)$$

sendo \mathbf{I}_M a matriz identidade de M elementos.

A amostragem da distribuição *a posteriori* para \mathbf{m} não pode ser feita de forma simples. Optou-se por usar o algoritmo de Metropolis-Hastings, em que a distribuição proposta $q(\mathbf{m})$ é fixa e escolhida a partir de uma aproximação gaussiana de $p(\mathbf{m}|\mathbf{y}, \mathbf{a}, \sigma_e^2)$.

Definimos, inicialmente

$$P(\mathbf{m}) = \sum_{n=P+1}^N \log [1 + \mathbf{h}_n \mathbf{m}]. \quad (6.45)$$

Escolhemos a distribuição proposta $q(\mathbf{m})$ definida a partir da aproximação de segunda ordem de $P(\mathbf{m})$ em torno \mathbf{m}_0 :

$$P'(\mathbf{m}) = P(\mathbf{m}_0) + \nabla P(\mathbf{m}_0) (\mathbf{m} - \mathbf{m}_0)^T + (\mathbf{m} - \mathbf{m}_0)^T \mathbf{H}_P(\mathbf{m}_0) (\mathbf{m} - \mathbf{m}_0), \quad (6.46)$$

em que $\nabla P(\mathbf{m}_0)$ e $\mathbf{H}_P(\mathbf{m}_0)$ são respectivamente o gradiente e a hessiana de $P(\mathbf{m}_0)$, obtidos por:

$$\nabla P(\mathbf{m}_0) = \sum_{n=P+1}^N \frac{\mathbf{h}_n^T}{1 + \mathbf{h}_n \mathbf{m}_0} \quad (6.47)$$

e

$$H_{ij}(\mathbf{m}_0) = \frac{\partial^2 P}{\partial m_i \partial m_j}(\mathbf{m}_0) = - \sum_{n=P+1}^N \left(\frac{h_{m_i} h_{m_j}}{(1 + \mathbf{h}_n \mathbf{m}_0)^2} \right), \text{ para } i, j = \{1, 2, \dots, M\}. \quad (6.48)$$

Substituindo-se $P(\mathbf{m})$ por $P'(\mathbf{m})$ na Equação (6.42), obtemos a distribuição

proposta:

$$\begin{aligned}
q(\mathbf{m}) &\propto & (6.49) \\
&\exp \left\{ P(\mathbf{m}_0) + \nabla P(\mathbf{m}_0) (\mathbf{m} - \mathbf{m}_0)^T + (\mathbf{m} - \mathbf{m}_0)^T \mathbf{H}_P(\mathbf{m}_0) (\mathbf{m} - \mathbf{m}_0) \right\} \\
&\times \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} + \mathbf{Y}\mathbf{m})^T \mathbf{A}^T \mathbf{A} (\mathbf{y} + \mathbf{Y}\mathbf{m}) \right\} \exp \left\{ -\frac{1}{2\sigma_m^2} \mathbf{m}^T \mathbf{m} \right\},
\end{aligned}$$

que pode ser escrita na forma de uma gaussiana padrão:

$$q(\mathbf{m}) \propto \exp \left(-\frac{1}{2} (\mathbf{m} - \boldsymbol{\mu}_m)^T \mathbf{C}_m^{-1} (\mathbf{m} - \boldsymbol{\mu}_m) \right) \quad (6.50)$$

com

$$\boldsymbol{\mu}_m = - \left(\frac{1}{\sigma_e^2} \mathbf{Y}^T \mathbf{A}^T \mathbf{A} \mathbf{Y} + \frac{\mathbf{I}_M}{\sigma_m^2} - \mathbf{H}_P(\mathbf{m}_0) \right)^{-1} \left(\frac{1}{\sigma_e^2} \mathbf{y}^T \mathbf{A}^T \mathbf{A} \mathbf{Y} + \mathbf{H}_P(\mathbf{m}_0) \mathbf{m}_0 - \nabla P(\mathbf{m}_0) \right) \quad (6.51)$$

e

$$\mathbf{C}_m = \left(\frac{1}{\sigma_e^2} \mathbf{Y}^T \mathbf{A}^T \mathbf{A} \mathbf{Y} + \frac{\mathbf{I}_M}{\sigma_m^2} - \mathbf{H}_P(\mathbf{m}_0) \right)^{-1} \quad (6.52)$$

Amostragem de \mathbf{m} . No instante j , gera-se uma amostra \mathbf{m}^* a partir de $q(\mathbf{m})$. A amostra é aceita de acordo com probabilidade α dada por

$$\alpha(\mathbf{m}^{(j)}, m^*) = \min \left(1, \frac{\pi_j(\mathbf{m}^*) q(\mathbf{m}^{(j)})}{\pi_j(\mathbf{m}^{(j)}) q(\mathbf{m}^*)} \right), \quad (6.53)$$

em que

$$\pi_j(\mathbf{m}) = p(\mathbf{m} | \mathbf{y}, \mathbf{a}^{(j)}, \sigma_e^{2(j)}). \quad (6.54)$$

Para ilustrar o método proposto, geramos um sinal distorcido com os mesmos parâmetros usados na seção anterior. A Figura 6.7 mostra os sinais original, distorcido e restaurado pelo algoritmo proposto. A comparação entre a curva usada para gerar o sinal distorcido e a curva estimada é feita na Figura 6.8, na qual podemos ver que elas praticamente se sobrepõem. As primeiras iterações dos parâmetros geradas pelo amostrador de Gibbs são mostradas nas Figuras 6.9, 6.10 e 6.11, a partir das quais podemos concluir que o algoritmo converge após poucas iterações. Com base em 900 iterações após a convergência, construímos os histogramas das Figuras 6.12, 6.13 e 6.14.

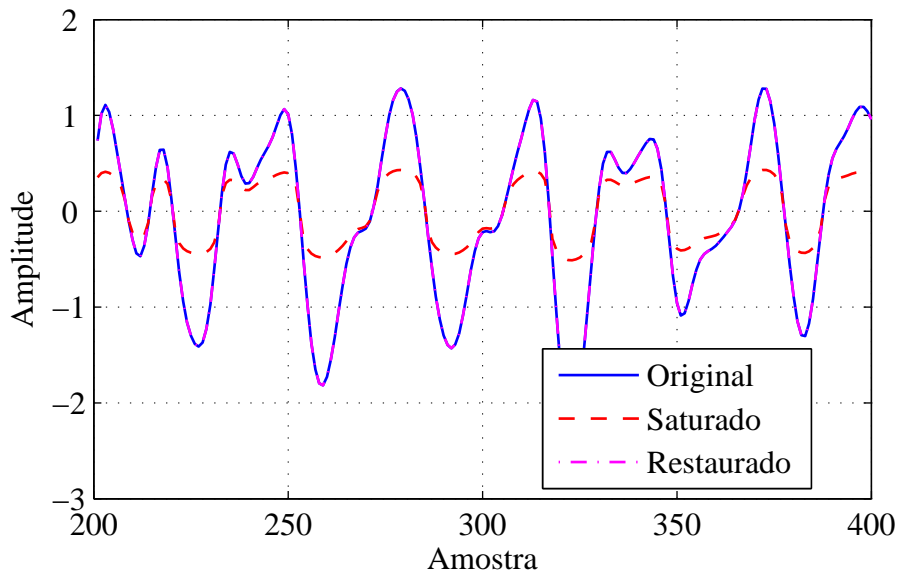


Figura 6.7: Comparação entre sinais original, saturado e restaurado por método bayesiano.

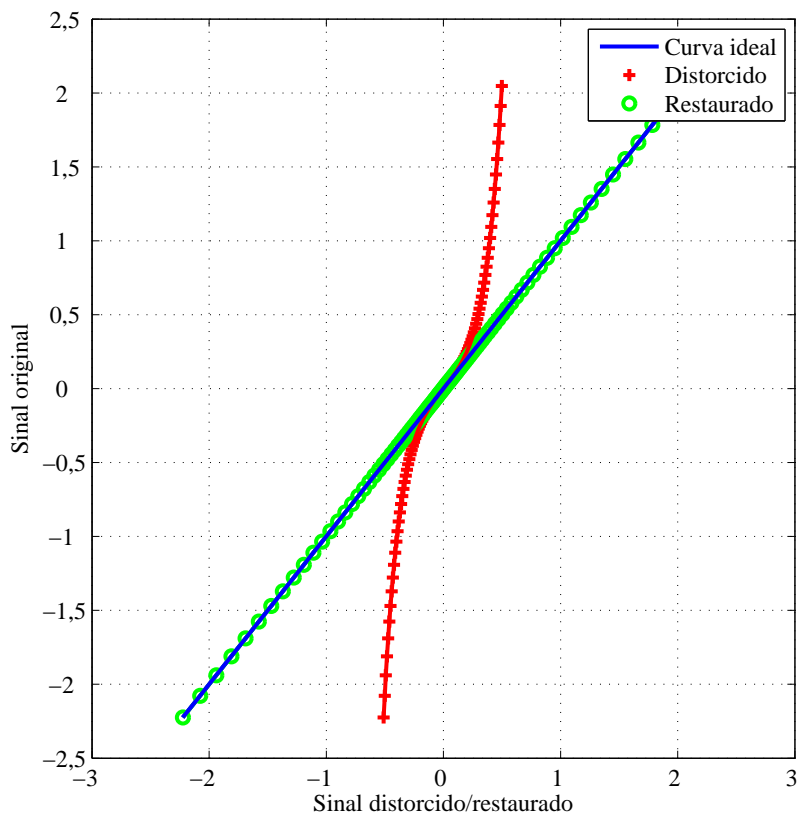


Figura 6.8: Sinal original contra sinal restaurado obtido com método de bayesiano, comparados com curva ideal (identidade).

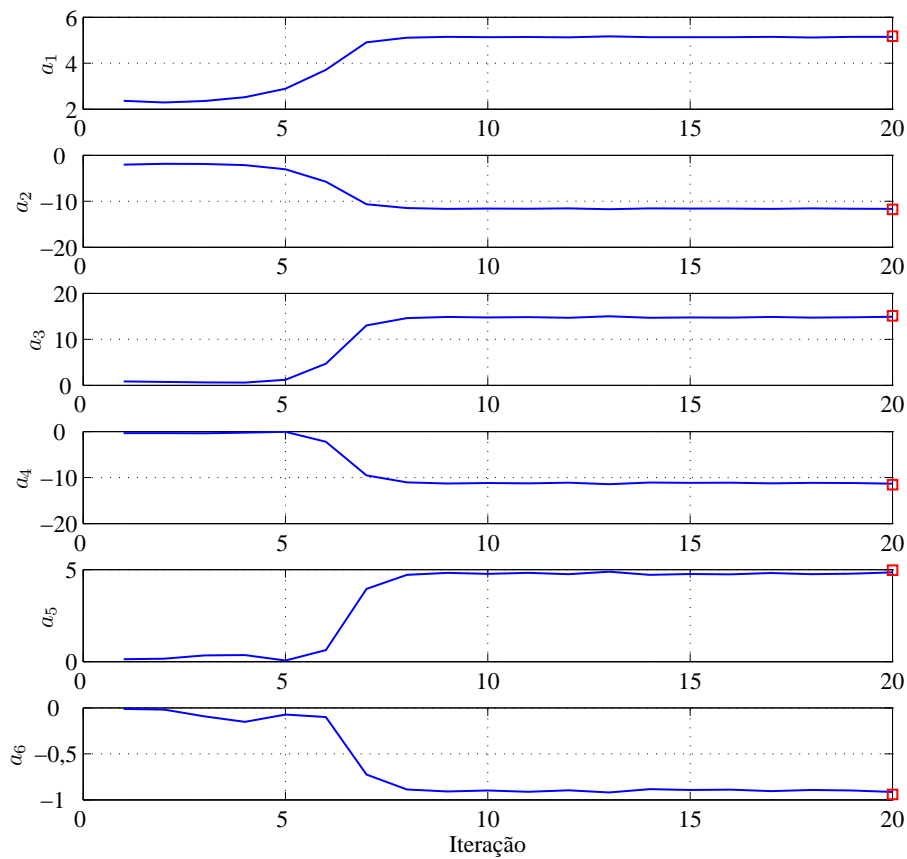


Figura 6.9: Evolução dos coeficientes do modelo AR através de amostrador de Gibbs.

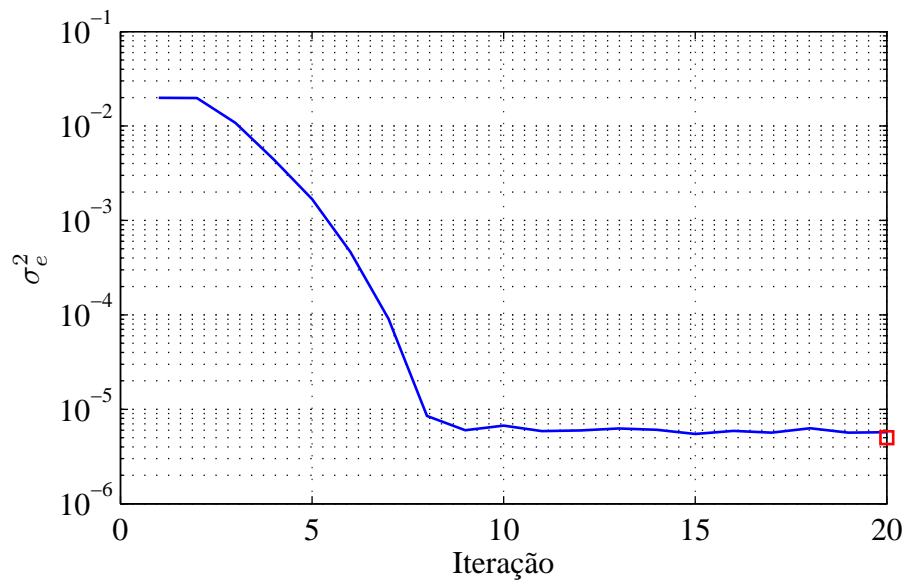


Figura 6.10: Evolução da variância de excitação do modelo AR através de amostrador de Gibbs.

Aplicando o algoritmo a sinais reais com diversos tipos de distorção artificial-

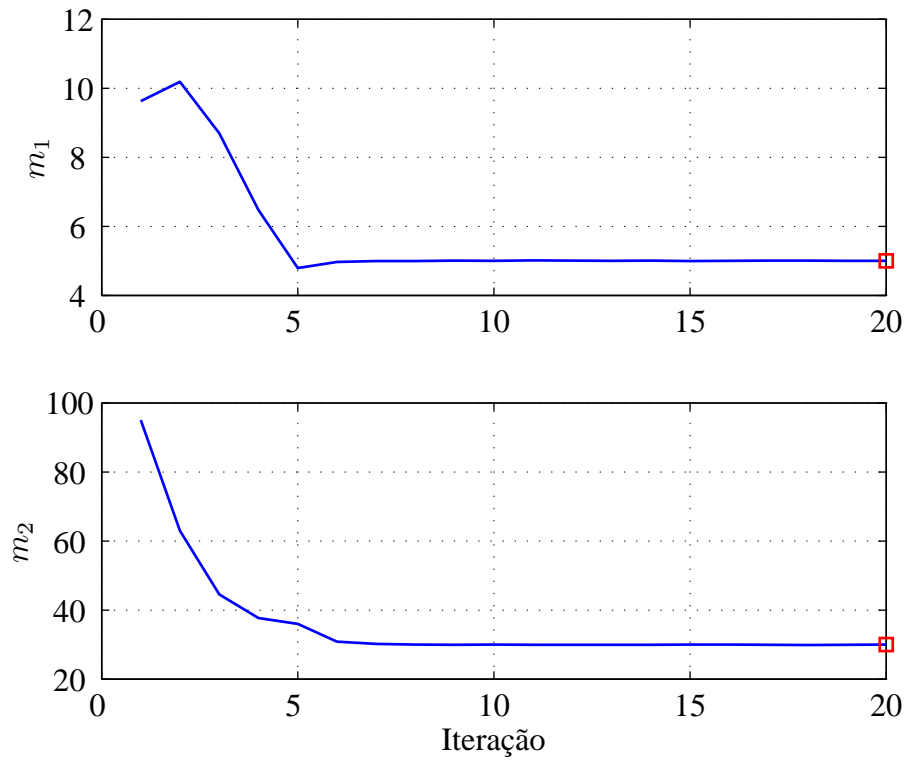


Figura 6.11: Evolução dos coeficientes da curva não linear através de amostrador de Gibbs.

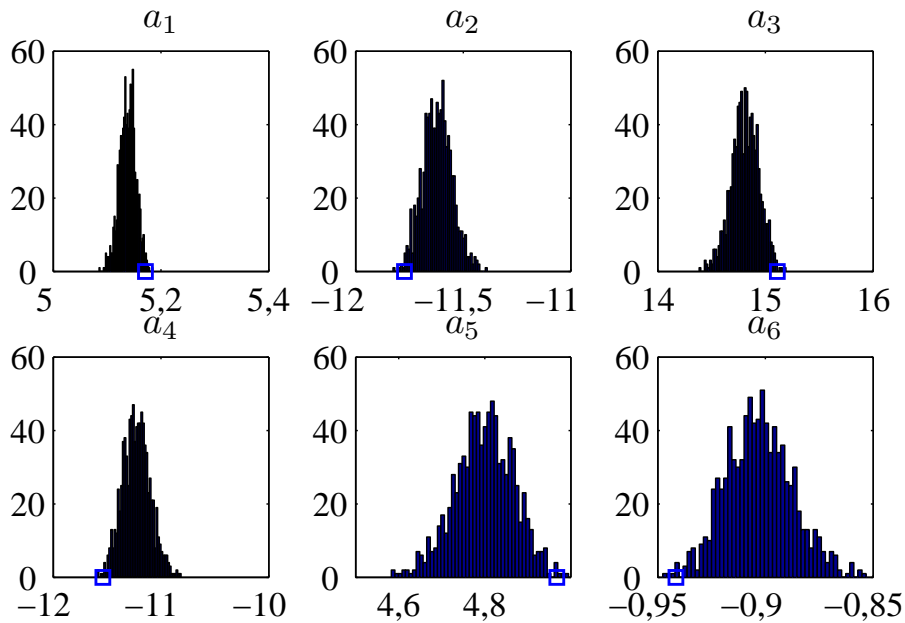


Figura 6.12: Histogramas de cada elemento dos coeficientes do modelo AR, pelo amostrador de Gibbs.

mente inseridas, observamos que o algoritmo converge para valores muito próximos dos corretos, e os sinais recuperados a partir das transformações inversas são pratica-

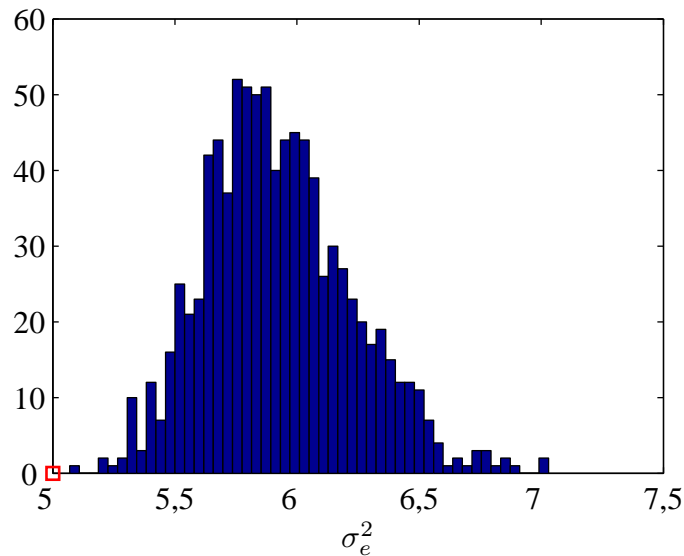


Figura 6.13: Histograma da excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs.

mente indistinguíveis dos originais. Entretanto, aplicando o algoritmo a um conjunto de sinais originalmente distorcidos (com características desconhecidas), os resultados não foram bons, e muitas vezes a aplicação da transformação estimada provocava aumento da distorção. Isso parece indicar que, ao menos para esse conjunto de sinais, a hipótese de distorção sem-memória não era válida.

6.6 Modelo linear por partes

Outra possibilidade de aproximação da distorção não-linear é através de uma curva linear por partes, formada por uma concatenação de retas de inclinações desconhecidas, como mostra a Figura 6.15. Naturalmente, quando maior o número de intervalos, mais acurada será a aproximação.

Assim como a modelagem por série de Taylor, trabalharemos com a função inversa $x(n) = g(y(n))$. Divide-se o intervalo $[0, \text{máx}(y_{\min}, y_{\max})]$ em $M+1$ partes de mesma largura Δy , conforme mostrado na Figura 6.15. Assim como anteriormente, assumimos que a curva é antissimétrica e possui inclinação unitária em torno da origem. Definimos a inclinação da reta do i -ésimo intervalo por $m_i = \text{tg}(\theta_i)$, $i \in \{0, 1, \dots, M\}$, com $m_0 = 1$. A um ponto genérico y^* pertencente ao i -ésimo intervalo,

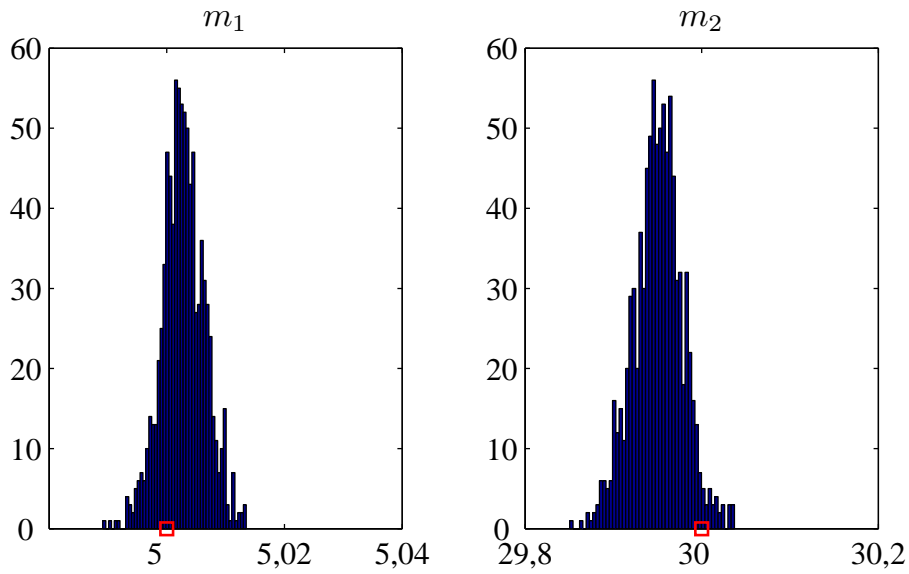


Figura 6.14: Histogramas dos coeficientes da curva não-linear, obtida pelo amostrador de Gibbs.

associa-se um ponto x^* , dado por:

$$x^* = g_i(y^*) = x_i + \text{sign}(y^*)(y^* - y_i)m_i, \quad (6.55)$$

em que $x_0 = 0$ e $\text{sign}(y^*)$ indica o sinal de y^* .

Como $x_i = \text{sign}(y^i)\Delta y \sum_{i=0}^{i-1} m_i$, segue que podemos escrever a Equação (6.55) na forma

$$x^* = u^* + \mathbf{p}(y^*)\mathbf{m}, \quad (6.56)$$

com $\mathbf{m} = [m_1 \ m_2 \ \dots \ m_M]^T$. Portanto, para o vetor \mathbf{x} completo:

$$\mathbf{x} = \mathbf{u} + \mathbf{P}\mathbf{m}, \quad (6.57)$$

em que tanto o vetor \mathbf{u} quanto a matriz \mathbf{P} são formados por elementos de \mathbf{y} e constantes conhecidas.

Com base nessa expressão, podemos calcular a verossimilhança:

$$\begin{aligned} p(\mathbf{y}|\mathbf{m}, \sigma_e^2, \mathbf{a}) &\propto \prod_{i=1}^M m_i^{N_i} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{u} + \mathbf{P}\mathbf{m})^T \mathbf{A}^T \mathbf{A} (\mathbf{u} + \mathbf{P}\mathbf{m}) \right\} \\ &= \exp \left\{ \sum_{i=1}^M N_i \ln m_i - \frac{1}{2\sigma_e^2} (\mathbf{u} + \mathbf{P}\mathbf{m})^T \mathbf{A}^T \mathbf{A} (\mathbf{u} + \mathbf{P}\mathbf{m}) \right\} \end{aligned} \quad (6.58)$$

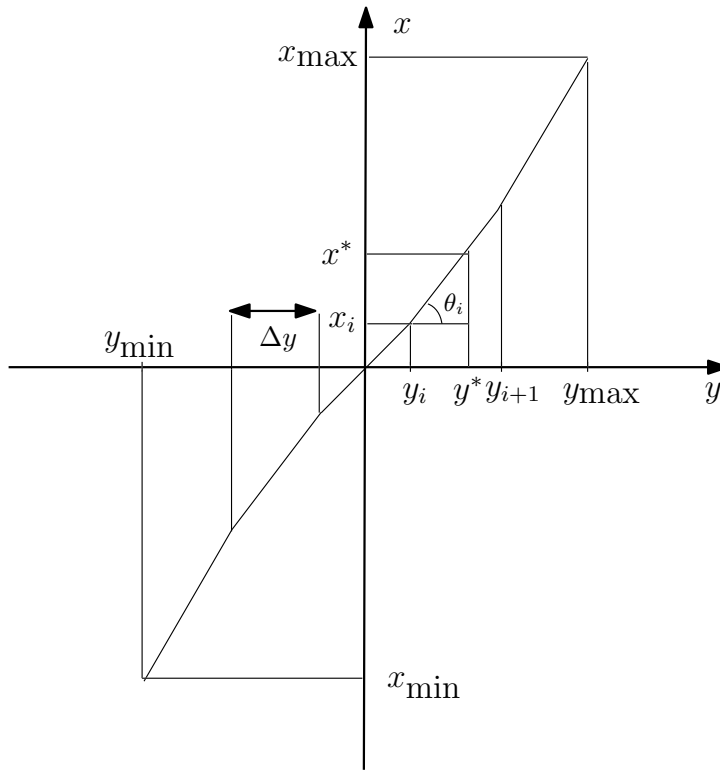


Figura 6.15: Ilustração do modelo linear por partes.

A distribuição condicional total de \mathbf{m} pode ser obtida pela regra de Bayes, resultando em:

$$p(\mathbf{m}|\mathbf{y}, \sigma_e^2, \mathbf{a}) \propto p(\mathbf{y}|\mathbf{m}, \sigma_e^2, \mathbf{a})p(\mathbf{m}), \quad (6.59)$$

Neste trabalho usamos uma priori vaga para \mathbf{m} idêntica à que usamos na seção anterior, mas futuramente pretendemos adotar prioris mais informativas, que incorporem por exemplo o fato de que devemos ter $m_i > m_{i+1}$ para que a curva seja do formato desejado. Para obtenção numérica da distribuição *a posteriori*, propomos um algoritmo muito similar ao da seção anterior, usando também uma aproximação gaussiana para a distribuição condicional total de \mathbf{m} .

Testamos o algoritmo num sinal artificialmente gerado com os parâmetros da Subseção 4.1.1 e distorcemos o sinal com uma curva não-linear gerada pelo modelo linear por partes com coeficientes $m_1 = 1$ e $m_2 = 2$. As Figuras de 6.16 a 6.21 mostram gráficos das amostras dos parâmetros e seus histogramas gerados pelo algoritmo proposto, a partir dos quais podemos verificar sua correção.

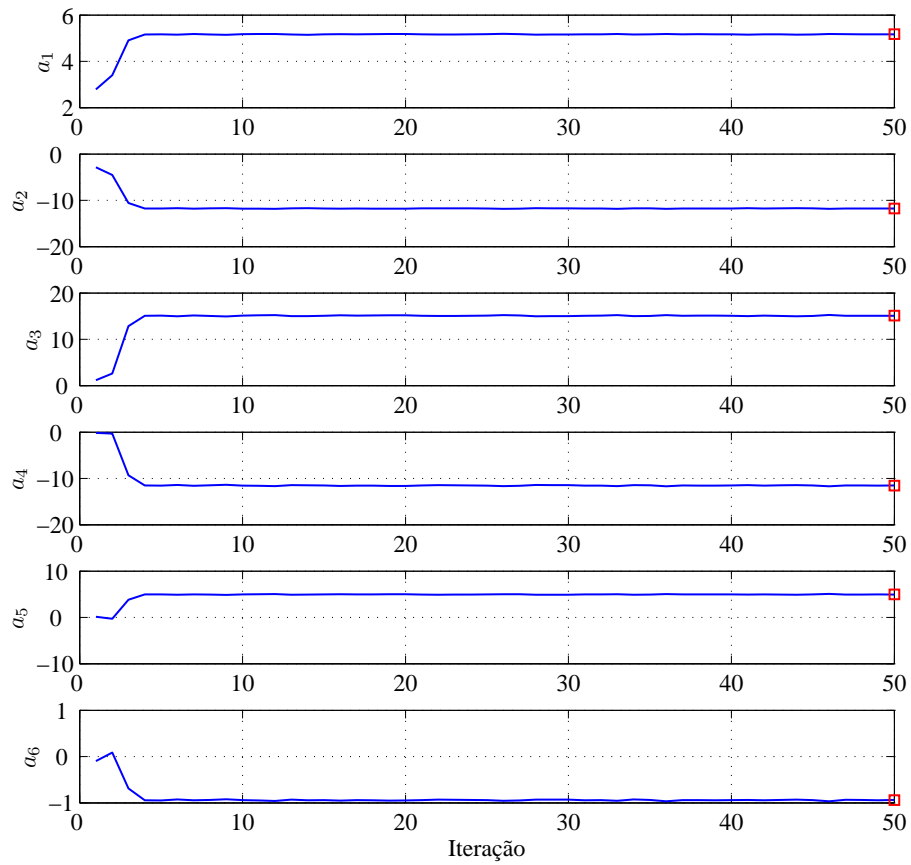


Figura 6.16: Evolução dos coeficientes do modelo AR através de amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.

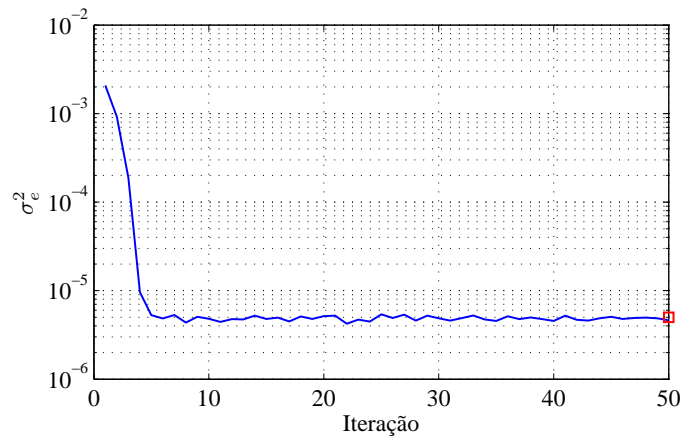


Figura 6.17: Evolução da variância de excitação do modelo AR através de amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.

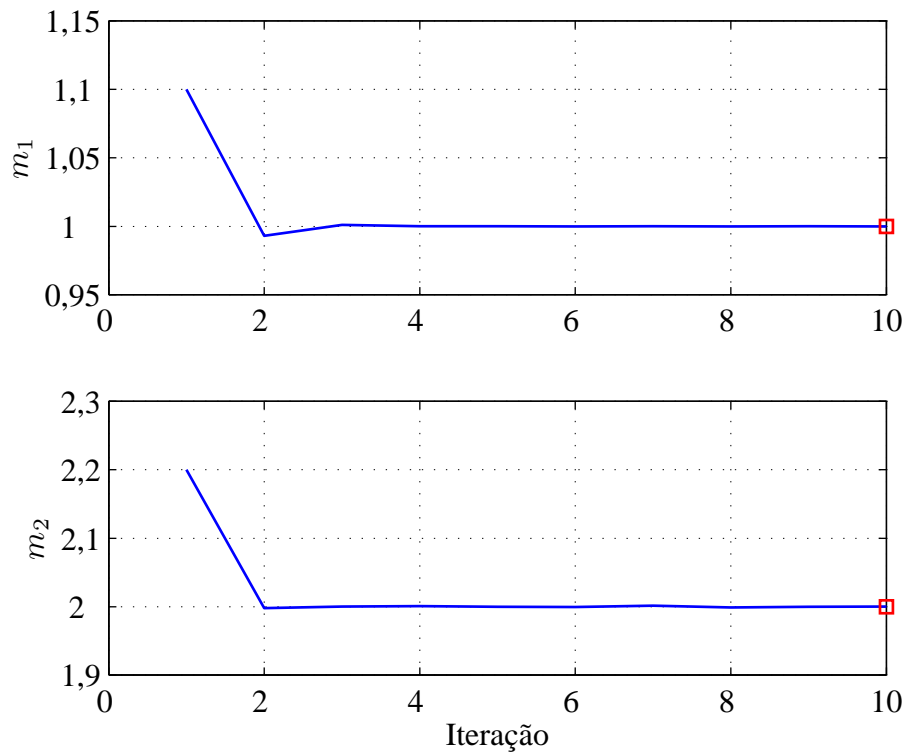


Figura 6.18: Evolução dos coeficientes das retas do modelo linear por partes, através de amostrador de Gibbs.

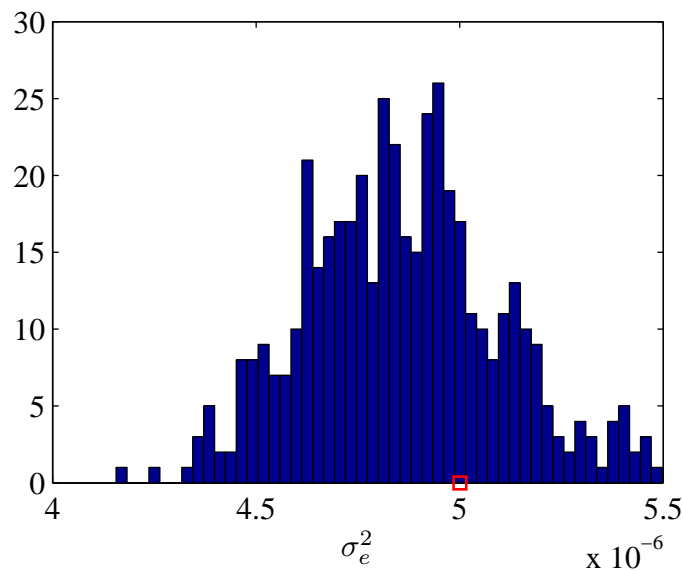


Figura 6.19: Histograma da variância de excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.

6.7 Modelagem de ruído aditivo

A Figura 6.22 mostra o modelo para distorções não-lineares sem memória em que o sinal degradado é observado na presença de ruído de medição, $v(n)$. A recuperação

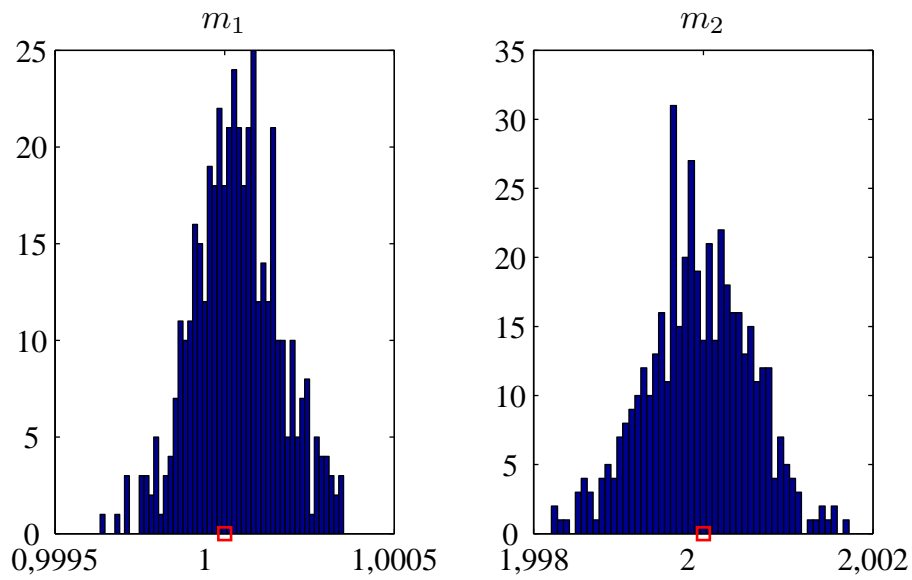


Figura 6.20: Histogramas dos coeficientes das retas do modelo linear por partes.

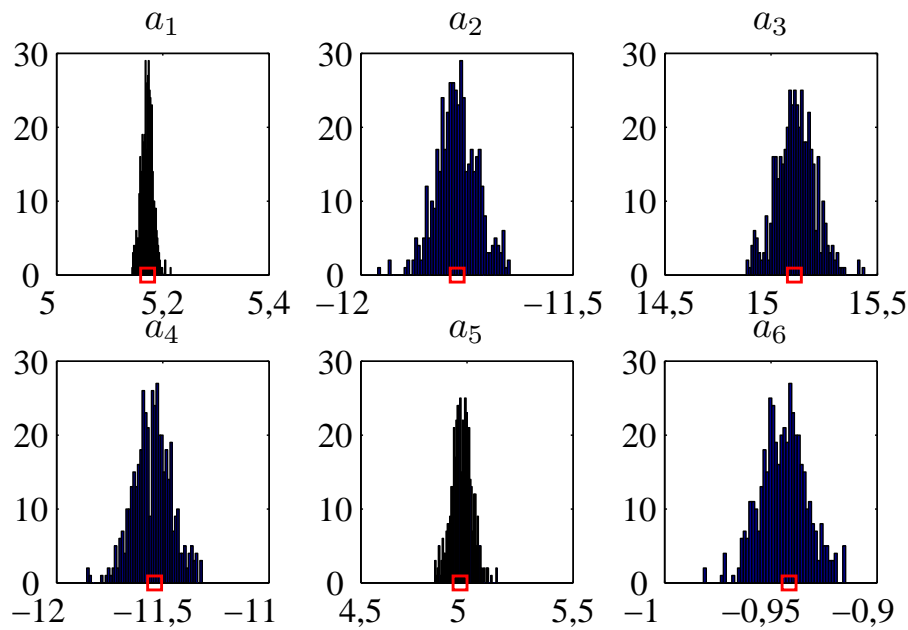


Figura 6.21: Histogramas de cada elemento dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs a partir de um sinal com distorção não-linear descrito por modelo linear por partes.

do sinal original $x(n)$ nesse caso é mais crítica pois a estimação dos parâmetros da não-linearidade não é mais suficiente para se obter $x(n)$. Nesta seção, descrevemos um algoritmo que permite estimar os parâmetros da distorção em conjunto com o sinal original.

Inicialmente pensamos que, para este problema, seria mais conveniente traba-

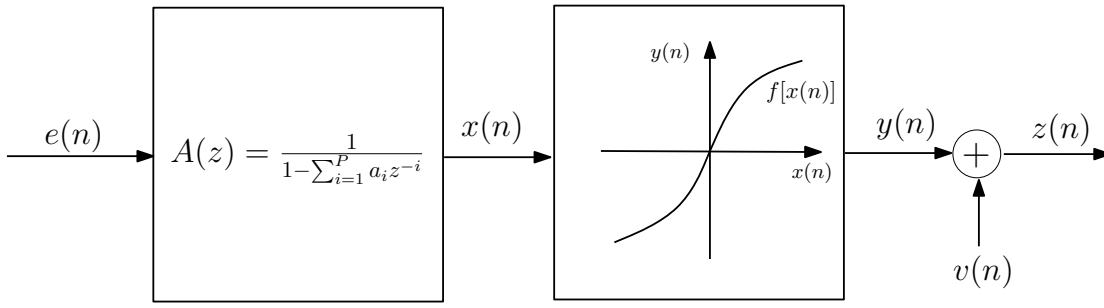


Figura 6.22: Modelo de distorção não-linear sem memória com observação ruidosa.

lhar com uma aproximação da função direta $f(x(n))$ em vez de sua inversa $g(y(n))$, já que agora $x(n)$ é uma das variáveis do algoritmo. Mas observamos que essa escolha iria requerer uma ordem maior para modelar as distorções com formato típicos. Decidimos então continuar trabalhando com a função inversa, embora isso torne não-gaussiana a distribuição condicional dos parâmetros da curva, e conseqüentemente mais difícil de amostrar, como veremos mais adiante.

Sendo assim, definimos $f(x(n)) = f(x(n), \mathbf{m})$ através de sua inversa $g(y(n)) = g(y(n), \mathbf{m}) = y(n) + \sum_{i=1}^M m_i y(n)^{2i+1}$. Como veremos, não precisaremos determinar $f(x(n), \mathbf{m})$ de forma fechada, desde que possamos calcular suas derivadas indiretamente através de $g(y(n), \mathbf{m})$.

O sinal degradado é observado na presença de ruído, ou seja:

$$z(n) = y(n) + v(n) = f(x(n), \mathbf{m}) + v(n). \quad (6.60)$$

O objetivo é estimar o sinal $x(n)$ com base em $z(n)$ tratando todas as demais variáveis como *nuisances*. Começamos calculando a verossimilhança do sinal observado condicionado ao sinal $x(n)$ e aos parâmetros \mathbf{a} , σ_e^2 e \mathbf{m} . Pela Equação (6.60), vemos que a distribuição de $z(n)$ é igual à distribuição de $v(n)$ com média deslocada de $f(x(n), \mathbf{m})$, isto é,

$$p(\mathbf{z}|\mathbf{x}, \mathbf{a}, \sigma_e^2, \mathbf{m}) = p(\mathbf{z}|\mathbf{x}, \mathbf{m}) = p_v(\mathbf{z} - f(\mathbf{x}, \mathbf{m})), \quad (6.61)$$

em que, por um abuso de notação $f(\mathbf{x}, \mathbf{m})$, representa um vetor formado por elementos de $f(x(n), \mathbf{m})$.

A distribuição *a posteriori* conjunta será portanto:

$$p(\mathbf{x}, \mathbf{a}, \sigma_e^2, \mathbf{m} | \mathbf{z}) \propto p_v(\mathbf{z} - f(\mathbf{x}, \mathbf{m}))p(\mathbf{x} | \mathbf{a}, \sigma_e^2)p(\mathbf{a}, \sigma_e^2)p(\mathbf{m}). \quad (6.62)$$

Novamente adotamos o amostrador de Gibbs como estrutura geral para obtenção de amostras de cada variável, recorrendo ao algoritmo de Metropolis-Hastings com distribuição proposta gerada por aproximação gaussiana quando a distribuição não puder ser amostrada de forma simples. Os parâmetros \mathbf{a} e σ_e^2 são amostrados de forma idêntica ao que foi feito na Subseção 4.1.1. Vamos analisar a amostragem dos demais parâmetros.

Amostragem de \mathbf{m} : Assumindo para este parâmetro uma distribuição *a priori* gaussiana de média zero e matriz de covariância $\mathbf{C}_m = \sigma_M^2 \mathbf{I}_M$, sua distribuição condicional total é dada por:

$$\begin{aligned} p(\mathbf{m} | \mathbf{x}, \mathbf{a}, \sigma_e^2, \mathbf{m}) &\propto p_v(\mathbf{z} - f(\mathbf{x}, \mathbf{m}))p(\mathbf{m}) \\ &\propto \exp \{L(\mathbf{m})\}, \end{aligned} \quad (6.63)$$

em que:

$$L(\mathbf{m}) = -\frac{1}{2\sigma_v^2} \sum_{n=1}^N [z(n) - f(x(n), \mathbf{m})]^2 - \frac{1}{2\sigma_M^2} \mathbf{m}^T \mathbf{m}. \quad (6.64)$$

Assim como antes, a estratégia de amostragem é aproximar essa distribuição por uma gaussiana através da aproximação quadrática de $L(\mathbf{m})$ em torno de um ponto \mathbf{m}_0 (idealmente próximo a seu mínimo), e então usar o resultado como distribuição proposta no contexto do algoritmo de Metropolis-Hastings.

Nesse estágio do algoritmo, $x(n)$ é conhecido exatamente, sendo portanto uma constante. Usamos esse fato para calcular as derivadas parciais de $f(x(n), \mathbf{m})$ em relação \mathbf{m}_i , $i = \{1, \dots, M\}$, necessárias para a aproximação de $L(\mathbf{m})$ por série de Taylor. Escrevendo o resultado da aproximação no formato de uma gaussiana padrão, obtemos a distribuição usada como proposta $q(\mathbf{m}) = \mathbf{N}(\mathbf{m} | \boldsymbol{\mu}_m, \mathbf{C}_m)$, cujos parâmetros podem ser obtidos pelo formulário abaixo:

$$\mathbf{C}_m = - \left(\mathbf{H}(\mathbf{m}_0) - \frac{1}{\sigma_M^2} \mathbf{I}_M \right)^{-1}; \quad (6.65)$$

$$\boldsymbol{\mu}_m = -\mathbf{C}_m(\mathbf{H}(\mathbf{m}_0)\mathbf{m}_0 - \mathbf{G}(\mathbf{m}_0)); \quad (6.66)$$

$$\mathbf{G}_i = \frac{1}{\sigma_v^2} \sum_{n=P+1}^N (z(n) - y(n)) \frac{\partial y(n)}{\partial m_i}; \quad (6.67)$$

$$\mathbf{H}_{ij} = \frac{1}{\sigma_v^2} \sum_{n=P+1}^N \left[(z(n) - y(n)) \frac{\partial^2 y(n)}{\partial m_i \partial m_j} - \frac{\partial y(n)}{\partial m_i} \frac{\partial y(n)}{\partial m_j} \right]; \quad (6.68)$$

$$\frac{\partial y(n)}{\partial m_i} = -\frac{y(n)^{2i+1}}{g'(y(n))}; \quad (6.69)$$

$$\frac{\partial^2 y(n)}{\partial m_i \partial m_j} = -(2i+1)y(n)^{2i} \frac{\partial y(n)}{\partial m_i} + (2j+1)y(n)^{2j} \frac{\partial y(n)}{\partial m_j} + \frac{\frac{\partial y(n)}{\partial m_i} \frac{\partial y(n)}{\partial m_j}}{g'(y(n))}. \quad (6.70)$$

Na Equação (6.67), G_i é o i -ésimo elemento do vetor gradiente \mathbf{G} .

Amostragem de \mathbf{x} : Olhando para a Equação (6.62), reconhecemos que a distribuição condicional total para \mathbf{x} é dada por:

$$p(\mathbf{x}|\mathbf{z}, \mathbf{a}, \sigma_e^2, \mathbf{m}) \propto p_v(\mathbf{z} - f(\mathbf{x}))p(\mathbf{x}|\mathbf{a}, \sigma_e^2). \quad (6.71)$$

Devido ao fato de a não-linearidade ser aplicada a \mathbf{x} , sua distribuição condicional total não tem um formato bem conhecido, tornando a amostragem dessa variável mais complicada. Como o número de parâmetros (no caso, a variável \mathbf{x}) é igual ao número de dados, não há garantia de que uma aproximação quadrática será acurada, exceto para níveis de ruído e não-linearidade moderados.

O algoritmo proposto para amostragem de \mathbf{x} se baseia no cálculo de uma aproximação de segunda ordem para $v^2(n) = [z(n) - f(x(n))]^2$, na escolha da distribuição resultante (que será gaussiana) como proposta e no uso do algoritmo de Metropolis-Hastings poderia ser usado para decidir sobre a aceitação ou rejeição de novas amostras.

O máximo da distribuição de $x(n)$ ocorre quando $z(n) - f(x(n)) = 0$, ou seja, quando $x(n) = f^{-1}(z(n))$. A aproximação de $v^2(n)$ em torno desse ponto resulta numa distribuição gaussiana que aproxima $p_v(\mathbf{z} - f(\mathbf{x}))$, a qual possui média $\boldsymbol{\mu} = f^{-1}(\mathbf{z}) = g(\mathbf{z})$ e matriz de covariância $\mathbf{C} = \sigma_v^2 g'(\mathbf{z})^2 \mathbf{I}_N$. Podemos então substituir essa aproximação na Equação (6.71) para gerar a distribuição proposta.

Testando o algoritmo, confirmamos que a aproximação é boa para ruído e distorção moderados, mas ruim nos demais casos, com um alto percentual de amos-

tras rejeitadas. Em geral, existe um compromisso entre a quantidade de ruído e o nível de distorção para os quais o método funciona adequadamente.

As Figuras de 6.23 a 6.30 mostram gráficos de simulações obtidos com a implementação do algoritmo para um sinal distorcido com parâmetros $m_1 = 5$ e $m_2 = 30$ e com ruído aditivo de 30 dB SNR.

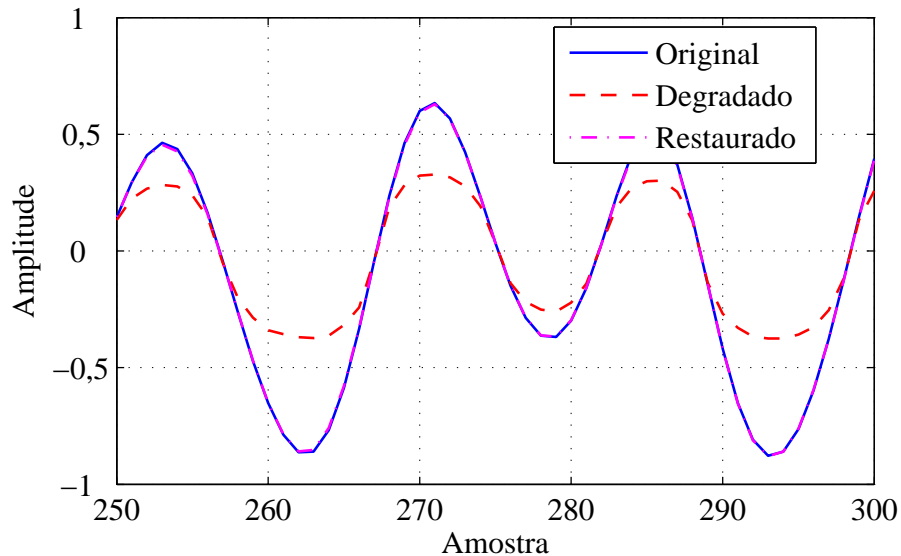


Figura 6.23: Comparação entre sinais original, saturado com ruído e restaurado por método bayesiano.

6.8 Conclusões

Neste capítulo apresentamos diversos algoritmos para lidar com o problema de restauração de sinais corrompidos com saturação gradual. Modelando a curva não-linear por sua expansão em série de Taylor, elaboramos diversos algoritmos que permitem estimar os coeficientes dessa série, e assim recuperar o sinal original. Dois métodos simples e rápidos permitem obter uma estimativa inicial de boa qualidade para esses coeficientes, as quais podem ser usadas para os métodos iterativos mais sofisticados. Além disso, propusemos um algoritmo que realiza a maximização da verossimilhança através do método de Gauss-Newton e outro que, no contexto bayesiano, obtém amostras da distribuição *a posteriori* das variáveis desconhecidas.

Para solução do problema na presença de ruído aditivo, apresentamos uma variação do algoritmo bayesiano que se mostrou adequada para níveis de distorção

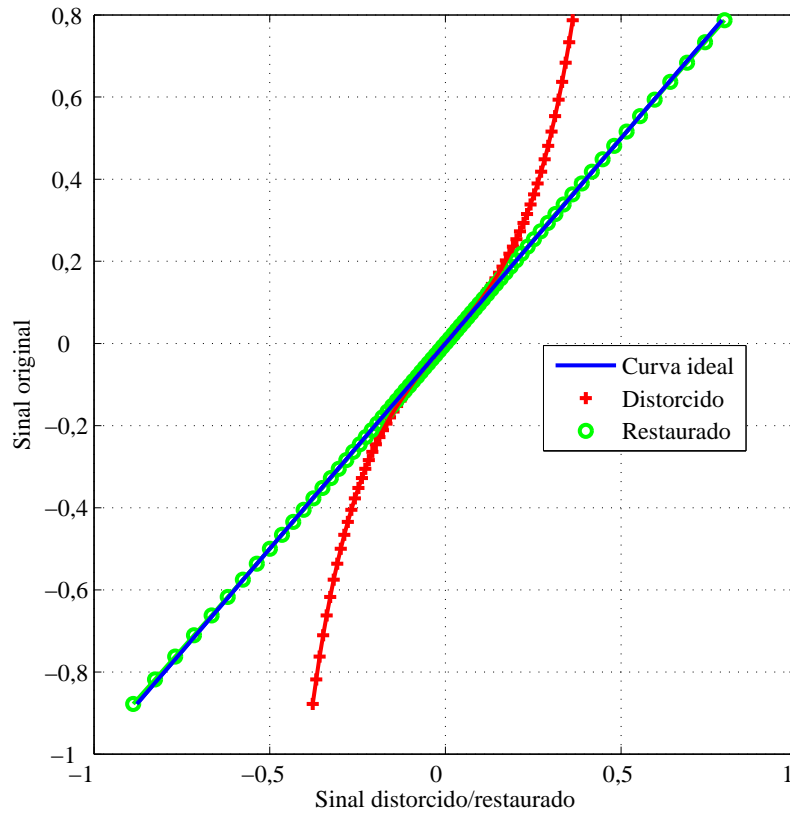


Figura 6.24: Sinal original contra sinal restaurado obtido com método Bayesiano, a partir de sinal distorcido com ruído, comparado com curva ideal (identidade).

ou ruído moderado. Também introduzimos um modelo linear por partes para aproximação genérica de curvas não-lineares invertíveis sem-memória, conduzindo-nos a um algoritmo bayesiano que apresentou resultados promissores.

Exemplos da aplicação dos algoritmos deste capítulo a sinais reais são apresentados no Apêndice B.

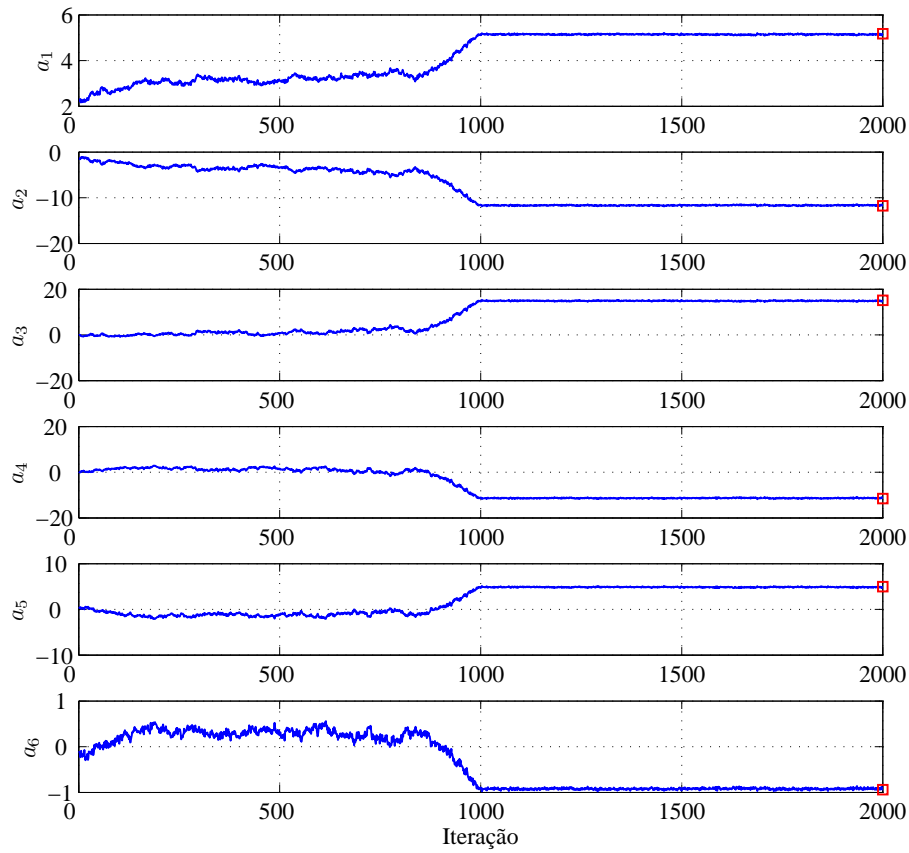


Figura 6.25: Evolução dos coeficientes do modelo AR através de amostrador de Gibbs, obtidos a partir de sinal saturado com ruído.

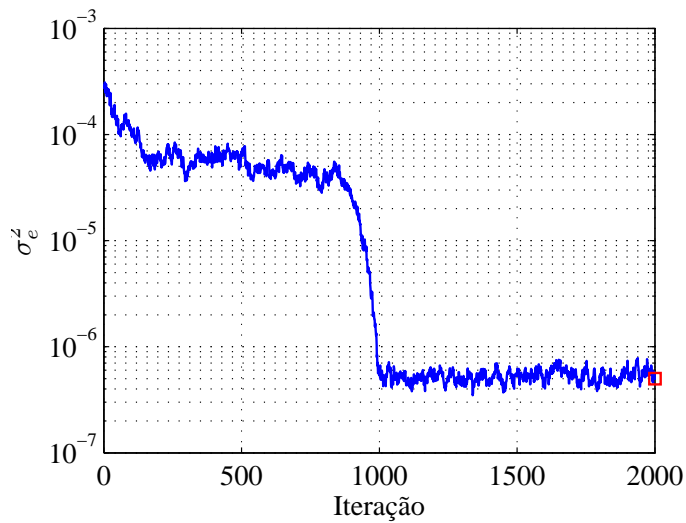


Figura 6.26: Evolução da variância de excitação do modelo AR através de amostrador de Gibbs, obtida a partir de sinal saturado com ruído.

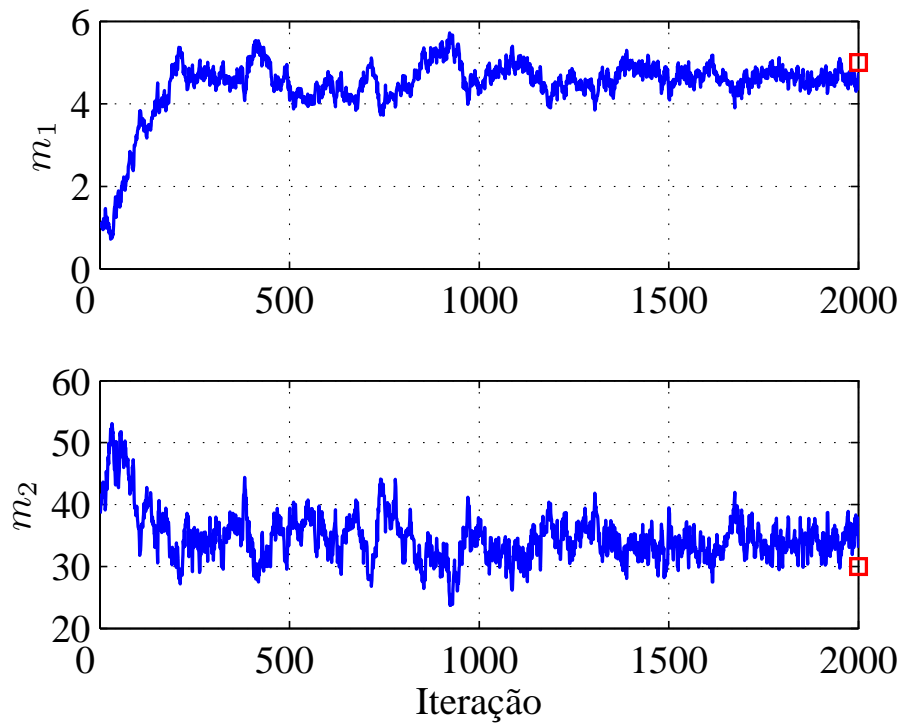


Figura 6.27: Evolução dos coeficientes da curva não linear através de amostrador de Gibbs, obtidos a partir de sinal saturado com ruído.

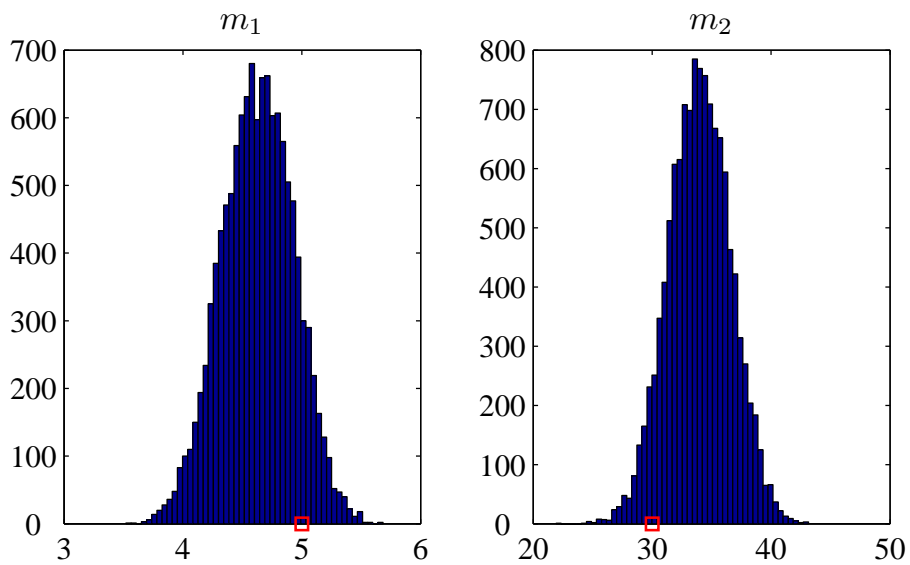


Figura 6.28: Histogramas dos coeficientes da curva não-linear obtida pelo amostrador de Gibbs, a partir de sinal saturado com ruído.

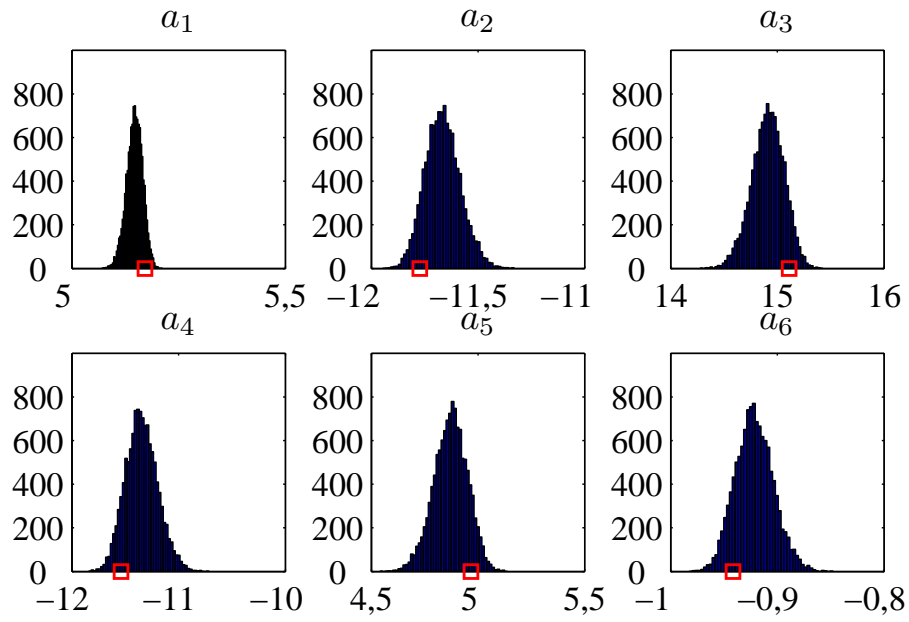


Figura 6.29: Histogramas de cada elemento dos coeficientes do modelo AR obtidos pelo amostrador de Gibbs, a partir de um sinal saturado com ruído.

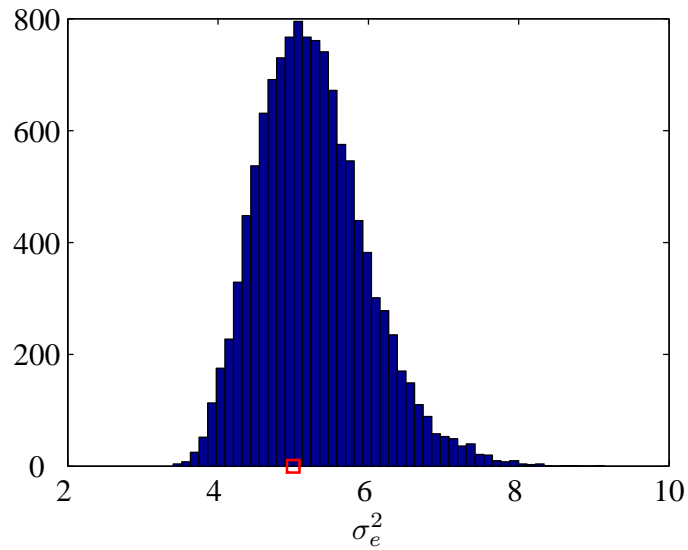


Figura 6.30: Histograma da excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs, a partir de um sinal saturado com ruído.

Capítulo 7

Restauração de Sinais

Corrompidos com Distorção

Não-linear com Memória

No capítulo anterior apresentamos algoritmos para lidar com o problema de restauração de sinais com saturação, em que se assumiu um modelo não-linear sem memória para a distorção. Em muitos casos, contudo, a presença concomitante de outras distorções introduz memória no sistema. É o que acontece, por exemplo, nos sistemas que usam circuitos de equalização nos equipamentos de reprodução.

Em [22], sugere-se que um filtro linear em cascata com a curva não-linear sem memória, como mostrado na Figura 7.1, seria geral o suficiente para a modelagem realista de sinais de áudio com distorção não-linear com memória. Em [34], reportam-se testes em que essa estrutura se mostrou adequada para sinais de áudio com distorções causadas por amplificadores e gravadores magnéticos reais. Neste mesmo trabalho, os autores propõem um método para identificação da curva não-linear e do filtro linear. O método assume que a forma do espectro do sinal de áudio é constante ao longo do tempo, só variando a energia, e também assume conhecida a autocorrelação do sinal original. Reportam-se bons resultados com distorções artificialmente inseridas, mas as hipóteses do método nos fazem crer que esta solução está longe da ideal. Em particular, o método assume conhecimento sobre a densidade espectral de potência do sinal original, que na prática não seria facilmente medida.

Um método mais geral, aplicável em princípio a qualquer tipo de distorção não-linear com ou sem memória, proposto em [35] e mais detalhado em [36], baseia-se na modelagem do sinal degradado como a cascata de um modelo AR (associado ao sinal original) com um modelo NAR (*Non-Linear AR*, associado a um “canal” não-linear). Apesar de sua generalidade, esse método obteve sucesso ao ser aplicado a sinais reais, devido à dificuldade em se identificar os termos da série e também porque não modela adequadamente outros tipos de distorções usualmente presentes em conjunto com a não-linearidade.

Neste capítulo, vamos apresentar a formulação estatística do problema de restauração, considerando a estrutura da Figura 7.1. Propomos duas soluções, uma baseada na maximização da verossimilhança através do algoritmo de Gauss-Newton, e outra baseada na obtenção de amostras da distribuição *a posteriori* através de um algoritmo baseado em MCMC. Como ambas as soluções contêm elementos similares aos das soluções correspondentes no capítulo anterior, vamos nos concentrar nas partes dos algoritmos relacionadas à estimação do filtro linear $B(z)$.

As vantagens de nossa abordagem em relação aos métodos existentes são principalmente duas: (1) não precisa assumir conhecimento sobre a densidade espectral de potência do sinal original, ao contrário de [34] e (2) requer a estimação de um conjunto de parâmetros bastante reduzido em relação ao número de amostras do sinal, ao contrário de [36].

Organizamos o capítulo da seguinte forma. Na Seção 7.1, detalhamos o modelo adotado e formulamos mais precisamente o problema de restauração. Na seção seguinte, apresentamos o cálculo da verossimilhança para os parâmetros do modelo tendo como dados trechos curtos do sinal degradado. Na Seção 7.3, através de manipulação da verossimilhança, obtemos uma função-objetivo que servirá para a maximização da verossimilhança através do algoritmo de Gauss-Newton, também descrito nessa seção. Em seguida, na Seção 7.4, apresentamos a solução bayesiana para o problema através de um algoritmo baseado em MCMC, que pode usar como ponto de partida o resultado do algoritmo de Gauss-Newton, e encerramos o capítulo na Seção 7.5 com suas conclusões.

7.1 Formulação do problema

De acordo com a Figura 7.1, o sinal original $x(n)$ é gerado por um modelo AR, descrito por um filtro só-pólos excitado por ruído branco gaussiano $e(n)$ de variância σ_e^2 , assumidos fixos em trechos curtos do sinal. Esse sinal é distorcido pela combinação da curva não-linear sem memória descrita por $f[x(n)]$ com o filtro só-pólos $B(z)$, gerando o sinal observado $z(n)$, que, neste contexto, é conhecido exatamente.

O objetivo do algoritmo de restauração é identificar a curva $f[x(n)]$ e o filtro $B(z)$ com base no sinal $z(n)$, pois, uma vez identificada a distorção, o sinal original pode ser recuperado trivialmente através da transformação inversa de $B(z)$ e $f[x(n)]$.

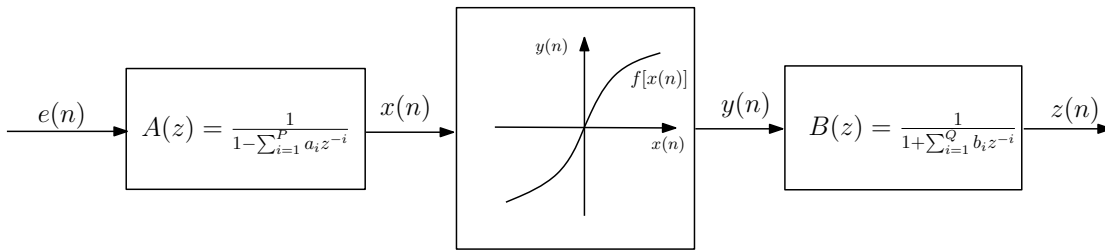


Figura 7.1: Modelo de distorção não-linear com memória.

Assim como no capítulo anterior, a função não-linear é modelada por sua expansão em série de Taylor, isto é:

$$x(n) = g[y(n)] = y(n) + \sum_{i=1}^M m_i y^{2i+1}(n). \quad (7.1)$$

No escopo desta tese, iremos assumir nos cálculos subsequentes que as ordens P , M e Q são conhecidas. Em [34], afirma-se que valores de M entre 7 e 9 e valores de Q entre 10 e 15 são adequadas para modelagem de sinais reais. Nos próximos passos do trabalho, planejamos tratar as ordens como variáveis e tentar estimá-las num contexto bayesiano, o que pode ser feito usando-se, por exemplo, o algoritmo de Metropolis-Hastings com saltos reversíveis [18].

7.2 Cálculo da verossimilhança

Como visto na Seção 6.3, a verossimilhança de \mathbf{y} condicionada aos parâmetros \mathbf{a} , \mathbf{m} e σ_e^2 e à condição inicial \mathbf{y}_0 é dada por:

$$p(\mathbf{y}|\mathbf{y}_0, \mathbf{a}, \mathbf{m}, \sigma_e^2) = \left(\frac{1}{2\pi\sigma_e^2} \right)^{\frac{N-P}{2}} \prod_{n=P+1}^N |g'(y(n), \mathbf{m})| \times \exp \left\{ -\frac{1}{2\sigma_e^2} \sum_{n=P+1}^N e^2(n, \mathbf{m}, \mathbf{a}) \right\}. \quad (7.2)$$

Como

$$z(n) = y(n) + \sum_{i=1}^Q b_i z(n-i), \quad (7.3)$$

e considerando que o jacobiano da transformação de $y(n)$ para $z(n)$ é unitário, a verossimilhança de $\mathbf{z} = [z(R+1)z(R+2) \dots z(N)]^T$ (sendo $R = P+Q$), condicionada a σ_e^2 , \mathbf{a} , \mathbf{m} e $\mathbf{b} = [b_1 b_2 \dots b_Q]^T$ e à condição inicial $\mathbf{z}_0 = [z(1)z(2) \dots z(R)]^T$ é obtida substituindo-se $y(n)$ em termos de $z(n)$ e \mathbf{b} em (7.2), o que resulta em:

$$p(\mathbf{z}|\mathbf{z}_0, \mathbf{a}, \mathbf{b}, \mathbf{m}, \sigma_e^2) = \left(\frac{1}{\sqrt{2\pi\sigma_e^2}} \right)^{N-R} \left| \prod_{n=R+1}^N s'[z(n), \mathbf{m}, \mathbf{b}] \right| \times \exp \left\{ -\frac{1}{2\sigma_e^2} e^2(n, \mathbf{b}, \mathbf{m}, \mathbf{a}) \right\}, \quad (7.4)$$

em que definimos

$$x(n) = s[z(n), \mathbf{m}, \mathbf{b}] = g \left[z(n) - \sum_{i=1}^Q b_i z(n-i), \mathbf{m} \right]. \quad (7.5)$$

O sinal de excitação $e(n)$ pode ser escrito em função de $z(n)$ da seguinte forma:

$$e(n) = x(n) - \sum_{i=1}^P a_i x(n-i) = s[z(n), \mathbf{m}, \mathbf{b}] - \sum_{i=1}^P a_i s[z(n-i), \mathbf{m}, \mathbf{b}]. \quad (7.6)$$

7.3 Maximização da verossimilhança

De forma similar ao que foi feito no capítulo anterior, através da manipulação da Equação (7.4), é possível mostrar que maximizar a verossimilhança é equivalente a

maximizar a seguinte função-objetivo:

$$V(\mathbf{b}, \mathbf{m}, \mathbf{a}) = \sum_{n=R+1}^N r_n^2, \quad (7.7)$$

$$r_n = g(\mathbf{b}, \mathbf{m})e(n, \mathbf{b}, \mathbf{m}, \mathbf{a}), \quad (7.8)$$

com

$$g(\mathbf{b}, \mathbf{m}) = \exp \left\{ -\frac{1}{N-R} \sum_{n=R+1}^N |\log(s'[z(n), \mathbf{m}, \mathbf{b}])| \right\}. \quad (7.9)$$

Para a maximização dessa função através do algoritmo de Gauss-Newton, precisamos do jacobiano, como definido na Equação (6.36). As derivadas de r_n em relação \mathbf{a} e \mathbf{m} se calculam como mostrado na Seção A.2, com \mathbf{y} obtido através da filtragem inversa de \mathbf{z} sendo $B(z)$ definido com base no valor atual de \mathbf{b} na iteração do algoritmo.

Calculamos as derivadas parciais de r_n em relação a cada um dos elementos de \mathbf{b} da seguinte forma:

$$\frac{\partial r_n}{\partial b_i} = \frac{\partial g}{\partial b_i} e(n) + g \frac{\partial e(n)}{\partial b_i}, \quad (7.10)$$

com

$$\begin{aligned} \frac{\partial g}{\partial b_i} &= \exp \left\{ -\frac{1}{N-R} \sum_{n=R+1}^N \log |s'[z(n), \mathbf{m}, \mathbf{b}]| \right\} \\ &\times \frac{1}{N-R} \frac{s''[z(n), \mathbf{m}, \mathbf{b}]z(n-i)}{s'[z(n), \mathbf{m}, \mathbf{b}]}. \end{aligned} \quad (7.11)$$

$$\frac{\partial e(n)}{\partial b_i} = -s'(z(n), \mathbf{b}, \mathbf{m})z(n-i) + \sum_{k=1}^P s'(z(n-k), \mathbf{b}, \mathbf{m})z(n-k-i). \quad (7.12)$$

A inicialização do algoritmo é problemática porque, ao contrário do que foi feito no capítulo anterior, não temos um método simples para obter uma estimativa inicial de boa qualidade para os parâmetros da não-linearidade e do filtro linear. Quanto aos coeficientes do modelo AR, verificamos em simulações que podemos obter uma estimativa inicial razoável tomando diretamente o sinal degradado $z(n)$ como se fosse o sinal original. Para os demais parâmetros, adotamos uma inicialização aleatória.

Apresentamos a seguir simulações para ilustrar o algoritmo proposto. O sinal

obtido com o modelo AR da Subseção 4.1.1 é distorcido com uma curva determinada pelos coeficientes $m_1 = 5$ e $m_2 = 30$ e em seguida por um filtro linear com coeficientes $b_1 = 1,13$ e $b_2 = 0,64$, que corresponde a polos com fase $+\pi/4$ e $-\pi/4$ e módulo 0,8.

Executando o algoritmo diversas vezes, observamos que são frequentes os casos em que os parâmetros convergem para valores distantes dos corretos, o que indica a presença de diversos mínimos locais. Entretanto, pelo menos nesse exemplo, o algoritmo converge para os valores esperados na maioria das vezes. Também observamos que se a inicialização for feita razoavelmente próxima dos valores corretos, os parâmetros convergem sempre para o máximo global. Isso indica que uma estimativa inicial de boa qualidade garantiria a convergência correta do algoritmo. Dentre os trabalhos futuros dessa tese, iremos incluir a elaboração de métodos simples para tal estimativa.

As Figuras 7.2, 7.3, 7.4 mostram a evolução dos parâmetros \mathbf{b} , \mathbf{m} e \mathbf{a} , respectivamente, para inicialização aleatória dos parâmetros \mathbf{b} e \mathbf{m} . Nesse caso, observamos convergência para o máximo global, mas é importante notar que nem sempre isso ocorre.

Recuperando o sinal original através das transformações inversas com base nos parâmetros estimados, pudemos observar que o sinal restaurado praticamente se confunde com o original, como mostrado nas Figuras 7.5 e 7.6. Isso significa que, aplicando o método a um sinal real, o sinal restaurado seria perceptualmente indistinguível do sinal original.

7.4 Solução bayesiana

Nesta seção apresentamos um procedimento para solução do problema num contexto bayesiano. Inicialmente, calculamos a distribuição *a posteriori*:

$$p(\mathbf{a}, \mathbf{b}, \mathbf{m}, \sigma_e^2 | \mathbf{z}) \propto p(\mathbf{z} | \mathbf{a}, \mathbf{b}, \mathbf{m}, \sigma_e^2) p(\mathbf{a}, \sigma_e^2) p(\mathbf{b}) p(\mathbf{m}), \quad (7.13)$$

em que se assumiu que os parâmetros do modelo AR (\mathbf{a} e σ_e^2), \mathbf{b} e \mathbf{m} são independentes *a priori*.

A solução do problema pelo amostrador de Gibbs envolve a amostragem iterativa a partir das distribuições condicionais totais para cada um dos parâmetros

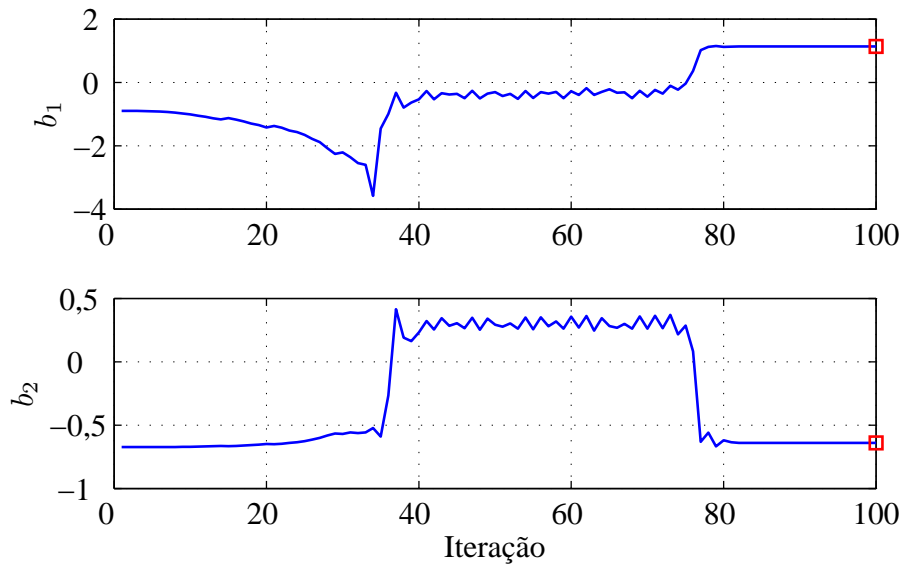


Figura 7.2: Evolução dos coeficientes do filtro linear $B(z)$ pelo método de Gauss-Newton.

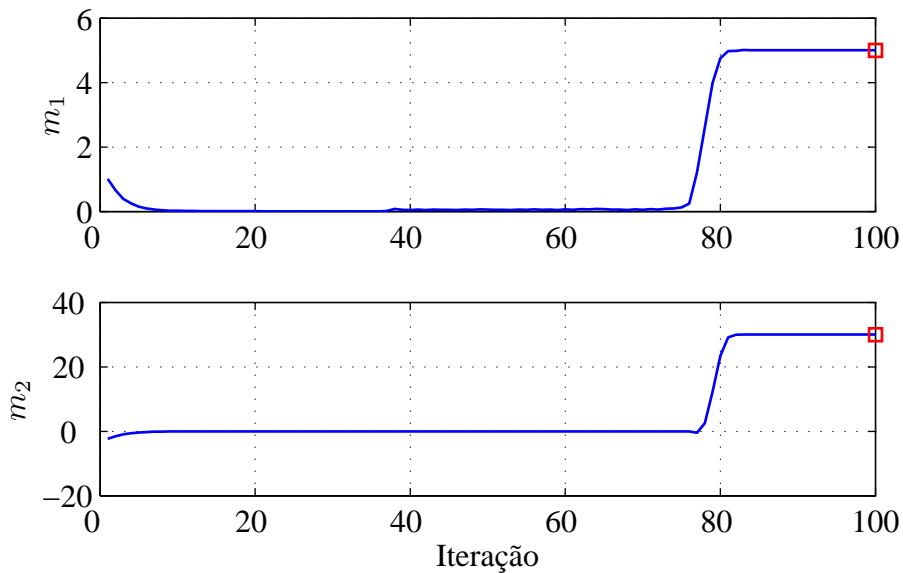


Figura 7.3: Evolução dos parâmetros da não-linearidade pelo método de Gauss-Newton.

envolvidos. A amostragem de \mathbf{a} , σ_e^2 e \mathbf{m} não apresenta qualquer dificuldade adicional em relação ao caso em que o filtro $B(z)$ não estava presente no modelo. Com efeito, como \mathbf{b} é suposto conhecido para a amostragem desses parâmetros (\mathbf{a} , σ_e^2 e \mathbf{m}), o sinal $y(n)$ será também conhecido, bastando para isso processar $z(n)$ pelo filtro inverso de $B(z)$.

A dificuldade reside na amostragem dos parâmetros \mathbf{b} , pois, devido à não-

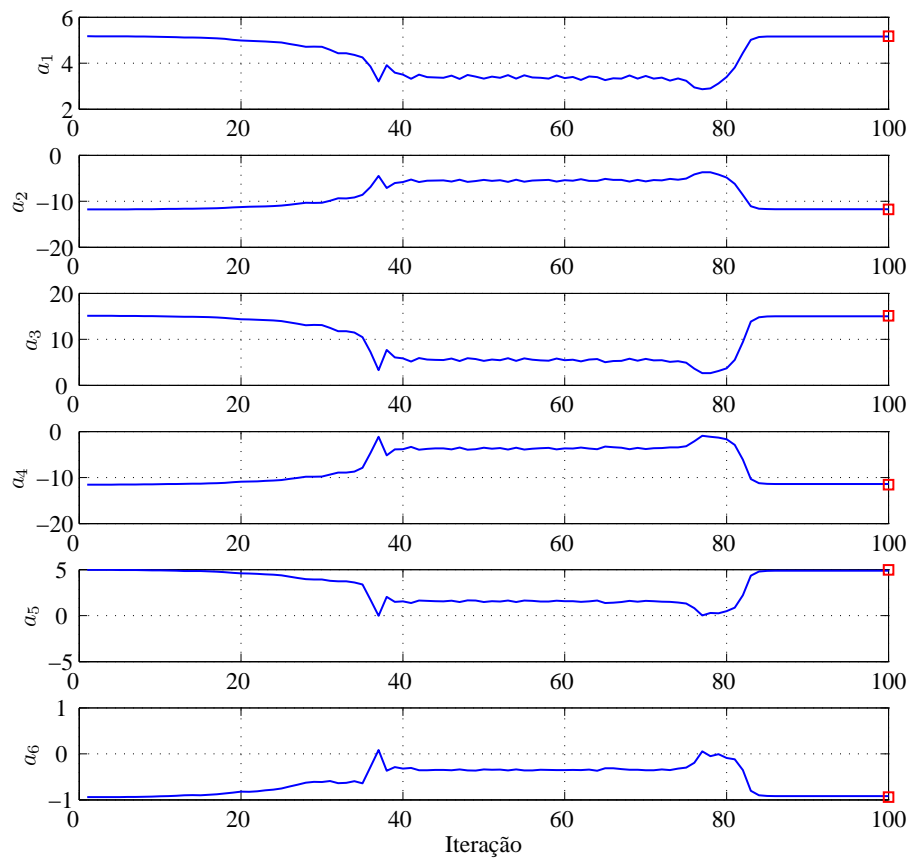


Figura 7.4: Evolução dos coeficientes do modelo AR pelo método de Gauss-Newton.

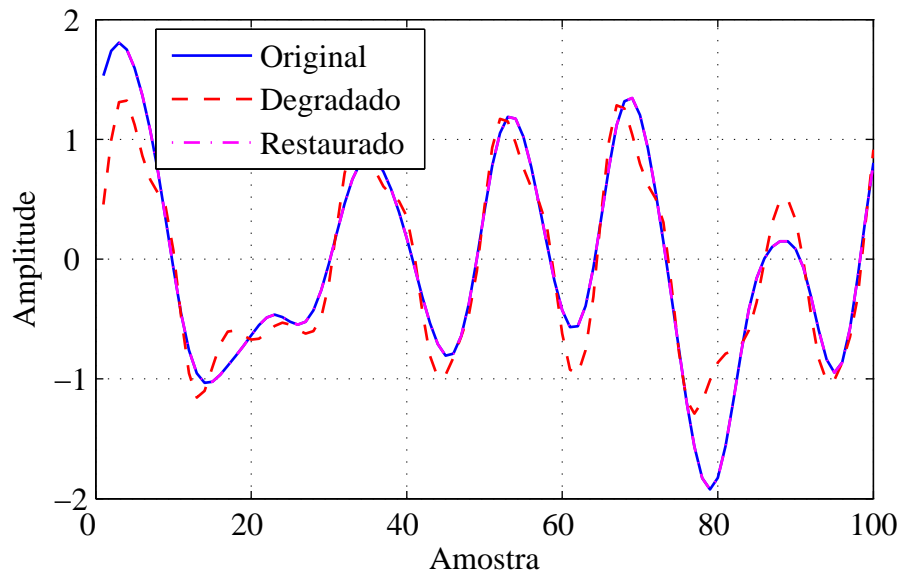


Figura 7.5: Comparação entre sinais original, degradado e restaurado pelo método de Gauss-Newton.

linearidade, a distribuição de \mathbf{b} não estará na forma de uma gaussiana. A distribui-

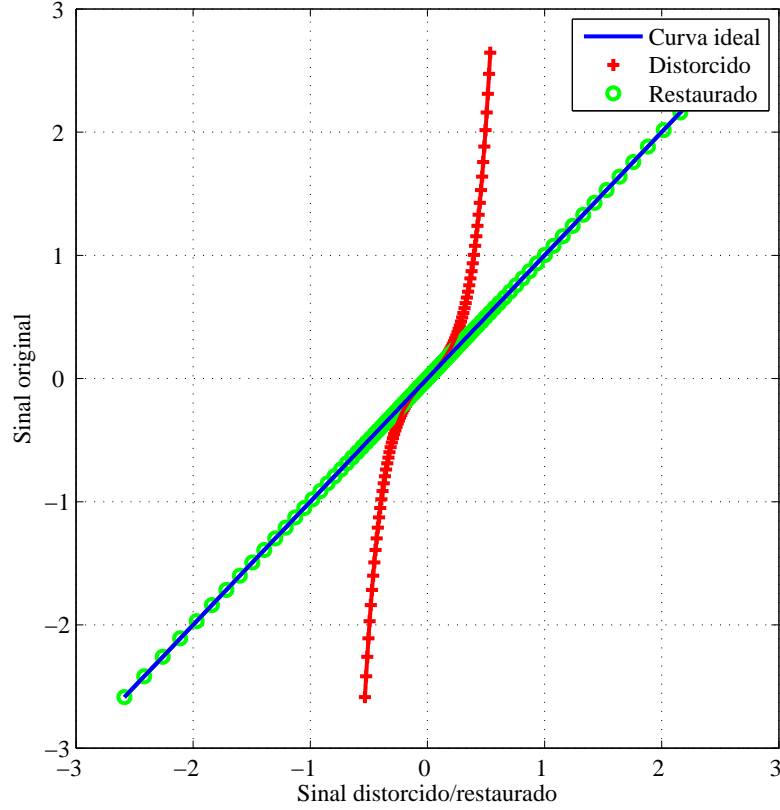


Figura 7.6: Sinal original contra sinal restaurado obtido com método de Gauss-Newton, comparados com curva ideal (identidade).

ção condicional total para \mathbf{b} é dada por:

$$p(\mathbf{b}|\mathbf{z}, \mathbf{a}, \sigma_e^2, \mathbf{m}) \propto \left| \prod_{n=R+1}^N s'[z(n), \mathbf{m}, \mathbf{b}] \right| \exp \left\{ -\frac{1}{2\sigma_e^2} \mathbf{e}^T \mathbf{e} \right\} p(\mathbf{b}). \quad (7.14)$$

Propomos aqui uma solução similar à adotada na Seção 6.5 para amostragem dos parâmetros \mathbf{m} . Lá, escrevemos a distribuição de \mathbf{m} na forma $\exp(L(\mathbf{m}))$ e calculamos uma aproximação de segunda ordem para o componente não quadrático de $L(\mathbf{m})$ de modo que o resultado fosse escrito como uma gaussiana, a partir da qual fosse fácil obter amostras. Podemos escrever a Equação (7.14) da seguinte forma:

$$p(\mathbf{b}|\mathbf{z}, \mathbf{a}, \sigma_e^2, \mathbf{m}) \propto \exp \left\{ \sum_{n=R+1}^N \log(s'[z(n), \mathbf{m}, \mathbf{b}]) \right\} \exp \left\{ -\frac{1}{2\sigma_e^2} \mathbf{e}^T \mathbf{e} \right\} p(\mathbf{b}). \quad (7.15)$$

Assumindo, por conveniência matemática, uma *priori* gaussiana para \mathbf{b} caracteri-

zada por

$$p(\mathbf{b}) \propto \exp \left\{ -\frac{1}{2}(\mathbf{b} - \boldsymbol{\mu}_b)^T \mathbf{C}_b^{-1}(\mathbf{b} - \boldsymbol{\mu}_b) \right\}, \quad (7.16)$$

a distribuição *a posteriori* de \mathbf{b} pode ser reescrita como:

$$p(\mathbf{b}|\mathbf{z}, \mathbf{a}, \sigma_e^2, \mathbf{m}) = \exp \left\{ L(\mathbf{b}) - \frac{1}{2}(\mathbf{b} - \boldsymbol{\mu}_b)^T \mathbf{C}_b^{-1}(\mathbf{b} - \boldsymbol{\mu}_b) \right\}, \quad (7.17)$$

com

$$L(\mathbf{b}) = \sum_{n=R+1}^N \log(s'[z(n), \mathbf{m}, \mathbf{b}]) - \frac{1}{2\sigma_e^2} \mathbf{e}^T \mathbf{e}, \quad (7.18)$$

que pode ser aproximada em torno de \mathbf{b}_0 por:

$$\hat{L}(\mathbf{b}) = L(\mathbf{b}_0) + \nabla_L(\mathbf{b}_0)(\mathbf{b} - \mathbf{b}_0) + \frac{1}{2}(\mathbf{b} - \mathbf{b}_0)^T \mathbf{H}_L(\mathbf{b}_0)(\mathbf{b} - \mathbf{b}_0). \quad (7.19)$$

Mostramos os detalhes do cálculo do gradiente e da hessiana na Seção A.3. Uma vez obtida $\hat{L}(\mathbf{b})$, pode-se obter uma aproximação gaussiana para $p(\mathbf{b})$ ao substituir $L(\mathbf{b})$ por $\hat{L}(\mathbf{b})$ em (7.14). A distribuição resultante pode ser usada como proposta para obtenção de amostras para \mathbf{b} , que então seria aceita ou rejeitada de acordo com o critério da Equação (3.21).

A distribuição proposta $q(\mathbf{b})$, gerada dessa forma, pode ser escrita como uma gaussiana padrão $q(\mathbf{b}) = \mathbf{N}(\mathbf{b}|\boldsymbol{\mu}_b, \mathbf{C}_b)$, com:

$$\mathbf{C}_b = - \left(\mathbf{H}_L(\mathbf{b}_0) - \frac{1}{\sigma_b^2} \mathbf{I}_Q \right)^{-1}, \quad (7.20)$$

$$\boldsymbol{\mu}_b = -\mathbf{C}_b (\mathbf{H}_L(\mathbf{b}_0)\mathbf{b}_0 - \nabla_L(\mathbf{b}_0)). \quad (7.21)$$

A estrutura do algoritmo proposto é similar à do descrito na Seção 6.5, com o acréscimo da amostragem de \mathbf{b} de acordo com o procedimento que acabamos de descrever.

Para ilustrar o método proposto, geramos um sinal distorcido com os mesmos parâmetros usados na seção anterior. A Figura 7.7 mostra os sinais original, distorcido e restaurado pelo algoritmo proposto. Na Figura 7.8, comparamos a curva usada para gerar o sinal distorcido com a curva estimada, podendo concluir que elas são praticamente iguais. As primeiras iterações dos parâmetros geradas pelo amostrador de Gibbs são mostradas nas Figuras 7.9, 7.10, 7.11 e 7.12, a partir das quais

podemos inferir que o algoritmo converge após cerca de 30 iterações. Para comprovar a convergência correta do algoritmo, construímos os histogramas dos parâmetros amostrados com base em 900 iterações após a a convergência, e os mostramos nas Figuras 7.13, 7.14, 7.15 e 7.12.

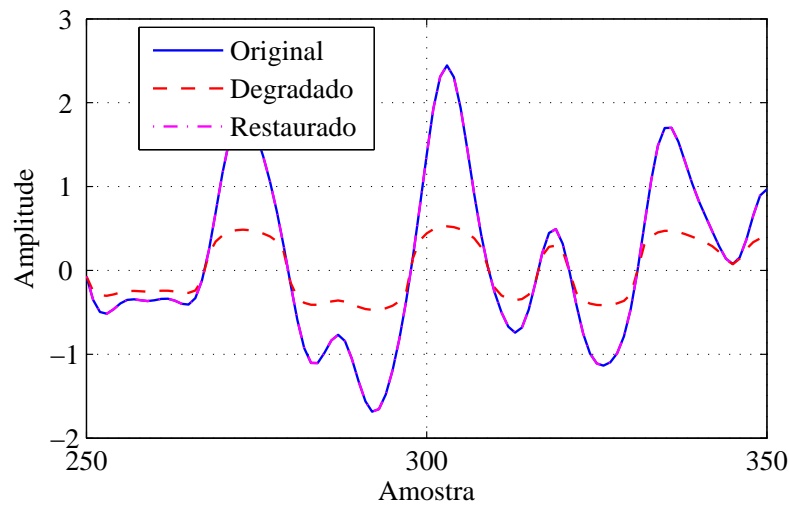


Figura 7.7: Comparação entre sinais original, degradado e restaurado por método bayesiano.

7.5 Conclusões

Neste capítulo descrevemos uma estrutura aplicável para a modelagem de sinais de áudio com distorções não-lineares com memória. A descrição matemática do modelo é apresentada juntamente com duas soluções estatísticas, uma baseada na maximização da verossimilhança através do método de Gauss-Newton, e outra bayesiana que permite obter amostras dos parâmetros relevantes através de métodos MCMC.

Uma dificuldade que deverá ser enfrentada para que se tenha uma solução efetiva são a estimativa inicial dos coeficientes do filtro $B(z)$ e dos parâmetros da distorção não-linear, necessários para se obter boas aproximações para as distribuições desses parâmetros. Além disso, a escolha de um *priori* adequadas para \mathbf{b} , idealmente inspirada num modelo físico, parece ser necessária para permitir que o filtro seja identificado.

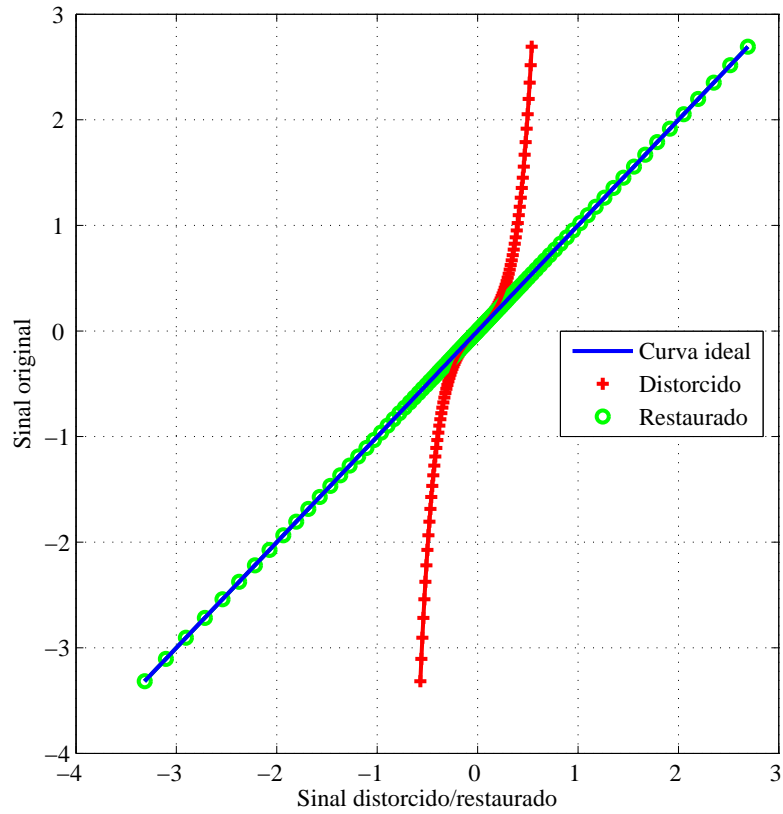


Figura 7.8: Sinal original contra sinal restaurado obtido com método de bayesiano, comparados com curva ideal (identidade).

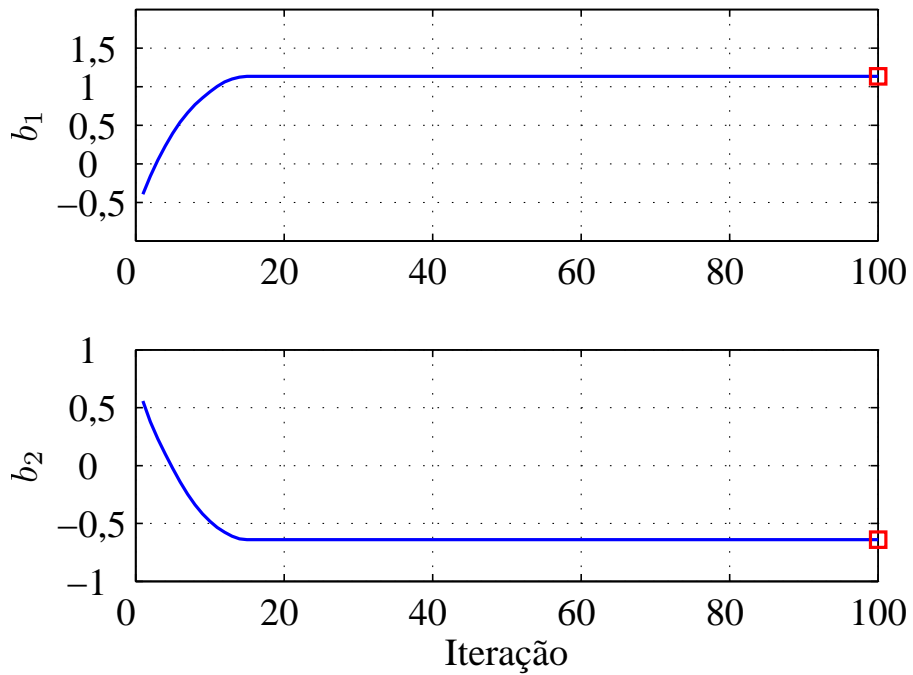


Figura 7.9: Evolução dos coeficientes do filtro linear através de amostrador de Gibbs.

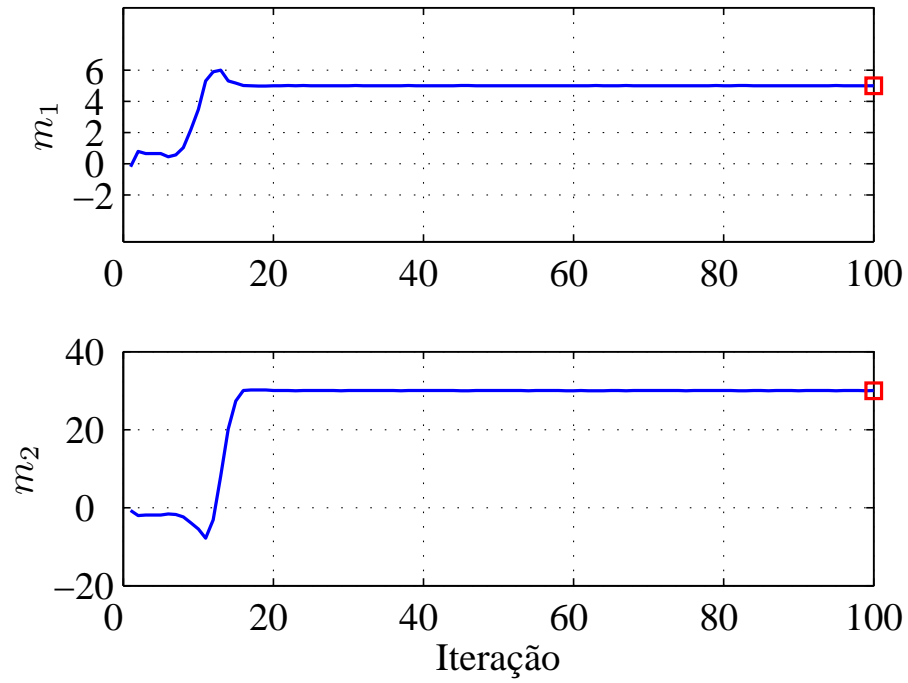


Figura 7.10: Evolução dos coeficientes da curva não linear através de amostrador de Gibbs.

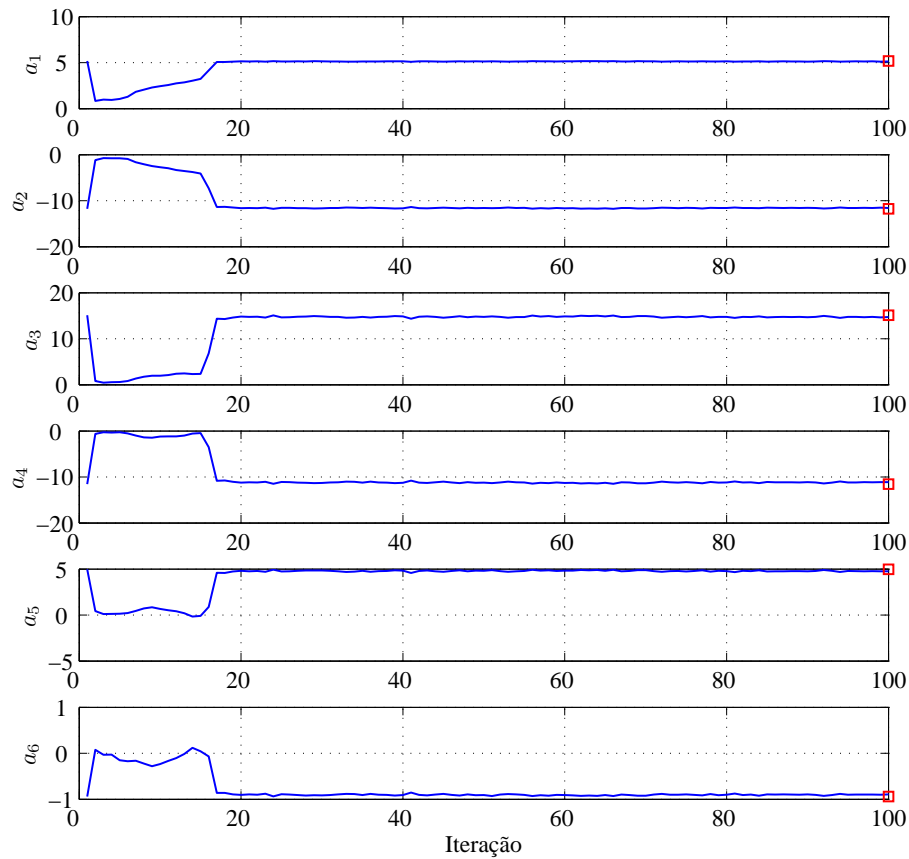


Figura 7.11: Evolução dos coeficientes do modelo AR através de amostrador de Gibbs.

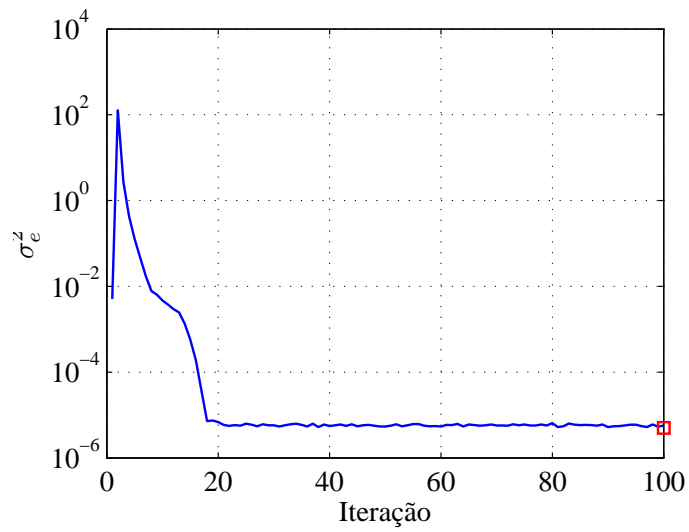


Figura 7.12: Evolução da variância de excitação do modelo AR através de amostrador de Gibbs.

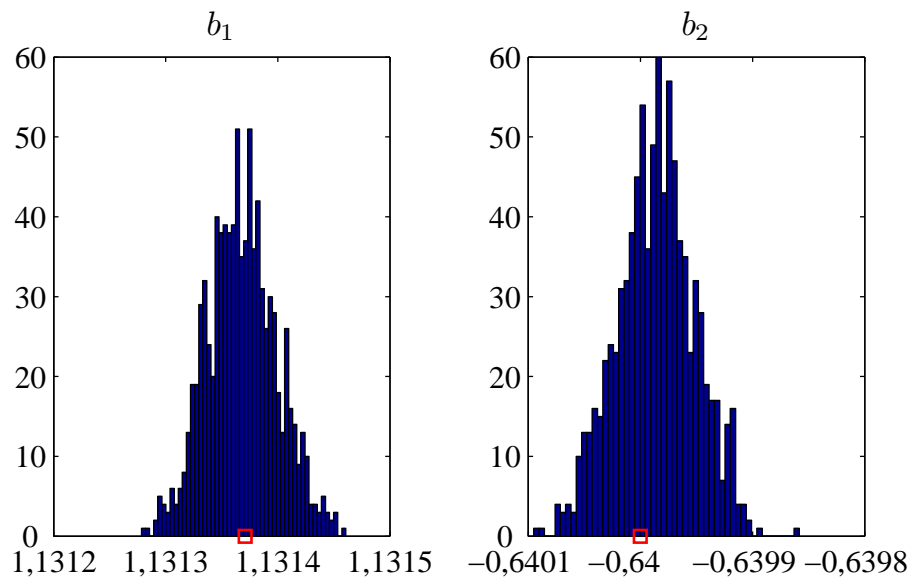


Figura 7.13: Histogramas dos coeficientes do filtro linear, obtida pelo amostrador de Gibbs.

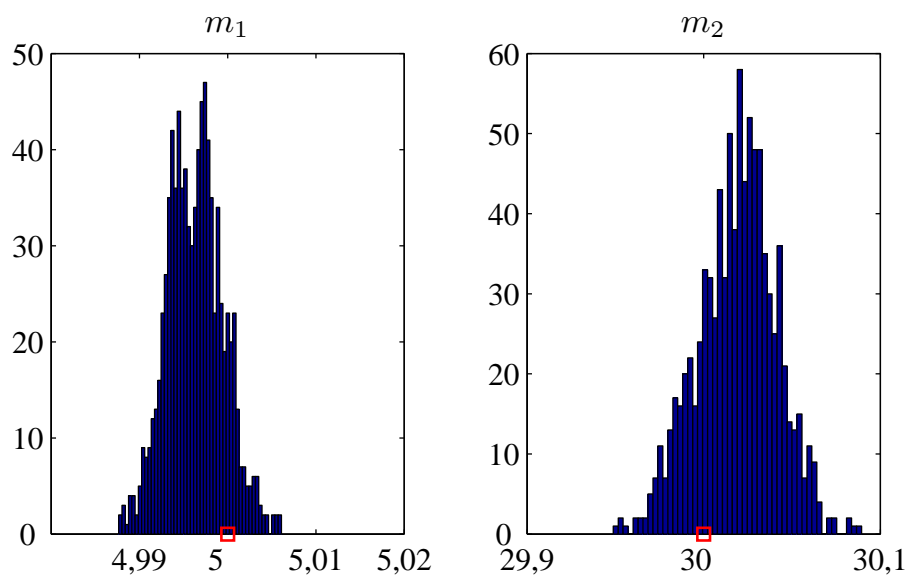


Figura 7.14: Histogramas dos coeficientes da curva não-linear, obtida pelo amostrador de Gibbs.

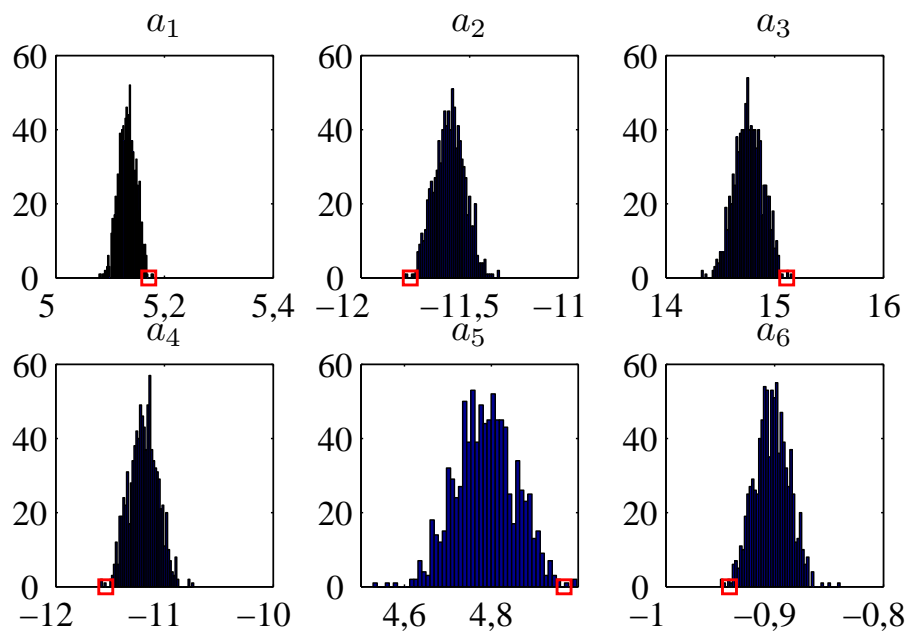


Figura 7.15: Histogramas de cada elemento dos coeficientes do modelo AR, pelo amostrador de Gibbs.

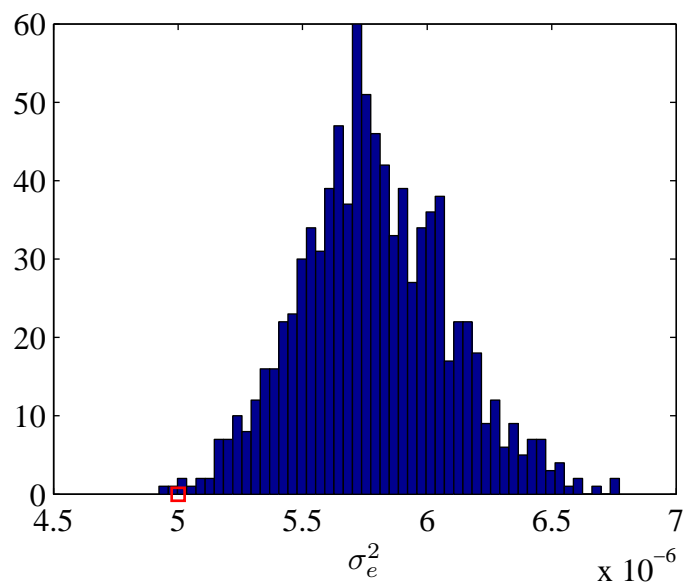


Figura 7.16: Histograma da excitação do modelo AR, obtido a partir das amostras geradas pelo amostrador de Gibbs.

Capítulo 8

Conclusões e Trabalhos Futuros

Esta tese se dedicou à análise e restauração de diversos tipos de distorções não-lineares presentes em sinais de áudio. Para os problemas de limitação de amplitude, saturação gradual sem memória e distorções não-lineares com memória com estrutura do tipo Hammerstein, propusemos soluções que variavam desde algoritmos simples e computacionalmente eficientes até bayesianos baseados em modelos. Os métodos mais simples podem ser usados tanto isoladamente quanto para gerar um ponto de partida favorável para os métodos mais sofisticados.

Nas próximas seções, resumimos as soluções propostas em cada uma das distorções abordadas, destacamos suas principais contribuições e indicamos algumas possíveis extensões. Concluimos o capítulo descrevendo duas novas abordagens para o tratamento de distorções não-lineares que planejamos desenvolver após a conclusão desta tese.

8.1 Limitação de amplitude

Limitação de amplitude, ou saturação radical, ocorre quando o sinal possui faixa dinâmica maior que a permitida pelo sistema de armazenamento, transmissão ou reprodução. Para este problema, consideramos dois casos principais: (1) sinal saturado sem ruído e (2) sinal saturado com ruído. Este último é sub-dividido em dois casos: (a) ruído antes ou (b) depois da saturação.

Dentro do contexto de processamento estatístico de sinais, o primeiro caso pode ser tratado de forma sub-ótima como um problema de interpolação de sinais

com amostras perdidas, ignorando a amplitude da região degradada. Implementamos o algoritmo descrito em [23] para termos uma solução de referência. Formulando o problema como interpolação com restrições, a solução proposta nesta tese permite resolver o problema de forma ótima, levando em consideração as restrições na amplitude e integrando numericamente as variáveis de *nuisance* através do amostrador de Gibbs.

Observamos que a solução proposta tende a resultar em sinais com maior SNR com relação à solução de referência, mas os ganhos se mostraram modestos na maioria dos sinais em que a testamos. Como na maioria das aplicações práticas a quantidade de amostras corretas é bastante grande, o método sub-ótimo já oferecia um resultado bastante acurado, que não podia ser muito melhorado por métodos mais sofisticados. Entretanto, pudemos observar diferenças substanciais de SNR quando introduzimos distorções muito severas.

Para a solução do problema na presença de ruído, até onde sabemos, não existe solução proposta na literatura. Em ambos os casos considerados, conseguimos aumentar significativamente a SNR dos sinais tratados através de algoritmos iterativos baseados no amostrador de Gibbs. Como os algoritmos envolvem a amostragem de gaussianas da ordem do número de amostras dos sinais e o número de iterações necessários para convergência é bastante alto, a complexidade final da versão mais básica da solução proposta é bastante elevada. Conseguimos reduzir o número de iterações para convergência ao inicializar os coeficientes do modelo AR em um bloco do sinal com os valores estimados no bloco anterior, assumindo que estes parâmetros não variam muito de um bloco para outro. Além disso, propusemos uma variação simples do algoritmo em que cada ciclo do sinal é processada por vez, aumentando a probabilidade de aceitação nas regiões com saturação, sem aumentar significativamente a correlação na cadeia de Markov.

Como prosseguimento dessa linha de trabalho, pensamos na adaptação dos algoritmos para o problema de restauração de sinais quantizados com *dither*, que também pode ser descrito como uma não-linearidade não-invertível combinada com ruído aditivo.

8.2 Não-linearidade invertível sem memória

Consideramos um caso particular em que a curva é antissimétrica, que acreditamos corresponder à maioria dos casos, embora a generalização não seja difícil. Modelamos a curva de distorção de duas maneiras: (1) expansão em séries de Taylor em torno da origem com coeficientes desconhecidos e (2) curva linear por partes com coeficientes angulares desconhecidos. Desenvolvemos mais o modelo 1, usando-o em algoritmos de maximização da verossimilhança dos coeficientes pelo método de Gauss-Newton, obtenção de amostras da distribuição *a posteriori* dos coeficientes por um algoritmo baseado em MCMC e na extensão deste para a recuperação do sinal original quando ruído aditivo está presente no sinal degradado. Aplicamos o modelo 2 apenas num algoritmo baseado em MCMC para estimação dos coeficientes angulares das retas, assumindo ausência de ruído, mas acreditamos que seu desempenho nos mesmos casos considerados com o modelo 1 seria igualmente bem-sucedida.

Embora a aproximação da curva não-linear por séries polinomiais já tenha sido proposta em várias aplicações, até onde sabemos, nosso trabalho é o primeiro a resolver o problema de identificação desse tipo de distorção no contexto de inferência bayesiana. Acreditamos que o modelo linear por partes, aplicado nesse contexto, seja inteiramente original. Além disso, a modelagem estatística do ruído aditivo em combinação com a não-linearidade sem memória é uma contribuição original da tese.

Como possível melhoria no algoritmo, planeja-se permitir que a ordem da função não-linear seja variável, podendo ser determinada a partir do sinal observado, através do algoritmo de Metropolis-Hastings com saltos reversíveis [37].

8.3 Distorção não-linear com memória

Para este problema, adotamos um modelo formado pela combinação em cascata de uma curva não-linear sem memória com um filtro linear (estrutura conhecida como de Hammerstein), o qual acreditamos ser capaz de descrever boa parte das distorções não-lineares encontradas em sistemas reais. Duas soluções foram propostas: uma baseada na maximização da verossimilhança dos coeficientes da curva não-linear e do filtro linear pelo algoritmo de Gauss-Newton e outra na estimação bayesiana desses coeficientes por um algoritmo baseado em MCMC.

Os algoritmos propostos funcionaram bem com distorções artificialmente introduzidas nos sinais, respeitando a estrutura assumida pelo algoritmo, mas ainda não os testamos extensivamente em sinais reais. A dificuldade é encontrar exemplos em que saibamos antecipadamente que a distorção possa ser caracterizada pela estrutura proposta, visto que tanto a análise visual quanto auditiva não fornece muitas pistas. Para testar os algoritmos, uma alternativa intermediária entre sinais artificiais e distorções reais seria usar modelos de não-linearidade representativos de dispositivos reais para gerar um conjunto de sinais distorcidos.

8.4 Trabalhos futuros

Além dos avanços naturais nos algoritmos propostos, como mencionamos anteriormente, planejamos continuar a pesquisa com distorções não-lineares buscando na literatura, em particular na área de análise de séries temporais e modelagem esparsa de sinais, técnicas que permitam lidar com os casos para os quais os métodos propostos nesta tese se mostraram insatisfatórios.

Propomos continuar a pesquisa em duas linhas alternativas, uma em que o sinal de áudio é tratado no domínio do tempo e outra em que o tratamos no domínio da transformada. Além disso, propomos levar em consideração elementos psicoacústicos associados à percepção de distorções não-lineares. Nas seções seguintes, discutimos alguns caminhos que pretendemos seguir em cada uma dessas linhas de pesquisa e sugerimos de que forma modelos psicoacústicos podem ser incorporados aos algoritmos de restauração.

8.4.1 Solução no domínio do tempo

Dentro dessa linha de modelagem no domínio do tempo, nossa proposta é estudar a literatura de séries temporais não-lineares, buscando uma estrutura que seja adequada à modelagem de distorções não-lineares em sinais de áudio. Um candidato inicial a ser estudado e implementado é o modelo autorregressivo com limiar, TAR (do inglês, *Threshold Autorregressive*) [38], que pode ser entendido como uma extensão não-linear do modelo AR tradicional. Proposto por H. Tong na década de 80 [39], esse modelo vem sendo usado com sucesso em diversas áreas de estatística

aplicada e engenharia [40], mas seu uso na área de processamento de sinais ainda é limitado.

Matematicamente, o modelo é definido da seguinte forma:

$$x(n) = a_0^{(j_n)} + \sum_{i=1}^P a_i^{(j_n)} x(n-i) + b^{(j_n)} e(n), \quad (8.1)$$

onde J_n , $n = \{1, \dots, J\}$, indica um modo particular permitido pelo modelo. Se os valores de J_n forem determinados pelas saídas passadas de $x(n)$, o modelo é chamado SETAR (do inglês *Self-Exciting Threshold Autoregressive*). Um exemplo com $J = 2$ e ordem 1 seria o seguinte:

$$x(n) = \begin{cases} a_0 + a_1 x(n-1) + e(n), & x(n-d) \leq r; \\ a_2 + a_3 x(n-1) + b e(n), & x(n-d) > r. \end{cases} \quad (8.2)$$

Foram publicados diversos trabalhos sobre a análise do modelo TAR e foram propostas aplicações suas em diversos problemas, especialmente em modelagem de séries econômicas. Nosso plano é estudar essa literatura e adaptar o modelo à descrição de sinais de áudio com distorções não-lineares. Uma vez tendo um modelo adequado para o sinal distorcido, a dificuldade maior seria separar a parte linear da parte não-linear do modelo resultante, considerando que só temos acesso à combinação das duas, que é não-linear.

8.4.2 Solução no domínio da transformada

Uma outra possível abordagem para a solução do mesmo problema seria explorar as propriedades do sinal de áudio no domínio de uma transformada apropriadamente escolhida. Sabe-se que sinais de áudio podem ser representados em bases esparsas, isto é, com poucos coeficientes diferentes de zero. Como a não-linearidade tende a destruir essa esparsidade, o sinal original poderia, em princípio, ser recuperado, através de uma operação inversa que a regerasse.

No contexto bayesiano, a esparsidade pode ser assegurada através das chamadas *prioris indutoras de esparsidade* (*sparsity inducing priors*). Soluções para representação e restauração de sinais esparsos foram propostas em diversos contextos,

por exemplo [41]. Em particular, a transformada de cosseno modificada (MDCT) com prioris indutoras de esparsidade foram usadas para modelagem e restauração de sinais de áudio corrompidos por ruído de fundo [42]. Uma idéia similar foi aplicada [43] ao problema de restauração de imagens ruidosas. Nosso plano é usar os métodos mencionados como ponto de partida para a solução de nosso problema.

8.4.3 Elementos psicoacústicos

Uma área ativa de pesquisa em processamento de sinais é a avaliação automática de sinais de áudio e voz, cujo objetivo é criar métodos capazes de gerar notas para sinais com degradação que sejam altamente correlacionados com notas médias geradas por um grupo de ouvintes. Algoritmos padronizados pela União Internacional de Telecomunicações (ITU), como o PESQ [44], o P.563 [45] e o POLQA [46], permitem medidas de qualidade confiáveis e replicáveis para uma faixa ampla de cenários de distorções. Embora essa área de pesquisa esteja bem desenvolvida, o uso de fenômenos psicoacústicos para restauração ainda é limitada. Algumas tentativas bem-sucedidas [47–50], tanto heurísticas quanto estatísticas, lidam apenas com o problema de melhoramento de sinais com ruído aditivo de faixa larga, e não podem ser facilmente generalizados para outros tipos de distorção.

Estudos sobre o efeito perceptivo de distorções não-lineares em sinais de áudio aparecem numa série de artigos publicados entre 2003 e 2004 [51], [52] e [53]. Neles, os autores reportam experimentos em que ouvintes avaliam um conjunto abrangente de sinais com distorções não-lineares, os quais servem de base para o desenvolvimento de um método objetivo para avaliação automática de distorções não-lineares.

A maneira como iremos usar aspectos psicoacústicos em nossos trabalhos ainda não está definida, mas uma idéia inicial seria gerar uma base de dados contendo um conjunto de sinais com distorções artificialmente introduzidas em sinais sem distorção, e através do algoritmo *Rnonlin* [53], estudar o efeito de parâmetros dos modelo de distorção na qualidade do sinal degradado. Com base nessa informação, esperamos ter condições de propor funções-objetivo que dêem um peso maior para os erros mais perceptíveis.

Referências Bibliográficas

- [1] BISCAINHO, L. W., ESQUEF, P. “DSP Techniques for Sound Enhancement of Old Recording”. In: Pérez-Meana, H. (Ed.), *Advances in Audio and Speech Signal Processing*, Idea Group, pp. 93–128, Hershey, EUA, 2007.
- [2] GODSILL, S. J., RAYNER, P. J. W. *Digital Audio Restoration - A Statistical Model Based Approach*. Cambridge, Reino Unido, Cambridge University Press, 1998.
- [3] CHRISTMAS, J., EVERSON, R. “Robust Autoregression: Student-t Innovations Using Variational Bayes”, *IEEE Trans.on Signal Proc.*, v. 59, n. 1, pp. 48–57, Jan. 2011.
- [4] VERMAAK, J., ANDRIEUW, C., DOUCET, A., et al. *Non-Stationary Bayesian Modelling and Enhancement of Speech Signals*. Technical Report CUED/F-INFENG/TR.351, Cambridge University Engineering Department, Cambridge, Reino Unido, 1999.
- [5] P. A. A. ESQUEF, V. V., KARJALAINEN, M. “Restoration and Enhancement of Solo Guitar Recordings Based on Sound Source Modeling”, *Journal of the Audio Engineering Society*, v. 50, n. 4, pp. 227–236, Abril 2002.
- [6] MOORE, B. C. J. *An Introduction to the Psychology of Hearing*. San Diego, EUA, Elsevier, 2004.
- [7] KABAL, P. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. Technical report, McGill University, Montreal, Canadá, 2002.
- [8] ÁVILA, F. R. *Algoritmos Baseados em Modelos Bayesianos para Restauração de Sinais de Áudio*. Tese de mestrado, COPPE/UFRJ Laboratório de Processamento de Sinais, Rio de Janeiro, RJ, Brasil, 2008. <http://www.pee.ufrj.br/teses/index.php?Resumo=2008033101>.
- [9] BERNARDO, J. M., SMITH, A. F. M. *Bayesian Theory*. New Jersey, USA, John Wiley & Sons, 2007.

- [10] MIGON, H. S., GAMERMAN, D. *Statistical Inference: An Integrated Approach*. Londres, Reino Unido, Hodder Arnold, 1999.
- [11] HAJEK, A. “Interpretations of Probability”. 2007. <http://plato.stanford.edu/entries/probability-interpret/>.
- [12] R. O. DUDA, P. E. HART, D. G. S. *Pattern Classification*. New Jersey, USA, Jhon Wiley & Sons, 2001.
- [13] KAY, S. M. *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. New Jersey, USA, Prentice Hall, 1993.
- [14] JAYNES, E. “Prior Probabilities”, *IEEE Transactions on Systems Science and Cybernetics*, v. 4, n. 3, pp. 227–241, Set. 1978.
- [15] BERNARDO, J. M. “Reference Posterior Distributions for Bayesian Inference”, *Journal of the Royal Statistical Society*, v. 41, n. 2, pp. 113–147, Dez. 1979.
- [16] ANDRIEU, C., DE FREITAS, N., DOUCET, A., et al. “An Introduction to MCMC for Machine Learning”, *Machine Learning*, v. 50, n. 1-2, pp. 5–43, Jan.-Fev. 2003.
- [17] NEAL, R. *Probabilistic Inference Using Markov Chain Monte Carlo Methods*. Technical Report CRG-TR-93-1, University of Toronto, Department of Computer Science, Toronto, Canadá, 1993.
- [18] GAMERMAN, D., LOPES, H. *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. Boca Raton, USA, Chapman & Hall/CRC, 2006.
- [19] ROBERT, C., CASELLA, G. *Monte Carlo Statistical Methods*. Nova Iorque, EUA, Springer, 2004.
- [20] GEMAN, S., GEMAN, D. “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 6, n. 6, pp. 721–741, 1984.
- [21] SCHAFER, R. W., MERSEREAU, R. M., RICHARDS, M. A. “Constrained Iterative Restoration Algorithms”, *Proceedings of the IEEE*, v. 69, n. 4, pp. 432–450, Abril 1981.
- [22] TROUGHTON, T. *Simulation Methods for Linear and Nonlinear Time Series Models with Application to Distorted Audio Signals*. Tese de doutorado, University of Cambridge Department of Engineering, Cambridge, Reino Unido, 1999.

- [23] JANSEN, A. J. E. M., VELDHUIS, R., VRIES, L. B. “Adaptive Interpolation of Discrete-time Signals that Can be Modeled as AR Processes”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. 34, n. 2, pp. 317–330, Abril 1986.
- [24] ABEL, J. S., SMITH, J. O. “Restoring a Clipped Signal”. In: *International Conference in Acoustic, Speech and Signal Processing (ICASSP)*, pp. 1745–1748, Toronto, Canadá, Abril 1991. IEEE.
- [25] ADLER, A., EMIYA, V., JAFARI, M., et al. “A Constrained Matching Pursuit Approach to Audio Declipping”. In: *Proc. of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011*, pp. 329–333, Praga, República Tcheca, Maio 2011. IEEE.
- [26] OLOFSSON, T. “Deconvolution and Model-Based Restoration of Clipped Ultrasonic Signals”, *IEEE Trans. on Instrum. and Measur.*, v. 54, n. 3, pp. 1235–1240, Junho 2005.
- [27] FONG, W., GODSILL, S. “Monte Carlo Smoothing for Non-Linearly Distorted Signals”. In: *International Conference in Acoustic, Speech and Signal Processing (ICASSP)*, pp. 3997–4000, Salt Lake City, EUA, Maio 2001. IEEE.
- [28] A. P. DEMPSTER, N. M. L., RUBIN, D. B. “Maximum Likelihood from Incomplete Data via the EM Algorithm”, *Journal of the Royal Statistical Society*, v. 39, n. 1, pp. 1–38, 1977.
- [29] GODSILL, S. J., RAYNER, P. J. W., CAPPÉ, O. “Digital Audio Restoration”. In: Kahrs, M., Brandenburg, K. (Eds.), *Applications of Digital Signal Processing to Audio and Acoustics*, Kluwer Academic Publishers, ISBN 0-7923-8130-0, pp. 133–193, Nova Iorque, EUA, 1998.
- [30] L. VANBEYLEN, R. P., SCHOUKENS, J. “Blind Maximum-Likelihood Identification of Wiener Systems”, *IEEE Transactions on Signal Processing*, v. 57, n. 9, pp. 3017–3029, Agos. 2009.
- [31] PICARD, G., CAPPÉ, O. “Blind Identification of Hammerstein Nonlinear Distortion Models”. In: *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 17–20, New Paltz, NY, USA, October 2003. IEEE.
- [32] WHITE, S. A. “Restoration of Nonlinearly Distorted Audio by Histogram Equalization”, *Journal of the Audio Engineering Society*, v. 30, n. 11, pp. 828–832, Nov. 1982.

- [33] KLIPPEL, W. “Compensation for Nonlinear Distortion of Horn Loudspeakers by Digital Signal Processing”, *Journal of the Audio Engineering Society*, v. 44, n. 11, pp. 964–972, Nov. 1996.
- [34] PICARDI, G., CAPPÉ, O. “Blind Identification of Hammerstein Nonlinear Distortion Models”. In: *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 19–22, New Paltz, EUA, Out. 2003. IEEE.
- [35] TROUGHTON, P. T., GODSILL, S. J. “Restoration of Nonlinearly Distorted Audio Using Markov Chain Monte Carlo Methods”. In: *Proc. 104th AES Convention*, pp. 1–9, Amsterdã, Holanda, Maio 1998. AES.
- [36] TROUGHTON, P., GODSILL, S. “MCMC Methods for Restoration of Nonlinearly Distorted Autoregressive Signals”, *Special Section on Markov Chain Monte Carlo (MCMC) Methods for Signal Processing*, v. 81, n. 1, pp. 83–97, 2001.
- [37] GREEN, P. “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination”, *Biometrika*, v. 82, n. 4, pp. 711–732, Dez. 1995.
- [38] TONG, H. *Threshold Models in Time Series Analysis-30 Years On*. Technical Report 471, The University of Hong Kong, Hong Kong, 2010.
- [39] CHAN, K. S., TONG, H. “On Estimating Thresholds in Autoregressive Models”, *Journal of Time Series Analysis*, v. 7, n. 3, pp. 179–190, Maio 1986.
- [40] TSAY, R. S. “Testing and Modeling Threshold Autoregressive Processes”, *Journal of the American Statistical Association*, v. 84, n. 405, pp. 231–240, Março 1989.
- [41] BABACAN, S. D., MOLINA, R., KATSAGGELOS, A. “Bayesian Compressive Sensing Using Laplace Priors”, *IEEE Trans. on Image Processing*, v. 19, n. 1, pp. 53–63, Jan. 2010.
- [42] FEVOTTE, C., TORRÉSANI, B., DAUDET, L., et al. “Sparse Linear Regression with Structured Priors and Application to Denoising of Musical Audio”, *IEEE Transactions on Audio, Speech and Language Processing*, v. 16, n. 1, pp. 174–185, Jan. 2008.
- [43] DOBIGEON, N., HERO, A. O., TURNERET, J. Y. “Hierarchical Bayesian Sparse Image Reconstruction with Application to MRFM”, *IEEE Transactions on Image Processing*, v. 18, n. 9, pp. 2059–2070, Set. 2009.

- [44] RIX, A., HOLLIER, M., HEKSTRA, A., et al. “Perceptual Evaluation of Speech Quality (PESQ), The New ITU Standard for End-to-End Speech Quality Assessment, Part I – Time-Delay Compensation”, *Journal of Audio Engineering Society*, v. 50, n. 10, pp. 755–764, Out. 2002.
- [45] MALFAIT, L., BERGER, J., KASTNER, M. “P.563–The ITU-T Standard for Single-Ended Speech Quality Assessment”, *IEEE Trans. on Audio, Speech, and Lang. Processing*, v. 14, n. 6, pp. 1924–1934, Nov. 1996.
- [46] ITU-T REC. P.863. “Perceptual objective listening quality assessment”. International Telecommunication Union, Jan., Geneva, Suíça 2011.
- [47] WOLFE, P., GODSILL, S. “Perceptually Motivated Approaches to Music Restoration”, *Journal of the New Music Research*, v. 30, n. 1, pp. 83–92, Jan. 2001.
- [48] HANSEN, J. H., RADHAKRISHNAN, V., AREHART, K. H. “Speech Enhancement Based on Generalized Minimum Mean Square Errors Estimators and Masking Properties of the Auditory System”, *IEEE Trans. on Audio, Speech and Lang. Processing*, v. 14, n. 6, pp. 2049–2063, Nov. 2006.
- [49] LOIZOU, P. “Speech Enhancement Based on Perceptually Motivated Bayesian Estimators of the Magnitude Spectrum”, *IEEE Trans. on Audio, Speech and Lang. Processing*, v. 13, n. 5, pp. 857–869, Nov. 2005.
- [50] PLOURDE, E., CHAMPAGNE, B. “Auditory-Based Spectral Amplitude Estimators for Speech Enhancement”, *IEEE Trans. on Audio, Speech and Lang. Processing*, v. 16, n. 8, pp. 1614–1623, Nov. 2008.
- [51] TAN, C. T., BRIAN, A., MOORE, C. J., et al. “The Effect of Nonlinear Distortion on the Perceived Quality of Music and Speech Signals”, *Journal of the Audio Engineering Society*, v. 51, n. 11, pp. 1012–1031, Nov. 2003.
- [52] TAN, C. T., A, BRIAN, et al. “Measuring and Predicting the Perceived Quality of Music and Speech Subjected to Combined Linear and Nonlinear Distortion”, *Journal of the Audio Engineering Society*, v. 52, n. 12, pp. 1228–1244, Dez. 2004.
- [53] TAN, C. T., BRIAN, A., MOORE, C. J., et al. “Predicting the Perceived Quality of Nonlinearly Distorted Music and Speech Signals”, *Journal of the Audio Engineering Society*, v. 52, n. 7, pp. 699–711, Julho 2003.
- [57] ÁVILA, F. “Examples of Restored Signals”. 2012. <http://www02.lps.ufrj.br/flavio/dscthesis.html>.

Apêndice A

Algumas Demonstrações

Este apêndice é dedicado à demonstração de alguns teoremas e expressões usados na tese.

A.1 Produto de gaussianas multivariáveis

Seja \mathbf{x} uma variável aleatória multidimensional, cuja distribuição é definida como o produto de duas gaussianas com médias e matrizes de covariância quaisquer:

$$p(\mathbf{x}) = kp_1(\mathbf{x})p_2(\mathbf{x}), \quad (\text{A.1})$$

com $p_1(\mathbf{x}) = N(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{C}_1)$, $p_2(\mathbf{x}) = N(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{C}_2)$ e constante k tal que a integral de $p(\mathbf{x})$ resulte em 1.

Podemos escrever a distribuição resultante no formato $p(\mathbf{x}) = \exp\{L(\mathbf{x})\}$, com

$$L(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \mathbf{C}_1^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \mathbf{C}_2^{-1}(\mathbf{x} - \boldsymbol{\mu}_2). \quad (\text{A.2})$$

Como $L(\mathbf{x})$ é quadrático, segue que a distribuição resultante também é gaussiana, e sua distribuição pode ser escrita como:

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_p)^T \mathbf{C}_p^{-1}(\mathbf{x} - \boldsymbol{\mu}_p)\right\}. \quad (\text{A.3})$$

Podemos obter $\boldsymbol{\mu}_p$ igualando a zero o gradiente do argumento da exponencial

e resolvendo para $\mathbf{x} = \boldsymbol{\mu}_p$. Aplicando esse procedimento em $L(\mathbf{x})$, obtemos:

$$\boldsymbol{\mu}_p^T = (\boldsymbol{\mu}_1^T \mathbf{C}_1^{-1} + \boldsymbol{\mu}_2^T \mathbf{C}_2^{-1}) (\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})^{-1}. \quad (\text{A.4})$$

Calculando a hessiana do argumento da exponencial, obtemos $\mathbf{H} = -\mathbf{C}_p^{-1}$. Portanto a matriz de covariância da distribuição resultante é dada pelo negativo da inversa da hessiana de $L(\mathbf{x})$, resultando em:

$$\mathbf{C}_p = (\mathbf{C}_1^{-1} + \mathbf{C}_2^{-1})^{-1}. \quad (\text{A.5})$$

A.2 Cálculo do Jacobiano para o método de Gauss-Newton

Como vimos na Seção 6.4.2, a função-objetivo é dada por:

$$V(\mathbf{m}, \mathbf{a}) = \sum_{n=P+1}^N r_n^2, \quad (\text{A.6})$$

$$r_n = g(\mathbf{m})e(n, \mathbf{a}, \mathbf{m}), \quad (\text{A.7})$$

$$g(m) = \exp \left\{ -\frac{1}{N-P} \sum_{n=P+1}^N \log(1 + h_n \mathbf{m}) \right\}, \quad (\text{A.8})$$

$$e(n, \mathbf{a}, \mathbf{m}) = g[y(n)] - \sum_{i=1}^P a_i g[y(n-i)], \quad (\text{A.9})$$

$$\mathbf{e} = \mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{y} + \mathbf{Y}\mathbf{m}) \quad (\text{A.10})$$

Com base nessas expressões, podemos calcular o Jacobiano, definido por:

$$\mathbf{J}_{ni} = \frac{\partial r_n}{\partial \theta_i}, \quad (\text{A.11})$$

com $\boldsymbol{\theta} = \{\mathbf{m}, \mathbf{a}\}$. Definindo $\mathbf{r} = [r_1 \ r_2 \ \dots \ r_N]^T$, podemos escrever:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{r}}{\partial \mathbf{m}} & \frac{\partial \mathbf{r}}{\partial \mathbf{a}} \end{bmatrix} \quad (\text{A.12})$$

Usando a regra da cadeia, obtemos

$$\frac{\partial \mathbf{r}}{\partial \mathbf{m}} = g \frac{\partial \mathbf{e}}{\partial \mathbf{m}} + \frac{\partial g}{\partial \mathbf{m}} \mathbf{e} \quad (\text{A.13})$$

e

$$\frac{\partial \mathbf{r}}{\partial \mathbf{a}} = g \frac{\partial \mathbf{e}}{\partial \mathbf{a}}. \quad (\text{A.14})$$

Olhando a Equação (A.10), obtemos:

$$\frac{\partial \mathbf{e}}{\partial \mathbf{m}} = \mathbf{A} \mathbf{Y}. \quad (\text{A.15})$$

Pela Equação (A.8), obtemos:

$$\frac{\partial g}{\partial \mathbf{m}} = - \left(\frac{1}{N - P} \right) g \sum_{n=P+1}^N \frac{\mathbf{h}_n}{1 + \mathbf{h}_n \mathbf{m}}. \quad (\text{A.16})$$

Por fim, lembrando que também podemos escrever $\mathbf{e} = \mathbf{x} - \mathbf{X} \mathbf{a}$, segue que:

$$\frac{\partial \mathbf{e}}{\partial \mathbf{a}} = -\mathbf{X}, \quad (\text{A.17})$$

em que \mathbf{X} é definida na Equação (4.7).

A.3 Cálculo do gradiente e da hessiana para o modelo ICM

Para o cálculo do gradiente de $L(\mathbf{b})$ calculamos separadamente as derivadas de $L(\mathbf{b})$ em relação a cada elemento de \mathbf{b} , que podem ser obtidas pelo formulário abaixo:

$$\frac{\partial L}{\partial b_i} = \sum_{n=R+1}^N \frac{1}{s'[z(n), \mathbf{m}, \mathbf{b}]} \frac{\partial s'[z(n), \mathbf{m}, \mathbf{b}]}{\partial b_i} + \mathbf{e}^T \frac{\partial \mathbf{e}}{\partial b_i}, \quad (\text{A.18})$$

$$\frac{\partial s'[z(n), \mathbf{m}, \mathbf{b}]}{\partial b_i} = \frac{\partial g'[y(n), \mathbf{m}]}{\partial y(n)} \frac{\partial y(n)}{\partial b_i}, \quad (\text{A.19})$$

$$\frac{\partial g'[y(n), \mathbf{m}]}{\partial y(n)} = \frac{\partial}{\partial y(n)} \left(1 + \sum_{j=1}^M m_j (j+1) y^j(n) \right) = \sum_{j=1}^M m_j j (j+1) y^{j-1}(n), \quad (\text{A.20})$$

$$\frac{\partial y(n)}{\partial b_i} = -z(n-i), \quad (\text{A.21})$$

$$\frac{\partial \mathbf{e}(n)}{\partial b_i} = \frac{\partial x(n)}{\partial b_i} - \sum_{j=1}^P a_j \frac{\partial x(n-i)}{\partial b_i}, \quad (\text{A.22})$$

$$\frac{\partial x(n)}{\partial b_i} = \frac{\partial g[y(n), \mathbf{m}]}{\partial y(n)} \frac{\partial y(n)}{\partial b_i}, \quad (\text{A.23})$$

$$\frac{\partial g[y(n), \mathbf{m}]}{\partial y(n)} = 1 + \sum_{j=1}^M m_j (j+1) y^j(n). \quad (\text{A.24})$$

As derivadas de segunda ordem, necessárias para a obtenção da Hessiana, podem ser obtidas pelo formulário abaixo:

$$\begin{aligned} \frac{\partial^2 L}{\partial b_i \partial b_j} &= \sum_{n=R+1}^N \left(-\frac{\frac{\partial s'[z(n), \mathbf{m}, \mathbf{b}]}{\partial b_i}}{(s'[z(n), \mathbf{m}, \mathbf{b}])^2} \right) \\ &+ \sum_{n=R+1}^N \left(\frac{\partial^2 s'[z(n), \mathbf{m}, \mathbf{b}]}{\partial b_i \partial b_j} \frac{1}{(s'[z(n), \mathbf{m}, \mathbf{b}])} \right) \\ &+ \mathbf{e}^T \frac{\partial^2 \mathbf{e}}{\partial b_i \partial b_j} + \left(\frac{\partial \mathbf{e}^T}{\partial b_i} \right) \left(\frac{\partial \mathbf{e}}{\partial b_j} \right), \end{aligned} \quad (\text{A.25})$$

$$\begin{aligned} \frac{\partial^2 s'[z(n), \mathbf{m}, \mathbf{b}]}{\partial b_i \partial b_j} &= -z(n-i) \frac{\partial}{\partial y(n)} \left(\sum_{k=1}^M j(j+1) y^{j-1}(n) \right) \frac{\partial y(n)}{\partial b_j} \\ &= z(n-i) z(n-j) \sum_{j=1}^M (j-1) j (j+1) y^{j-2}(n), \end{aligned} \quad (\text{A.26})$$

$$\frac{\partial^2 e(n)}{\partial b_i \partial b_j} = \frac{\partial^2 x(n)}{\partial b_i \partial b_j} - \sum_{j=1}^P a_j \frac{\partial^2 x(n-i)}{\partial b_i \partial b_j}, \quad (\text{A.27})$$

$$\frac{\partial^2 x(n)}{\partial b_i \partial b_j} = z(n-i) z(n-j) \sum_{k=1}^M (j+1) j y^{j-1}(n). \quad (\text{A.28})$$

Apêndice B

Testes com Sinais Reais

Neste apêndice, descrevemos exemplos de aplicação de alguns dos algoritmos propostos em sinais reais, com distorções artificialmente introduzidas. Estes exemplos, que se encontram na página cujo endereço consta em [57], correspondem a algoritmos dos Capítulos 5 e 6. Deveremos inserir futuramente exemplos correspondentes a algoritmos do Capítulo 7. Recomendamos que se usem fones de ouvido para que se possa perceber melhor as distorções introduzidas e o efeito dos algoritmos de restauração.

Os sinais escolhidos para teste foram retirados de gravações de piano. Os trechos escolhidos, com duração em torno de 10 s, apresentam alta variação de amplitude, de tal forma que os efeitos não-lineares sejam mais relevantes. Nestes exemplos, as distorções foram inseridas artificialmente para que pudéssemos avaliar a correção dos algoritmos. Também fizemos testes com distorções obtidas em condições mais realísticas, tais como ruído não-branco presente em gravações antigas e distorções não-lineares representativas de curva de compressão de faixa dinâmica. Futuramente, pretendemos expandir esse conjunto de testes.

Para todos os exemplos escolhemos a ordem do modelo como $P = 30$ e tamanho de cada bloco como $N = 800$. Quando o ruído está presente, usamos como variância σ_v^2 o mesmo valor que foi usado para gerá-lo.

Sinal 1: Limitação de amplitude, com limiar em 30 % do valor máximo do trecho do sinal. Nesse exemplo, a distorção mais perceptível encontra-se em torno dos 5 s, no meio do sinal. O algoritmo bayesiano rodou por apenas 10 iterações, e o sinal restaurado é obtido pela média das 5 últimas. Como não há ruído presente, a

convergência é bastante rápida. Ouvindo os sinais original e restaurado, concluímos que o algoritmo de restauração torna a degradação praticamente imperceptível.

Sinal 2: Limitação de amplitude, com limiar em 30 % do valor máximo do trecho do sinal, com ruído inserido a 25 dB, antes da limitação. Ambas as distorções são bastante perceptíveis ao longo de todo o sinal. A presença do ruído torna a convergência do algoritmo bastante lenta: foram necessárias 10000 iterações, sendo que o sinal restaurado é gerado pela média das 9000 últimas. Observamos no sinal restaurado significativa redução de ambos os defeitos, sendo que o SNR no sinal recuperado é de 35 dB.

Sinal 3: Limitação de amplitude, com limiar em 30 % do valor máximo do trecho do sinal, com ruído inserido a 25 dB, após a limitação. As degradações e os resultados são perceptivamente muito similares ao caso anterior, assim como o tempo de convergência do algoritmo. A maior diferença é no tempo de processamento de cada iteração, que é maior para este exemplo devido à maior complexidade exigida pelo procedimento de estimação das regiões saturadas.

Sinal 4: Saturação gradual inserida artificialmente através de modelo baseado em série de Taylor com $m_1 = 5$ e $m_2 = 30$, sem ruído. Para estimação dos coeficientes da série de Taylor, aplicamos a solução bayesiana, a qual convergiu em poucas iterações, com erro de estimação desprezível. Ouvindo os sinais original e restaurado, não percebemos qualquer diferença.

Sinal 5: Saturação gradual com mesmos parâmetros do Sinal 4, porém com ruído inserido a 35 dB. Devido ao ruído, a convergência do algoritmo é bem mais lenta, em torno de 10000 iterações. Ouvindo os sinais original e restaurado, verificamos que o algoritmo praticamente elimina a não-linearidade e reduz sensivelmente a percepção de ruído.