



**Universidade Estadual de Campinas  
Faculdade de Engenharia Elétrica e de Computação  
Departamento de Comunicações**

# **Modelo Mel-Cepstral Generalizado para Envoltória Espectral de Fala**

**Ramiro Roque Antunes Barreira**  
**Orientador: Prof. Dr. Fábio Violaro**  
**Co-Orientador: Dr. Edmilson S. Morais**

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da UNICAMP como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.

Comissão Examinadora

Fábio Violaro, Dr. ....DECOM/FEEC/UNICAMP  
Fernando Gil Vianna Resende Jr., Dr. ....COPPE/POLI/UFRJ  
Renato da Rocha Lopes, Dr. ....DECOM/FEEC/UNICAMP

Campinas, SP  
Outubro/2010

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

B274m Barreira, Ramiro Roque Antunes  
Modelo mel-cepstral generalizado para envoltória  
espectral de fala / Ramiro Roque Antunes Barreira. --  
Campinas, SP: [s.n.], 2010.

Orientadores: Fábio Violaro, Edmilson da Silva  
Morais.

Dissertação de Mestrado - Universidade Estadual de  
Campinas, Faculdade de Engenharia Elétrica e de  
Computação.

1. Sistemas de telecomunicações. 2. Sistemas de  
processamento da fala. 3. Processamento de sinais -  
Técnicas digitais. 4. Processamento de sinais - Análise  
espectral. 5. Processos estocásticos - Modelos  
matemáticos. I. Violaro, Fábio. II. Morais, Edmilson da  
Silva. III. Universidade Estadual de Campinas.  
Faculdade de Engenharia Elétrica e de Computação. IV.  
Título.

Título em Inglês: Mel-generalized cepstral model for speech spectral envelope

Palavras-chave em Inglês: Telecommunication systems, Speech processing systems,  
Signal processing - Digital techniques, Signal processing -  
Spectral analysis, Stochastic processes - Mathematical  
models

Área de concentração: Telecomunicações e Telemática

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Fernando Gil Vianna Resende Júnior, Renato da Rocha Lopes

Data da defesa: 27/10/2010

Programa de Pós Graduação: Engenharia Elétrica

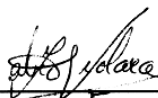
## COMISSÃO JULGADORA - TESE DE MESTRADO

**Candidato:** Ramiro Roque Antunes Barreira

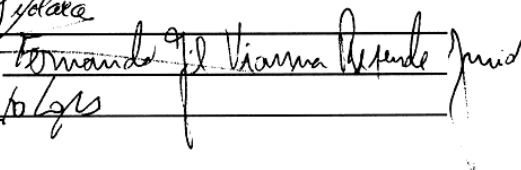
**Data da Defesa:** 27 de outubro de 2010

**Título da Tese:** "Modelo Mel-Cepstral Generalizado para Envoltória Espectral de Fala"

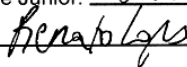
Prof. Dr. Fábio Violaro (Presidente):



Prof. Dr. Fernando Gil Vianna Resende Junior:



Prof. Dr. Renato da Rocha Lopes:



# Resumo

A análise Mel-Cepstral Generalizada (MGC) corresponde a uma abordagem para estimação de envoltória espectral de fala que unifica as análises LPC, Mel-LPC, Cepstral e Mel-Cepstral. A forma funcional do modelo MGC varia continuamente com dois parâmetros reais  $\gamma$  e  $\alpha$ , possibilitando que o modelo assuma diferentes características. A flexibilidade oferecida pelo modelo MGC aliada à sua estabilidade e bom desempenho sob manipulação de parâmetros tem feito com que os parâmetros MGC sejam empregados com sucesso em codificação de fala e síntese de fala via HMM (*Hidden Markov Models*). O presente trabalho foca os aspectos matemáticos da análise MGC, abordando e demonstrando, em extensão, a formulação em seus vieses analítico e computacional para a solução do modelo. As propriedades e formulações básicas da análise MGC são tratadas na perspectiva do espectro mel-logarítmico generalizado. Propõe-se um método para a computação dos coeficientes MGC e Mel-Cepstrais que não envolve o uso de fórmulas recursivas de transformação em frequência. As análises e experimentos relacionados ao método encontram-se em estágio inicial e devem ser completados no sentido de se identificar a relação *ganho computacional*  $\times$  *qualidade da representação*.

**Palavras-chave:** Análise Mel-Cepstral Generalizada, Envoltória Espectral de Fala, Análise e Síntese de Fala, Síntese de Fala Paramétrica Estatística, HMM.

# Abstract

Mel-Generalized Cepstral analysis (MGC) is an approach for speech spectral envelope estimation that unifies LPC, Mel-LPC, Cepstral and Mel-Cepstral Analysis. The functional form of the MGC model varies continuously with the real parameters  $\gamma$  e  $\alpha$ , enabling the model to acquire different characteristics. The flexibility of MGC model associated with its stability and good performance under parameter manipulation have made MGC parameters to be successfully employed in speech codification and HMM speech synthesis. The present study focuses on mathematical aspects of MGC analysis, treating and proving, in a fairly extended way, analytical and computational formulation for model solution. MGC analysis properties and basic formulation are treated in mel-generalized logarithmic spectrum perspective. A method for the computation of MGC and Mel-Cepstral coefficients that do not require frequency transformation recursion formulas is proposed. Experiments and analysis concerning the method are in their initial stage and needs to be completed in the sense to identify *computational*  $\times$  *representation* performances.

**Keywords:** Mel-Generalized Cepstral Analysis, Speech Spectral Envelope, Speech Analysis and Synthesis, Statistical Parametric Speech Synthesis, HMM.

# Agradecimentos

Agradeço imensamente ao meu orientador Fábio Violaro pelos ensinamentos, por toda a confiança e pelo incentivo permanente no decorrer dos trabalhos.

Agradeço à minha mãe, Maria Cristina, e ao meu pai, Amilton, por toda a força, paciência e suporte.

Pelo incentivo, ajuda e bons momentos, agradeço aos amigos e pessoas queridas: Cecílio Fráguas, Alexandre Magro, José Antônio Rezende, Guilherme Moreira, Gilberto Egydio, Matheus Urenha, Osvaldo Félix, Rebeca Andrade, Priscila Tsupal, Fabiana Gil, Adriana Catuta, Ana Carolina Roselino, Paula Skromov e tantos outros.

Agradeço aos meus irmãos Cristiano e Caetano pelo incentivo e tolerância em momentos difíceis.

Agradeço ao meu primo Mateus e à minha tia Maria Silvia pelo apoio afetivo.

Agradeço ao Edmilson Morais e à Jussara Vieira pelas diversas oportunidades proporcionadas na Vocalize (empresa de soluções em tecnologias da fala e da linguagem).

Agradeço aos professores Fernando Gil V. Resende (COPPE - UFRJ) e Keiichi Tokuda (*Nitech - Japan*) por cederem artigos importantes para a elaboração do trabalho, e ao professor Djalma M. Redondo (IFSC - USP) pelos esclarecimentos conceituais.

Agradeço à CAPES e aos professores e funcionários da FEEC pelo apoio.

*Aos meus pais*

# Sumário

<b>Lista de Figuras</b> .....	<b>xvii</b>
<b>Glossário</b> .....	<b>xix</b>
<b>Lista de Símbolos</b> .....	<b>xxi</b>
<b>Trabalhos Publicados durante o Mestrado</b> .....	<b>xxiii</b>
<b>Capítulo 1 Introdução</b> .....	<b>1</b>
<b>Capítulo 2 Cepstrum Generalizado</b> .....	<b>5</b>
2.1 Introdução .....	5
2.2 Função Logarítmica Generalizada .....	7
2.2.1 Função log-gen .....	7
2.2.2 Função exp-gen (log-gen inversa) .....	8
2.3 Espectro Logarítmico Generalizado .....	9
2.3.1 Cepstrum Generalizado .....	9
2.3.2 Magnitude Logarítmica Generalizada .....	10
2.4 Representação Espectral Através do Cepstrum Generalizado .....	13
2.4.1 Relação entre $x(0)$ e $x_\gamma(0)$ , e Representação do Espectro através do Cepstrum Generalizado Normalizado .....	14
2.5 Critério para a Unicidade de $\ln_\gamma(X(z))$ .....	16
2.5.1 Multivalência de $\ln_\gamma(X(z))$ .....	17
2.5.2 Critério para Unicidade de $\ln_\gamma(X(z))$ .....	18
2.5.3 Relação Implícita entre $x_\gamma(n)$ e $x(n)$ .....	21
2.5.4 Relação Implícita entre $x_{\gamma_2}(n)$ e $x_{\gamma_1}(n)$ .....	21
2.6 Critério de Fase Mínima para $x(n)$ .....	22
2.6.1 PSDs de $\ln_\gamma(X(z))$ e Condição de Fase Mínima .....	23
2.6.2 Cepstrum Complexo: Critério de Fase Mínima .....	25
2.7 Cálculo do Cepstrum Generalizado .....	27
2.7.1 Cálculo de $x(n)$ a partir de $x_\gamma(n)$ .....	28

2.7.2	Cálculo de $x_\gamma(n)$ a partir de $x(n)$ .....	29
2.7.3	Cálculo de $x_{\gamma_2}(n)$ a partir de $x_{\gamma_1}(n)$ .....	30
2.7.4	Transformação Logarítmica Generalizada e Suavização Espectral.....	31
2.7.5	Computação do Cepstrum Generalizado .....	32
2.8	Suavização Espectral .....	34
2.8.1	Suavização do Espectro log-gen.....	35
2.8.2	Deconvolução Homomórfica log-gen .....	36
2.8.3	Envoltória Espectral Característica e Remoção de Viés Estatístico.....	37

### Capítulo 3 Seqüências Distorcidas em Freqüência e o Mel-Cepstrum

#### Generalizado..... 41

3.1	Introdução .....	41
3.2	Espectro Distorcido em Freqüência .....	42
3.2.1	Seqüências Distorcidas em Freqüência .....	42
3.3	Representação do Espectro através da Seqüência Distorcida em Freqüência.....	45
3.4	Cálculo da Seqüência Distorcida em Freqüência.....	47
3.4.1	Cálculo de $\tilde{x}(n)$ a partir de $x(n)$ .....	47
3.4.2	Cálculo de $\tilde{x}_{\alpha_2}(n)$ a partir de $\tilde{x}_{\alpha_1}(n)$ .....	51
3.4.3	Transformação em Freqüência e Suavização Espectral.....	52
3.4.4	Análise Espectral com Resolução Não-Uniforme .....	54
3.4.5	DFT e Amostragem do Espectro Distorcido em Freqüência.....	61
3.5	Espectro Mel-Logarítmico Generalizado.....	64
3.5.1	Mel-Cepstrum Generalizado .....	64
3.5.2	Magnitude Mel-Logarítmica Generalizada .....	65
3.6	Representação Espectral Através do Mel-Cepstrum Generalizado .....	68
3.7	Cálculo do Mel-Cepstrum Generalizado e Suavização Espectral.....	69
3.7.1	Crítério de fase mínima com base no Mel-Cepstrum .....	69
3.7.2	Cálculo de $\tilde{x}_{\gamma_2, \alpha_2}(n)$ a partir de $\tilde{x}_{\gamma_1, \alpha_1}(n)$ .....	70
3.7.3	Suavização Espectral .....	71

#### Capítulo 4 Modelo Mel-Cepstral Generalizado..... 75

4.1	Introdução .....	75
4.2	Preliminares: Análise Cepstral Generalizada.....	77
4.2.1	Modelo.....	77
4.2.2	Normalização.....	77
4.2.3	Caso Cepstral ( $\gamma = 0$ ) .....	79



4.2.4	Caso LPC ( $\gamma = -1$ ) .....	79
4.3	Análise Mel-Cepstral Generalizada .....	80
4.3.1	Modelo.....	80
4.3.2	Modelo Baseado nos Coeficientes MGC Transformados $b_\gamma(n)$ .....	81
4.3.3	Normalização.....	83
4.3.4	Caso Mel-Cepstral ( $\gamma = 0$ ).....	85
4.3.5	Caso Mel-LPC ( $\gamma = -1$ ).....	85
4.4	Critério Espectral .....	86
4.4.1	Unicidade e Estabilidade da Solução.....	87
4.4.2	Computação do Periodograma .....	88
4.5	Método de Newton-Raphson para a Solução do Modelo .....	89
4.5.1	Método de Newton-Raphson .....	89
4.5.2	Caso Mel-Cepstral Generalizado ( $\alpha \neq 0, \gamma \neq 0$ ) .....	90
4.5.3	Caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ) .....	92
4.5.4	Caso Cepstral Generalizado ( $\alpha = 0, \gamma \neq 0$ ).....	93
4.5.5	Caso Cepstral ( $\alpha = 0, \gamma = 0$ ).....	96
4.6	Computação dos Coeficientes Mel-Cepstrais Generalizados ( $\alpha \neq 0$ ) por meio de FFTs e Fórmulas Recursivas.....	98
4.6.1	Caso Mel-Cepstral Generalizado ( $\alpha \neq 0, \gamma \neq 0$ ) .....	99
4.6.2	Caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ) .....	103
4.7	Características do Modelo e Emprego dos Parâmetros MGC em Síntese de Fala Via HMM 106	
4.7.1	Perfis do Modelo e o Espectro de Fala.....	106
4.7.2	Ressíntese do Sinal de Fala .....	113
4.7.3	Emprego dos Parâmetros MGC em Síntese de Fala via HMM.....	113
4.8	Método para a Computação dos Coeficientes Mel-Cepstrais Generalizados ( $\alpha \neq 0$ ) por meio de FFTs.....	117
4.8.1	Caso Mel-Cepstral Generalizado ( $\alpha \neq 0, \gamma \neq 0$ ) .....	118
4.8.2	Caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ) .....	120
4.8.3	Método para a Computação do Periodograma $\tilde{I}_N(\tilde{\omega})$ .....	122
4.9	Conclusão .....	128
<b>Apêndice A Função log-gen e sua Inversa .....</b>		<b>131</b>
A.1	Função log-gen ( $\ln_\gamma(\cdot)$ ) .....	131
A.1.1	Relações para a Função log-gen .....	132
A.2	Função exp-gen ( $\exp_\gamma(\cdot)$ ).....	133

A.2.1	Relações para a Função exp-gen .....	133
<b>Apêndice B Relações Úteis para <math>\Phi_n(e^{j\omega})</math> .....</b>		<b>135</b>
B.1	Relações entre $\Phi_k(e^{j\omega})$ e $\Psi(e^{j\omega})$ .....	135
B.2	Relações para os produtos $\Phi_m(e^{j\omega})\Phi_k^*(e^{j\omega})$ e $\Phi_m(e^{j\omega})\Phi_l(e^{j\omega})$ .....	137
B.3	Relações envolvendo $\Lambda(e^{j\omega})$ .....	138
<b>Apêndice C Modelo MGC baseado nos Coeficientes <math>b_\gamma(n)</math> e</b>		
<b>Transformações <math>\tilde{c}_\gamma(n) \rightleftharpoons b_\gamma(n)</math> e <math>c_\gamma(n) \rightleftharpoons b_\gamma(n)</math> .....</b>		<b>141</b>
C.1	Representação do Modelo MGC com base nos Coeficientes $b_\gamma(n)$ .....	141
C.2	Transformação $\tilde{c}_\gamma(n) \rightleftharpoons b_\gamma(n)$ .....	145
C.3	Transformação $c_\gamma(n) \rightleftharpoons b_\gamma(n)$ .....	147
<b>Apêndice D Gradiente e Hessiana de <math>\varepsilon</math> .....</b>		<b>151</b>
D.1	Análise Mel-Cepstral Generalizada ( $\alpha \neq 0, \gamma \neq 0$ ) .....	151
D.1.1	Cálculo do Vetor Gradiente $\nabla_{\mathbf{b}, \varepsilon}$ .....	152
D.1.2	Cálculo da Matriz Hessiana $\mathbf{H}$ .....	153
D.2	Análise Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ) .....	156
D.2.1	Vetor Gradiente $\nabla_{\mathbf{b}, \varepsilon}$ .....	156
D.2.2	Matriz Hessiana $\mathbf{H}$ .....	157
D.3	Análise Cepstral Generalizada ( $\alpha = 0, \gamma \neq 0$ ) .....	158
D.3.1	Vetor Gradiente $\nabla_{\mathbf{b}, \varepsilon}$ .....	159
D.3.2	Matriz Hessiana $\mathbf{H}$ .....	159
D.4	Análise Cepstral ( $\alpha = 0, \gamma = 0$ ) .....	160
D.4.1	Vetor Gradiente $\nabla_{\mathbf{b}, \varepsilon}$ .....	160
D.4.2	Matriz Hessiana $\mathbf{H}$ .....	161
<b>Apêndice E Computação do Gradiente e da Hessiana de <math>\varepsilon</math> via Fórmulas</b>		
<b>Recursivas .....</b>		<b>163</b>
E.1	Análise Mel-Cepstral Generalizada .....	164
E.1.1	Cálculo de $\tilde{r}_\gamma(k)$ .....	166
E.1.2	Cálculo de $\tilde{p}_\gamma(k)$ .....	167
E.1.3	Cálculo de $\tilde{q}_\gamma(k)$ .....	171
E.1.4	Cálculo de $\varepsilon$ .....	171

E.2 Análise Mel-Cepstral ( $\gamma = 0$ )..... 172

**Apêndice F Expressões para a Computação do Gradiente e da Hessiana de  $\varepsilon$  via FFTs..... 177**

F.1 Análise Mel-Cepstral Generalizada ..... 179

F.2 Análise Mel-Cepstral ..... 181

**Referências Bibliográficas ..... 183**

# Lista de Figuras

Figura 1.1: Análise Mel-Cepstral Generalizada [1]. .....	2
Figura 2.1: Curvas da função log-gen de uma variável real $x$ , com $\gamma = 1$ , $\gamma = 1/3$ , $\gamma = 0$ , $\gamma = -1/3$ e $\gamma = -1$ . .....	7
Figura 2.2: Curvas da função exp-gen de uma variável real $x$ , com $\gamma = 1$ , $\gamma = 1/3$ , $\gamma = 0$ , $\gamma = -1/3$ e $\gamma = -1$ . .....	8
Figura 2.3: (a) Magnitude espectral (dB) de um sinal de fala (voz masculina pronunciando a vogal “i”). (b) Magnitude Espectral (escala linear). (c) Curvas da função log-gen de uma variável real com $\gamma = 1/3$ , $\gamma = 0$ , $\gamma = -1/3$ . (d) Magnitude espectral log-gen com $\gamma = 1/3$ . (e) Magnitude espectral log-gen com $\gamma = 0$ . (f) Magnitude espectral log-gen com $\gamma = -1/3$ . <i>Observação: a magnitude log-gen é expressa por <math>\ln_{\gamma}( \cdot ^2)</math>.</i> .....	12
Figura 2.4: (a) (c) (e) Espectros log-gen suavizados (ordem 30) com base nos espectros log-gen associados a $\gamma = 1/3$ , $\gamma = 0$ e $\gamma = -1/3$ (em verde). (b) (d) (f) Magnitudes (dB) das envoltórias espectrais características associadas a $\gamma = 1/3$ , $\gamma = 0$ e $\gamma = -1/3$ sobrepostas ao espectro original (em verde). <i>Observação: a magnitude log-gen é expressa por <math>\ln_{\gamma}( \cdot ^2)</math>.</i> .....	39
Figura 3.1: Curvas de mapeamento do eixo de frequências .....	44
Figura 3.2: (a) Magnitude espectral (dB) de um sinal de fala. (b) Curvas de mapeamento do eixo de frequências com $\alpha = 0.5$ e $\alpha = -0.5$ . (c) Magnitude do espectro distorcido em frequência com parâmetro $\alpha = 0.5$ . (d) Magnitude do espectro distorcido em frequência com parâmetro $\alpha = -0.5$ . .....	45
Figura 3.3: Curvas de mapeamento do eixo de frequências .....	47
Figura 3.4: Banco de filtros em cascata com entrada $x(-i)$ e saída $\tilde{x}(n, i)$ em sua $(n+1)$ -ésima seção. Os valores $\tilde{x}(n, 0)$ das seqüências de saída constituem a seqüência $\tilde{x}(n)$ . .....	49
Figura 3.5: (a): Curvas de mapeamento de frequências $\tilde{\omega}(\omega)$ para $\alpha = 0.5$ , $\alpha = 0$ e $\alpha = -0.5$ . (b): Respectivas curvas de $d\tilde{\omega}(\omega)/d\omega$ . .....	59
Figura 3.6: Exemplo de obtenção de espectro com resolução não-uniforme (com $\alpha = 0.5$ ). (a): Magnitude (dB) do espectro $X(e^{j\omega})$ de uma seqüência $x(n)$ de comprimento 96. (b): Magnitude (dB) do espectro $\tilde{X}(e^{j\tilde{\omega}})$ da seqüência transformada $\tilde{x}(n)$ de comprimento infinito. (c): Magnitude (dB) do espectro $\tilde{X}_M(e^{j\tilde{\omega}})$ da seqüência $\tilde{x}_M(n)$ de comprimento 32 (versão truncada de $\tilde{x}(n)$ ). (d): Magnitude (dB) de $X_{nu}(e^{j\omega}) = \tilde{X}_M(e^{j\tilde{\omega}(\omega)})$ , uma versão suavizada com resolução espectral não-uniforme de $X(e^{j\omega})$ . (e) Magnitude (dB) do espectro $X_M(e^{j\omega})$ da seqüência $x(n)$ truncada com comprimento 32: resolução uniforme. ....	60
Figura 3.7: Curva de mapeamento de frequências $\tilde{\omega}(\omega)$ para $\alpha = 0.42$ e curva da escala mel de frequência perceptual (linha pontilhada). Frequência de amostragem de 16kHz. (figura extraída de [22]). .....	61

Figura 3.8: (a) Magnitude espectral (dB). (b) Magnitude Espectral (escala linear). (c) Curva da função log-gen de uma variável real com  $\gamma = -0.2$ . (d) Magnitude espectral log-gen com  $\gamma = -0.2$ . (e) Curvas de mapeamento do eixo de freqüências com parâmetros  $\alpha = 0.5$  e  $\alpha = -0.5$ . (g) Magnitude espectral mel-log-gen com  $\alpha = 0.5$  e  $\gamma = -0.2$ . (g) Magnitude espectral mel-log-gen com  $\alpha = -0.5$  e  $\gamma = -0.2$ . *Observação: a magnitude log-gen é expressa por  $\ln_{\gamma}(|\cdot|^2)$ .* ..... 67

Figura 3.9: (a) (c) Espectros mel-log-gen suavizados (ordem 30) com base nos espectros mel-log-gen associados a  $(\alpha, \gamma) = (0.5, -0.2)$  e  $(\alpha, \gamma) = (-0.5, -0.2)$ . ..... 73

Figura 4.1: Filtro inverso:  $D(z)$  é determinado minimizando-se  $E[e^2(n)]$  ..... 87

Figura 4.2: Curvas para  $\gamma = -1$ ,  $\gamma = -1/2$  e  $\gamma = 0$  da função  $10 \log_{10}(\exp_{\gamma}(x))$ . Para  $\gamma = 0$  a função se torna linear na forma  $10 \log_{10}(e) \cdot x$ . ..... 108

Figura 4.3: Curvas para  $\gamma = -1$ ,  $\gamma = -1/2$  e  $\gamma = 0$  da função  $10 \log_{10}(\exp_{\gamma}(\tilde{f}_{\gamma}(\tilde{\omega}(\omega))))$  utilizando o harmônico  $k = 5$  de  $\tilde{f}_{\gamma}(\tilde{\omega})$ . (a) com  $\alpha = 0$  ( $\tilde{\omega}(\omega) = \omega$ ). (b) com  $\alpha = 0.42$ . .... 108

Figura 4.4: Envoltórias espectrais obtidas por análises MGC de ordem 20 aplicadas à emissão da vogal “ó” amostrada a 16kHz (espectro em verde). (a) Com  $(\alpha, \gamma) = (0, 0)$ , o modelo não faz distinção entre o par de formantes  $f_1 \sim 500\text{Hz}$  e  $f_2 \sim 900\text{Hz}$ , o mesmo vale para o par  $f_3 \sim 3000\text{Hz}$  e  $f_4 \sim 3500\text{Hz}$ . (b) Com  $(\alpha, \gamma) = (0, -1/3)$ , o modelo passa a distinguir  $f_3$  e  $f_4$ . (c) Com  $(\alpha, \gamma) = (0, -1/2)$ , adicionalmente, o modelo passa a distinguir entre  $f_1$  e  $f_2$ . (d) Com  $(\alpha, \gamma) = (0, -1)$ , bandas dos formantes seguem se estreitando e o vale em  $\sim 4300\text{Hz}$  tem sua descrição mais pobre. (e) Com  $(\alpha, \gamma) = (0.42, -1/3)$ , o modelo descreve freqüências baixas com maior resolução e passa a distinguir os formantes  $f_1$  e  $f_2$ , antes ambíguos no caso  $(\alpha, \gamma) = (0, -1/3)$ . ..... 111

Figura 4.5: Envoltórias espectrais obtidas por análises MGC de ordem 32 aplicadas à emissão da soante nasal “m” amostrada a 16kHz (espectro em verde). (a) Com  $(\alpha, \gamma) = (0, -1)$  o modelo não descreve os vales associados aos anti-formantes  $af_1 \sim 600\text{Hz}$  e  $af_2 \sim 3000\text{Hz}$  (b)  $(\alpha, \gamma) = (0, -1/3)$ : o aumento em  $\gamma$  ainda não contempla tais vales. (c) Com  $(\alpha, \gamma) = (0, 0)$ , os vales em  $af_1$  e  $af_2$  são contemplados pelo modelo. (d) Com  $(\alpha, \gamma) = (0.42, -1/3)$ , o vale em  $af_1$ , não contemplado no caso  $(\alpha, \gamma) = (0, -1/3)$ , passa a ser bem descrito em função da resolução superior na região de baixas freqüências. .... 112

Figura 4.6: envoltórias obtidas por análise mel-cepstral ( $\alpha = 0$ ) aplicada diretamente a um espectro de fala (FFT) (ordem 24 e 39) e ao espectro pré-suavizado STRAIGHT (ordem 39) [34]. ..... 115

Figura 4.7: Envoltórias espectrais obtidas por análises MGC de ordem 32 aplicadas à emissão da vogal “i” (curva espectral em verde). Curvas em azul: calculadas pelo método *FFT/recursão* [2][7]. Curvas em vermelho: calculadas pelo método *somente FFT*. (a) (c) (e): Análises efetuadas respectivamente com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ , a 16kHz, com  $\alpha = 0.42$ ,  $N = 384$  e  $N_1 = N_2 = 512$ . (b) (d) (f) Análises efetuadas respectivamente com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ , a 22.050kHz com  $\alpha = 0.45$  e  $N = N_1 = N_2 = 512$ . ..... 127

# Glossário

FFT – Transformada Rápida de Fourier (*Fast Fourier Transform*)

DFT – Transformada Discreta de Fourier (*Discrete Fourier Transform*)

CRU – Circunferência de Raio Unitário

RC – Região de Convergência

PSD – Ponto Singular Divergente

MGC – Mel-Cepstral Generalizado (*Mel-Generalized Cepstral*)

GC – Cepstral Generalizado (*Generalized Cepstral*)

LPC – Linear Predictive Coding

LSP – Line Spectrum Pair

HMM – Modelos Ocultos de Markov (*Hidden Markov Models*)

UELS – Unbiased Estimation of Log-Spectrum

F0 – Frequência Fundamental

cepstrum-gen – cepstrum generalizado

mel-cepstrum-gen – mel-cepstrum generalizado

log-gen – logaritmo generalizado

exp-gen – exponencial generalizada

# Lista de Símbolos

- $\ln_\gamma(\cdot)$  - função logarítmica generalizada
- $\exp_\gamma(\cdot)$  - função exponencial generalizada
- $\omega$  - eixo regular de frequências
- $\omega(\cdot)$  - eixo regular de frequências (variável dependente)
- $\tilde{\omega}$  - eixo distorcido de frequências
- $\tilde{\omega}(\cdot)$  - eixo distorcido de frequências (variável dependente)
- $z$  - plano complexo regular
- $z(\cdot)$  - plano complexo regular (variável dependente)
- $\tilde{z}$  - plano complexo transformado
- $\tilde{z}(\cdot)$  - plano complexo transformado (variável dependente)
- $\alpha$  - parâmetro de distorção em frequência
- $\gamma$  - parâmetro de distorção em amplitude
- $H(z)$  - modelo Mel-Cepstral Generalizado
- $D(z)$  - modelo Mel-Cepstral Generalizado normalizado em ganho
- $\varepsilon$  - critério a ser minimizado para a solução do modelo MGC
- $I_N(\omega)$  - periodograma modificado
- $c_\gamma(n)$  - seqüência de coeficientes cepstrais generalizados (GC)
- $\tilde{c}_\gamma(n)$  - seqüência de coeficientes mel-cepstrais generalizados (MGC)
- $b_\gamma(n)$  - seqüência de coeficientes mel-cepstrais generalizados (MGC) transformados

## Trabalhos Publicados durante o Mestrado

Ramiro R. A. Barreira and Fábio Violaro, "Uma Avaliação da Análise Cepstral Generalizada Aplicada ao Espectro de Vogais," in *Anais do XXVII Simpósio Brasileiro de Telecomunicações, 2009*, Blumenau, 2009.

Ramiro R. A. Barreira and Fábio Violaro, "Avaliação da Análise Cepstral Generalizada Aplicada à Modelagem de Vogais," in *Anais do 7o Congresso da AES Brasil 2009*, São Paulo, 2009, pp. 95-101.



# Capítulo 1

## Introdução

A análise Mel-Cepstral Generalizada [1][2] (MGC – *Mel-Generalized Cepstral*) é uma abordagem unificada para estimação de envoltória espectral de fala, que tem as análises LPC [3] (predição linear), Cepstral [4][5] (UELS – *Unbiased Estimation of Log-Spectrum*), Mel-Cepstral [6][7] e Mel-LPC [8] como casos particulares. Tais análises, modelos de uso clássico em processamento de fala, são obtidas, dentre uma diversidade de modelos, pela atribuição de diferentes combinações de valores a dois parâmetros:  $\alpha$  e  $\gamma$ . Estes são respectivamente responsáveis por regularem alterações nas escalas de frequência e amplitude espectral, proporcionando a diversidade de perfis do modelo MGC. O parâmetro  $\alpha$ , que assume valores no intervalo real  $]-1,1[$ , permite modificar a maneira como a resolução da análise é distribuída ao longo do eixo de frequências do espectro. Já o parâmetro  $\gamma$ , que assume valores no intervalo real  $[-1,1]$ , proporciona, pelo modelo, diferentes pesos à descrição de picos e vales espectrais. Para  $(\alpha, \gamma) = (0, -1)$ , a análise MGC se reduz à análise LPC, cujo modelo (de pólos - *all-pole*) descreve bem picos espectrais e é, porém, pouco sensível a vales. Já para  $(\alpha, \gamma) = (0, 0)$ , obtém-se a análise Cepstral (UELS), que modela picos e vales espectrais com pesos equivalentes. Com  $(\alpha, \gamma) = (0, 1)$  seria obtido um modelo de zeros (*all-zero*) caracterizado por favorecer a descrição de vales em detrimento de picos. Assim, variando-se o parâmetro  $\gamma$  de  $-1$  a  $1$  as características do modelo variam continuamente entre aquelas de um modelo *all-pole* e de um modelo *all-zero*, passando pelo modelo exponencial (Cepstral), que pode ser considerado um modelo de pólos e zeros (*pole-*

zero model). O modelo restrito a  $\alpha = 0$ , podendo assumir qualquer valor no intervalo  $\gamma = [-1,1]$ <sup>§</sup>, é denominado de modelo Cepstral Generalizado (GC) [9]. Com  $|\alpha| < 1$ , são obtidas as versões distorcidas em frequência dos casos mencionados. Destacam-se a análise Mel-LPC ( $|\alpha| < 1, \gamma = -1$ ) e a análise Mel-Cepstral ( $|\alpha| < 1, \gamma = 0$ ). A Figura 1.1 mostra como casos especiais estão situados, em termos dos valores de  $\alpha$  e  $\gamma$ , no contexto da análise MGC.

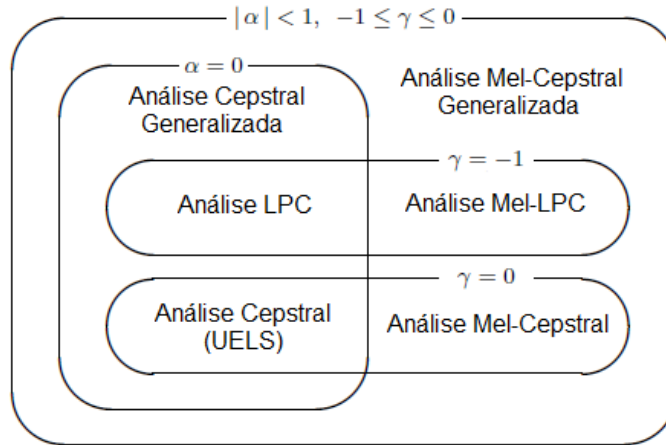


Figura 1.1: Análise Mel-Cepstral Generalizada [1].

O modelo MGC de ordem  $M$ , é descrito por  $M + 1$  coeficientes MGC. A gama de características propiciadas pela escolha dos parâmetros  $\alpha$  e  $\gamma$  permite que, com número fixo reduzido de coeficientes, possa ser escolhido um par  $(\alpha, \gamma)$  para uma representação eficiente com base em aspectos físicos descritos pelo espectro de fala e aspectos perceptuais da audição. Os últimos se referem à não-uniformidade da resolução auditiva em frequência e são contemplados uma vez que o parâmetro  $\alpha$  pode ser ajustado de maneira a aproximar a escala mel (ou bark) de frequência perceptual. Isso traz eficiência no uso do modelo em síntese de fala, por minimizar descrições redundantes para a percepção. No que se refere aos aspectos físicos, fixada a ordem do modelo, a escolha do parâmetro  $\gamma$  pesa uma relação compromissada entre as descrições de ressonâncias (formantes) e anti-ressonâncias (anti-formantes) associadas às dimensões do trato vocal. Com  $\gamma = -1$  (casos LPC e Mel-LPC), o modelo traz uma boa descrição de formantes, às custas de sub-modelar eventuais anti-formantes, característicos de emissões nasais. Com  $\gamma = 0$  (casos Cepstral e Mel-Cepstral), formantes e anti-formantes ganham descrições equivalentes, de maneira que as nasais

<sup>§</sup> Apesar de  $\gamma$  assumir valores no intervalo mencionado, o modelo MGC tem solução única somente para o intervalo  $\gamma = [-1,0]$ .

são contempladas pelo modelo com custo de uma abertura na banda dos formantes. A escolha de valores em  $-1 < \gamma < 0$  propicia descrições intermediárias às mencionadas. A descrição feita assume que o modelo tem ordem reduzida, de modo que sua envoltória se limite à descrição de características de ressonância do trato vocal (com o acréscimo perceptual), sem o registro de interferências da periodicidade do sinal de fala.

O modelo MGC, sob a forma de um filtro racional, pode ser usado para a ressíntese do sinal de fala (com base no modelo fonte-filtro). Porém, sob quantização, interpolação, ou quando gerados por estatísticas, os coeficientes MGC podem levar o modelo à instabilidade (nos casos  $-1 \leq \gamma < 0$ ), da qual se previne somente no caso  $\gamma = 0$  (coeficientes Mel-Cepstrais ou Cepstrais). Isso pode ser solucionado pela transformação dos coeficientes em parâmetros do tipo LSP, para os quais um critério simples garante estabilidade ao modelo. Adicionalmente os parâmetros MGC-LSP mostram bom desempenho sob quantização e interpolação [10].

A flexibilidade oferecida pelo modelo MGC mediante o ajuste dos parâmetros  $\alpha$  e  $\gamma$ , e a garantia de estabilidade e bom desempenho sob manipulações, têm feito dos MGCs, parâmetros de uso recorrente em diversas aplicações em fala, dentre elas: codificação [11] [12][13] e síntese de fala via HMM [14].

O presente trabalho foca os aspectos matemáticos da análise MGC, abordando e demonstrando, em extensão, a formulação em seus vieses analítico e computacional para a solução do modelo. Os casos Cepstral, Cepstral Generalizado (GC) e Mel-Cepstral, apesar de terem as respectivas formulações derivadas de uma formulação geral, por simples atribuições de valores aos parâmetros  $\alpha$  e  $\gamma$ , são individualmente abordados, já que exigem métodos computacionais dedicados. No intervalo  $\gamma = [-1,0]$ , ao qual nos restringiremos, a solução é obtida com base na minimização de um critério e computada por um processo iterativo de convergência quadrática. Em especial, os respectivos métodos para os cálculos dos coeficientes MGC ( $\alpha \neq 0, -1 \leq \gamma < 0$ ) e Mel-Cepstrais ( $\alpha \neq 0, \gamma = 0$ ) envolvem, a cada iteração, a computação de FFTs e fórmulas recursivas de transformação em frequência [2][7]. Tais métodos são, portanto, mais custosos do que aqueles para as soluções dos casos GC e Cepstral ( $\alpha = 0$ ), que exigem somente a computação de FFTs, a cada iteração, sem o uso de fórmulas recursivas.

Neste trabalho propõe-se um método para as respectivas computações dos coeficientes MGC e Mel-Cepstrais que, a exemplo dos casos GC e Cepstral, exige, a

cada iteração do processo, somente a computação de FFTs, sem necessidade do uso de fórmulas recursivas. O método, através do qual uma solução sem degradações significativas seria, a priori, esperada sob a exigência de comprimentos elevados para as FFTs, levou o modelo a curvas próximas àquelas obtidas pelo método de [2][7], utilizando FFTs de mesmo comprimento.

Como fundamento para a análise MGC, os capítulos 2 e 3 abordam o cepstrum generalizado e o mel-cepstrum generalizado. O mel-cepstrum generalizado é a seqüência que define os coeficientes MGC e cujo espectro é aquele tomado mutuamente em escalas não-lineares de amplitude e freqüência. Sua escala de amplitudes é definida pela função logarítmica generalizada (log-gen), cuja curva assume diferentes perfis com base em um parâmetro  $\gamma$  (ver seção 2.2.1). Já sua escala de freqüências é obtida por uma transformação bilinear complexa de primeira ordem e varia com base em um parâmetro  $\alpha$ . São abordadas as propriedades do cepstrum generalizado e do mel-cepstrum generalizado, bem como são deduzidas fórmulas recursivas de transformação em amplitude e freqüência, às quais se recorre para a solução do modelo MGC. A abordagem permite compreender características do modelo MGC pelo viés da suavização do espectro mel-logarítmico generalizado, tratando sua solução como um ajuste de um número reduzido de componentes do mel-cepstrum generalizado à versão mel-logarítmica generalizada do espectro em análise.

O presente texto corresponde ao primeiro de quatro capítulos da dissertação. O capítulo que se segue aborda o cepstrum generalizado e suas propriedades, assim como a transformação logarítmica generalizada e a suavização do espectro logarítmico generalizado. O terceiro capítulo aborda, de início, a distorção de seqüências em freqüência, e a transformação em freqüência e suas propriedades; aborda-se o uso mútuo das transformações log-gen e frequencial para o cálculo do mel-cepstrum generalizado, e a suavização do espectro mel-logarítmico generalizado. O quarto capítulo aborda o Modelo MGC, o critério de minimização para sua solução, o método de Newton Raphson para o seu cálculo, e as formulações analíticas e computacionais requeridas aos respectivos casos particulares do modelo; abordam-se também as características do modelo ante o espectro de fala, o uso dos parâmetros MGC em síntese de fala via HMM e, por fim, um método alternativo para a solução do modelo MGC (casos  $\alpha \neq 0$ ) é proposto.

# Capítulo 2

## Cepstrum Generalizado

### 2.1 Introdução

A representação de uma seqüência através de uma segunda seqüência, cujo espectro corresponde ao espectro da primeira em uma escala não-linear de amplitudes ou de freqüências, traz contextos diferenciados para análise e manipulação espectral. Quando o espectro de uma seqüência conhecida é adotado na escala logarítmica generalizada de amplitudes, ele corresponde ao espectro de uma seqüência denominada cepstrum generalizado. A função logarítmica generalizada, dada por (2.1), depende de um parâmetro  $\gamma$  que pode assumir valores no intervalo contínuo  $[-1,1]$  produzindo curvas diferenciadas (ver exemplos na Figura 2.1). Para  $\gamma = 0$ , a função logarítmica generalizada se reduz à função logarítmica, de maneira que o espectro logarítmico generalizado se reduz ao espectro logarítmico e, portanto, o cepstrum generalizado equivale ao cepstrum.

O cepstrum generalizado [15] (cepstrum-gen) é abordado no presente capítulo, uma vez que representa a base para as análises cepstral generalizada e mel-cepstral generalizada. No que se refere à obtenção de uma envoltória espectral de fala, a técnica de suavização linear do espectro logarítmico generalizado (log-gen) pode ser compreendida com base no SRDS (*Spectral Root Deconvolution System*) [16] e no LHDS (*Logarithmic Homomorphic Deconvolution System*) [17], sistemas de deconvolução que poderiam ser unificados teoricamente em um sistema que

denominaríamos de GLHDS (*Generalized Logarithmic Homomorphic Deconvolution System*). A suavização linear conta com a emergência de uma estrutura de bandas no domínio do cepstrum generalizado, de maneira que o conteúdo da envoltória do espectro possa ser selecionado sob filtragem (janelamento do cepstrum-gen). A envoltória espectral, obtida conduzindo-se o espectro log-gen suavizado à escala linear de amplitudes, é característica do valor de  $\gamma$  e fica sujeita a distorções em função da transição do corte estabelecido pela filtragem e em função da seleção de informações indesejáveis, já que, freqüentemente, estas não se encontram bem separadas no cepstrum-gen. Uma solução para tais problemas foi derivada de uma técnica denominada UELS (*Unbiased Estimation of Log-Spectrum*), descrita em [4] e [18], que conduziu às análises cepstral [5] e mel-cepstral [7], e se estendeu às análises cepstral generalizada [9] e mel-cepstral generalizada [1][2].

Uma vez que função log-gen é multivalente, são tratadas as condições para a unicidade do espectro log-gen, satisfeitas em relações implícitas entre seqüência original e cepstrum-gen e entre cepstrums com parâmetros  $\gamma$  distintos. A aplicação do critério de fase mínima para a seqüência original [17], aqui revisitada no contexto do cepstrum-gen, leva a cepstrums-gen causais. Isso permite que, a partir de suas relações implícitas, sejam derivadas fórmulas recursivas para o seu cálculo. Tal cálculo corresponde a uma transformação logarítmica generalizada e tem as propriedades abordadas ante o objetivo de suavização do espectro. Os problemas da suavização linear do espectro são discutidos na última seção do capítulo. A adoção de um método de ajuste de  $M$  amostras do cepstrum-gen ao espectro log-gen, se apresenta como solução para os problemas da suavização. No contexto do espectro regular, isso corresponde ao cálculo dos coeficientes do modelo Cepstral Generalizado (GC) de ordem  $M$ , onde as  $M$  amostras do cepstrum-gen entram como valores iniciais para o seu cálculo, efetuado por um método iterativo.

O termo *cepstrum* (caso  $\gamma = 0$  do cepstrum-gen) será adotado no texto, tanto para designar ambos o cepstrum real e o cepstrum complexo. Tais casos serão especificamente designados, quando necessário.

## 2.2 Função Logarítmica Generalizada

A função logarítmica generalizada e sua inversa, a exponencial generalizada, são definidas a seguir.

### 2.2.1 Função log-gen

A função logarítmica generalizada (log-gen) é dada por

$$\ln_{\gamma}(x) = \begin{cases} \frac{x^{\gamma} - 1}{\gamma}, & 0 < |\gamma| \leq 1 \\ \ln(x), & \gamma = 0 \end{cases} \quad (2.1)$$

onde  $\ln_0(x) = \ln(x)$ , já que

$$\lim_{\gamma \rightarrow 0} \left( \frac{x^{\gamma} - 1}{\gamma} \right) = \ln(x) \quad (2.2)$$

A Figura 2.1 mostra as curvas da função log-gen para diversos valores de  $\gamma$

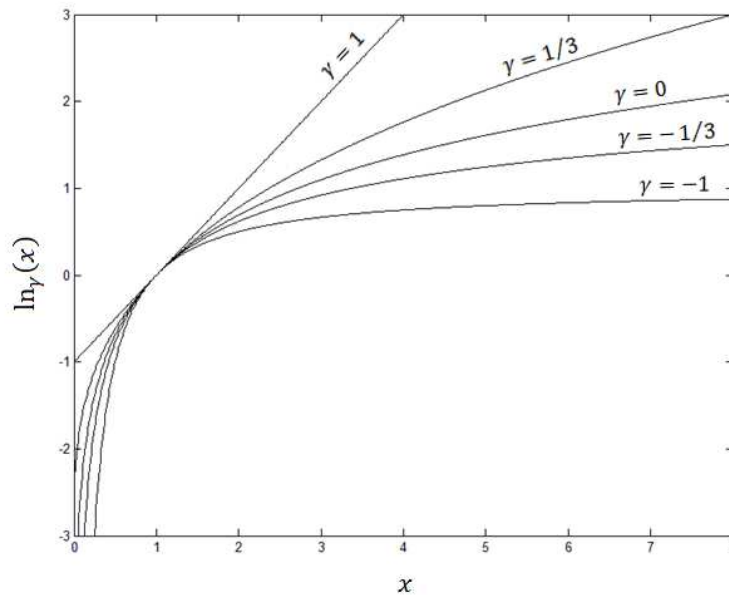


Figura 2.1: Curvas da função log-gen de uma variável real  $x$ , com  $\gamma = 1$ ,  $\gamma = 1/3$ ,  $\gamma = 0$ ,  $\gamma = -1/3$  e  $\gamma = -1$ .

Algumas das propriedades da função log-gen são dadas no Apêndice A.

## 2.2.2 Função exp-gen (log-gen inversa)

A inversa da função log-gen, a função exponencial generalizada (exp-gen), é dada por

$$\exp_{\gamma}(x) = \begin{cases} (1 + \gamma \cdot x)^{1/\gamma}, & 0 < |\gamma| \leq 1 \\ \exp(x), & \gamma = 0 \end{cases} \quad (2.3)$$

onde  $\exp_0(x) = e^x$ , já que

$$\lim_{\gamma \rightarrow 0} (1 + \gamma \cdot x)^{1/\gamma} = e^x \quad (2.4)$$

A Figura 2.2 mostra as curvas da função exp-gen para diversos valores de  $\gamma$

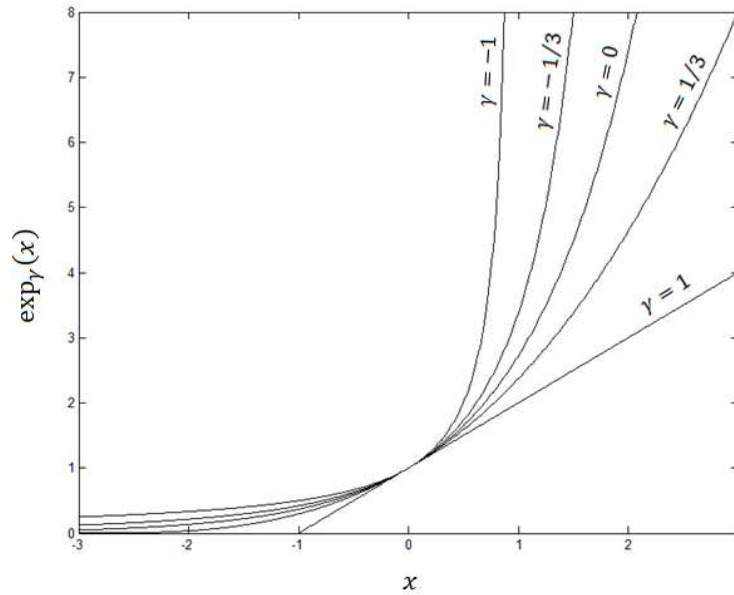


Figura 2.2: Curvas da função exp-gen de uma variável real  $x$ , com  $\gamma = 1$ ,  $\gamma = 1/3$ ,  $\gamma = 0$ ,  $\gamma = -1/3$  e  $\gamma = -1$ .

Algumas das propriedades da função exp-gen são dadas no Apêndice A.



## 2.3 Espectro Logarítmico Generalizado

### 2.3.1 Cepstrum Generalizado

Seja  $x(n)$  uma seqüência real e estável conhecida e,  $x_\gamma(n)$ , uma seqüência real e estável relacionada a  $x(n)$ , com base no que será exposto a seguir.

Considere as transformadas  $z$  de  $x(n)$  e  $x_\gamma(n)$ , dadas respectivamente por

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (2.5)$$

e

$$X_\gamma(z) = \sum_{n=-\infty}^{\infty} x_\gamma(n)z^{-n} \quad (2.6)$$

onde  $X_\gamma(z)$  está relacionada a  $X(z)$  por

$$X_\gamma(z) = \ln_\gamma(X(z)) \quad (2.7)$$

ou seja,  $X_\gamma(z)$  representa  $X(z)$  na escala logarítmica generalizada. A seqüência  $x_\gamma(n)$  corresponde, portanto, aos coeficientes de Fourier do espectro logarítmico generalizado  $\ln_\gamma(X(e^{j\omega}))$  e define o cepstrum generalizado. O cepstrum generalizado  $x_\gamma(n)$  é obtido pela transformada  $z$  inversa de (2.7), dada por

$$x_\gamma(n) = \frac{1}{2\pi j} \oint_C X_\gamma(z)z^{n-1}dz \quad (2.8)$$

onde  $C$  é uma curva fechada, tomada no sentido anti-horário, que inclui a origem e se encontra inteiramente na região de convergência (RC) estável do plano  $z$  (ou seja, a RC que inclui a circunferência de raio unitário). É importante considerar que  $\ln_\gamma(X(z))$ , em (2.7), deve estar unicamente definida por (2.6), ou seja, deve ser uma transformada  $z$  válida de uma seqüência real e estável. A restrição de validade da transformada  $z$  é necessária já que função log-gen complexa é multivalente, o que será abordado na seção 2.5.

---

§ A igualdade tem por objetivo mostrar a relação funcional entre  $X_\gamma(z)$  e  $X(z)$ , porém, dada a multivalência da função log-gen complexa, a definição dada por (2.6) é mais restritiva. Assim, a relação (2.7) será considerada válida sob a restrição de que  $\ln_\gamma(X(z))$  seja uma transformada  $z$  válida de uma seqüência real e estável.

### 2.3.2 Magnitude Logarítmica Generalizada

As representações gráficas do espectro logarítmico generalizado poderiam ser feitas com base na magnitude quadrática do espectro log-gen  $|X_\gamma(e^{j\omega})|^2$ . Porém, adota-se a curva  $\ln_\gamma(|X(e^{j\omega})|^2)$ , já que os aspectos de  $|X_\gamma(e^{j\omega})|^2$  relevantes para uma análise sobre  $x_\gamma(n)$ , são contemplados em  $\ln_\gamma(|X(e^{j\omega})|^2)$ , com a vantagem de que, com base nas curvas da Figura 2.1 é possível se prever como a função log-gen atuará sobre  $|X(e^{j\omega})|^2$  com diferentes valores de  $\gamma$ . Isso é razoável, já que  $\ln_\gamma(|X(e^{j\omega})|^2)$  é diretamente descrita por uma série de Fourier com coeficientes baseados no cepstrum generalizado. Utilizando a relação (A.5) do Apêndice A, conclui-se que  $\ln_\gamma(|X(e^{j\omega})|^2)$  é dada por

$$\ln_\gamma(|X(e^{j\omega})|^2) = 2. \operatorname{Re}[X_\gamma(e^{j\omega})] + \gamma. |X_\gamma(e^{j\omega})|^2 \quad (2.9)$$

A representação trigonométrica da expressão (2.9) (com base em (2.6)) resulta em

$$\ln_\gamma(|X(e^{j\omega})|^2) = d_\gamma(0) + 2. \sum_{n=0}^{\infty} d_\gamma(n). \cos(\omega. n) \quad (2.10)$$

onde

$$d_\gamma(n) = \begin{cases} 2. x_\gamma(0) + \gamma. \sum_{k=-\infty}^{\infty} x_\gamma^2(k), & n = 0 \\ x_\gamma(n) + x_\gamma(-n) + \gamma. \sum_{k=-\infty}^{\infty} x_\gamma(k+n)x_\gamma(k), & n > 0 \end{cases} \quad (2.11)$$

Em  $d_\gamma(n)$ , as contribuições de  $x_\gamma(m)$  são provenientes de  $\operatorname{Re}[X_\gamma(e^{j\omega})]$  e aquela de sua auto-correlação (termo do somatório) provém de  $|X_\gamma(e^{j\omega})|^2$ . Assim, a curva de  $\ln_\gamma(|X(e^{j\omega})|^2)$  é contemplada por uma série de Fourier com coeficientes baseados no cepstrum generalizado  $x_\gamma(n)$ . É interessante observar que, quando se assume  $\gamma = 0$ , a expressão (2.9) se reduz a

$$\ln(|X(e^{j\omega})|^2) = 2. \operatorname{Re}[X_0(e^{j\omega})] = 2. \operatorname{Re}[\ln(X(e^{j\omega}))] \quad (2.12)$$

que representa a magnitude espectral na escala logarítmica. Nesse caso ( $\gamma = 0$ ), a expressão (2.10) assume a forma

$$\ln(|X(e^{j\omega})|^2) = 2 \cdot \sum_{n=-\infty}^{\infty} x_0(n) \cdot \cos(\omega n) \quad (2.13)$$

Assim, a magnitude espectral logarítmica é representada por uma série trigonométrica ponderada diretamente pelo cepstrum  $x_0(n)$ .

A Figura 2.3 (a) mostra o espectro de fala em dB da emissão da vogal “i” por uma voz masculina<sup>§</sup>. Na Figura 2.3 (b), o espectro se apresenta na escala linear. As magnitudes espectrais log-gen observadas em (d), (e) e (f), correspondem respectivamente à aplicação das funções log-gen com parâmetros  $\gamma = 1/3$ ,  $\gamma = 0$  e  $\gamma = -1/3$  (curvas da Figura 2.3 (c)) ao espectro na escala linear em (b). Destaca-se a redundância das curvas em (a) e (e), ambas na escala logarítmica e obtidas respectivamente pela aplicação de  $10\log_{10}(|\cdot|^2)$  e  $\ln(|\cdot|^2)$  à magnitude do espectro.

---

<sup>§</sup> O espectro foi obtido a partir de um segmento de 24ms (384 amostras) de um sinal fala com frequência de amostragem de 16kHz. Adotou-se janela de Blackman.

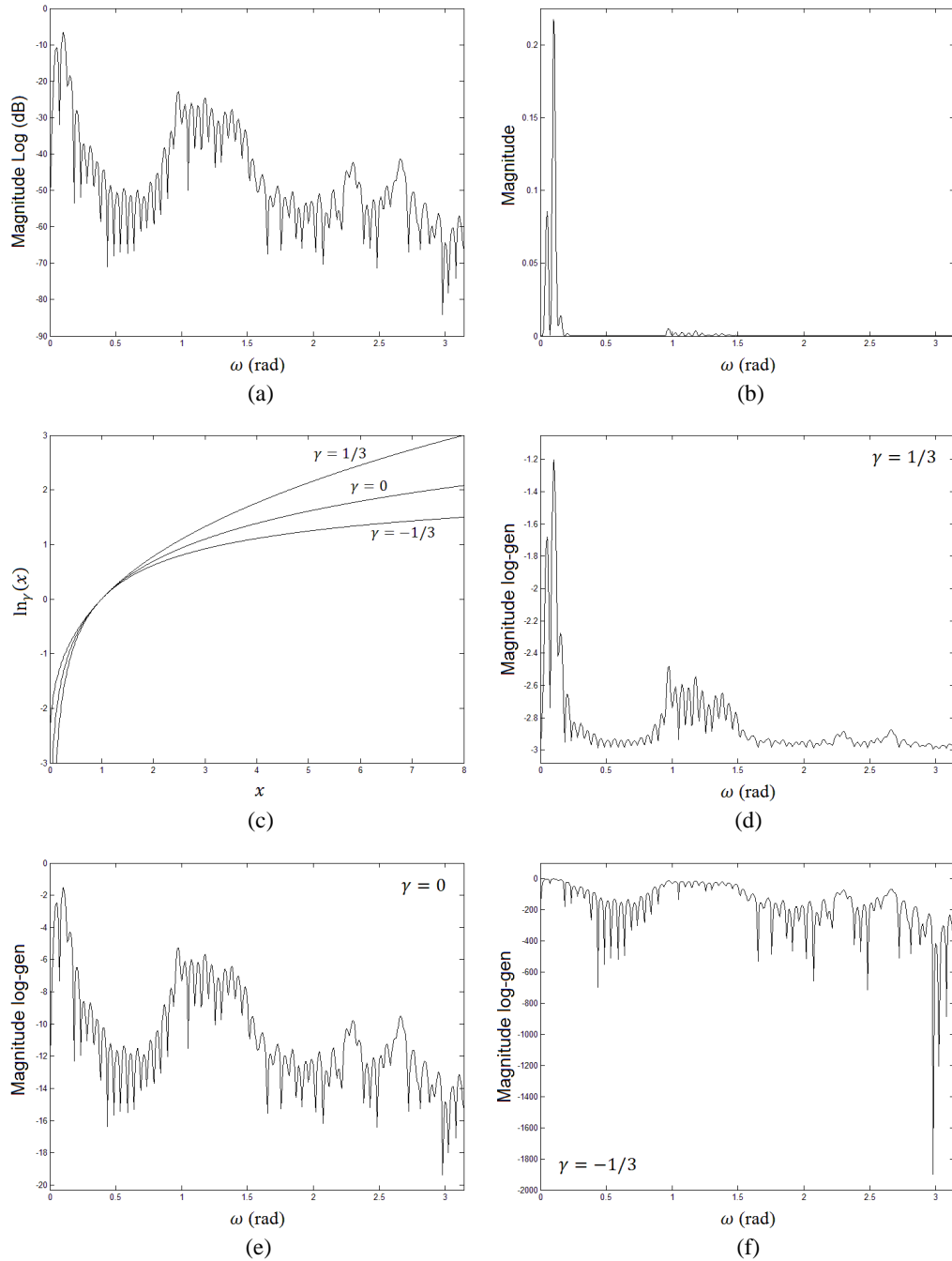


Figura 2.3: (a) Magnitude espectral (dB) de um sinal de fala (voz masculina pronunciando a vogal “i”). (b) Magnitude Espectral (escala linear). (c) Curvas da função log-gen de uma variável real com  $\gamma = 1/3$ ,  $\gamma = 0$ ,  $\gamma = -1/3$ . (d) Magnitude espectral log-gen com  $\gamma = 1/3$ . (e) Magnitude espectral log-gen com  $\gamma = 0$ . (f) Magnitude espectral log-gen com  $\gamma = -1/3$ . *Observação: a magnitude log-gen é expressa por  $\ln_\gamma(|\cdot|^2)$ .*

## 2.4 Representação Espectral Através do Cepstrum Generalizado

O espectro log-gen pode ser reconduzido à escala linear através da aplicação da função log-gen inversa, a exponencial generalizada (exp-gen). Aplicando-a em (2.7), obtém-se

$$X(z) = \exp_{\gamma} \left( X_{\gamma}(z) \right)^{\S} \quad (2.14)$$

Assim

$$X(z) = \exp_{\gamma} \left( \sum_{n=-\infty}^{\infty} x_{\gamma}(n)z^{-n} \right) = \begin{cases} \left( 1 + \gamma \cdot \sum_{n=-\infty}^{\infty} x_{\gamma}(n)z^{-n} \right)^{1/\gamma}, & 0 < |\gamma| \leq 1 \\ \exp \left( \sum_{n=-\infty}^{\infty} x_{\gamma}(n)z^{-n} \right), & \gamma = 0 \end{cases} \quad (2.15)$$

Desta maneira, temos  $X(z)$  representado com base no cepstrum generalizado  $x_{\gamma}(n)$ . Tal representação produz casos especiais cujas formas são conhecidas. Com  $\gamma = -1$ , a expressão (2.15) toma a forma

$$X(z) = \frac{1}{1 - \sum_{n=-\infty}^{\infty} x_{-1}(n)z^{-n}} \quad (2.16)$$

que remete ao modelo de pólos (*all-pole*). Com  $\gamma = 1$ , a expressão obtida é

$$X(z) = 1 + \sum_{n=-\infty}^{\infty} x_1(n)z^{-n} \quad (2.17)$$

que remete ao modelo de zeros (*all-zero*). Comparando tal expressão com a transformada  $z$  de  $x(n)$  (dada por (2.5)), pode-se ver que  $x_1(n)$  está relacionada a  $x(n)$  por

$$x_1(n) = \begin{cases} x(0) - 1, & n = 0 \\ x(n), & |n| > 0 \end{cases} \quad (2.18)$$

---

<sup>\S</sup> Dada a multivalência da função exp-gen complexa, a definição de  $X(z)$  dada por (2.5) é mais restritiva do que aquela dada pela relação (2.14) ou (2.15). Tais relações serão consideradas válidas sob a restrição de que  $\exp_{\gamma} \left( X_{\gamma}(z) \right)$  seja uma transformada  $z$  válida de uma seqüência real e estável.

### 2.4.1 Relação entre $x(0)$ e $x_\gamma(0)$ , e Representação do Espectro através do Cepstrum Generalizado Normalizado

Freqüentemente, será útil que se faça uso da versão normalizada do cepstrum generalizado. A normalização remove a informação de ganho, associada à amostra *zero* da seqüência envolvida, da representação do espectro. Isso permite que ganho e representação espectral normalizada sejam tratados de maneira independente. A condição de normalização é aplicável a  $x(n)$  e  $x_\gamma(n)$  causais, característica das seqüências de fase mínima às quais se restringirão as aplicações abordadas no presente texto. Sob a condição de causalidade, as representações de  $X(z)$ , através de  $x(n)$  e  $x_\gamma(n)$ , são dadas respectivamente por

$$X(z) = \sum_{n=0}^{\infty} x(n)z^{-n} \quad (2.19)$$

e

$$X(z) = \exp_\gamma \left( \sum_{n=0}^{\infty} x_\gamma(n)z^{-n} \right) \quad (2.20)$$

A condição de normalização exige que a amostra  $x'(0)$  da seqüência  $x'(n)$ , versão normalizada de  $x(n)$ , tenha valor unitário. O teorema do valor inicial [17], aplicável a seqüência causais, fornece a amostra inicial de uma seqüência e permite relacionar  $x(0)$  e  $x_\gamma(0)$ . Considerando  $x(n) = 0$  para  $n < 0$ , temos

$$\lim_{|z| \rightarrow \infty} X(z) = \lim_{|z| \rightarrow \infty} \left( \sum_{n=0}^{\infty} x(n)z^{-n} \right) = x(0) \quad (2.21)$$

O mesmo limite aplicado a  $X(z) = \exp_\gamma \left( X_\gamma(z) \right)$  (com  $x_\gamma(n) = 0$  para  $n < 0$ ) resulta em

$$\lim_{|z| \rightarrow \infty} X(z) = \lim_{|z| \rightarrow \infty} \left[ \exp_\gamma \left( \sum_{n=0}^{\infty} x_\gamma(n)z^{-n} \right) \right] = \exp_\gamma \left( x_\gamma(0) \right) \quad (2.22)$$

o que fornece a relação entre  $x(0)$  e  $x_\gamma(0)$

$$x(0) = \exp_\gamma \left( x_\gamma(0) \right) \quad (2.23)$$

A versão normalizada de  $X(z)$ , denotada por  $X'(z)$ , deve obedecer ao critério

$$\lim_{|z| \rightarrow \infty} X'(z) = 1 \quad (2.24)$$

que é obtido, como se pode deduzir de (2.21), dividindo-se  $X(z)$  por  $x(0)$ , assim

$$X'(z) = \frac{X(z)}{x(0)} = \sum_{n=0}^{\infty} \frac{x(n)}{x(0)} z^{-n} \quad (2.25)$$

$X(z)$  pode, então, ser representado por

$$X(z) = K \cdot X'(z) = K \cdot \left( 1 + \sum_{n=1}^{\infty} x'(n) z^{-n} \right) \quad (2.26)$$

onde  $K$  é a constante de ganho dada por

$$K = x(0) \quad (2.27)$$

e  $x'(n)$  representa a versão normalizada de  $x(n)$ , dada por

$$x'(n) = \frac{x(n)}{x(0)} \quad (x(0) \neq 0) \quad (2.28)$$

Aplicando-se o critério de normalização (2.24) a  $X'(z) = \exp_{\gamma}(X_{\gamma}'(z))$ , agora expresso por

$$\lim_{|z| \rightarrow \infty} X'(z) = \lim_{|z| \rightarrow \infty} \left[ \exp_{\gamma}(X_{\gamma}'(z)) \right] = 1 \quad (2.29)$$

conclui-se, como se pode deduzir de (2.22), que

$$X'(z) = \exp_{\gamma}(X_{\gamma}'(z)) = \frac{\exp_{\gamma}(X_{\gamma}(z))}{\exp_{\gamma}(x_{\gamma}(0))} \quad (2.30)$$

Utilizando a relação (A.17) (Apêndice A) para o quociente entre exponenciais generalizadas, chega-se a  $X'(z)$  dado por

$$X'(z) = \exp_{\gamma} \left( \sum_{n=1}^{\infty} \left( \frac{x_{\gamma}(n)}{1 + \gamma \cdot x_{\gamma}(0)} \right) z^{-n} \right) \quad (2.31)$$

Assim,  $X(z)$  pode ser representado por

$$X(z) = K \cdot \exp_{\gamma} \left( \sum_{n=1}^{\infty} x_{\gamma}'(n) z^{-n} \right) \quad (2.32)$$

onde

$$K = \exp_{\gamma} (x_{\gamma}(0)) \quad (2.33)$$

e

$$x_{\gamma}'(n) = \frac{x_{\gamma}(n)}{1 + \gamma \cdot x_{\gamma}(0)}, \quad n > 0 \quad (2.34)$$

corresponde ao cepstrum generalizado normalizado. A expressão (2.34) é responsável por remover a informação de ganho de  $x_{\gamma}(n)$ , fornecendo  $x_{\gamma}'(n)$ . Vale destacar que, quando  $\gamma = 0$  (cepstrum), a expressão (2.34) se reduz a  $x_0'(n) = x_0(n)$ , o que mostra que o cepstrum não carrega informação de ganho. Assim, basta que seja removido o termo  $x_0(0)$  do somatório em (2.20) para que se obtenha a representação normalizada  $X'(z)$ .

## 2.5 Critério para a Unicidade de $\ln_{\gamma}(X(z))$

A função log-gen de uma variável complexa é multivalente. Na seção 2.3.1, a validade de (2.7) ( $X_{\gamma}(z) = \ln_{\gamma}(X(z))$ ) é condicionada à restrição de que  $\ln_{\gamma}(X(z))$  seja uma transformada  $z$  válida. Mais do que isso, dado que  $X(z)$  é a transformada  $z$  de uma seqüência  $x(n)$  real e estável,  $\ln_{\gamma}(X(z))$  deve ser unicamente determinada pela transformada  $z$  de uma seqüência  $x_{\gamma}(n)$  real (e sua estabilidade decorrerá da estabilidade de  $x(n)$ ).

Estando a seqüência  $x(n)$  restrita à condição de ser real e estável, sua transformada  $z$ ,  $X(z)$ , possui condições específicas de simetria e sua região de convergência (RC) inclui a CRU. A última característica restringe a RC de  $\ln_{\gamma}(X(z))$  de maneira que ela também venha a incluir a CRU. Dadas tais restrições, é preciso lidar com a multivalência de  $\ln_{\gamma}(X(z))$ . A imposição de que  $\ln_{\gamma}(X(z))$  seja analítica no seu domínio de convergência faz com que ela admita representação por transformada  $z$ . A imposição de uma última condição de simetria a  $\ln_{\gamma}(X(z))$  e o fato de sua RC incluir a CRU, são as condições que levam  $\ln_{\gamma}(X(z))$  a ser definida pela transformada  $z$  de uma



seqüência  $x_\gamma(n)$  real e estável, o cepstrum generalizado. A multivalência de  $\ln_\gamma(X(z))$  e o critério para sua unicidade serão abordados na presente seção.

### 2.5.1 Multivalência de $\ln_\gamma(X(z))$

Considerando  $x(n)$  uma seqüência estável, restringimos a região de convergência de  $X(z)$  àquela que inclui a circunferência de raio unitário (CRU,  $|z| = 1$ ), de maneira que existem  $X(e^{j\omega})$  e  $\ln_\gamma(X(e^{j\omega}))$ . Dada tal restrição, consideremos

$$X(e^{j\omega}) = |X(e^{j\omega})|. e^{j.\arg(X(e^{j\omega}))} \quad (2.35)$$

Denotando por  $\theta(\omega)$  o argumento de  $X(e^{j\omega})$  definido unicamente pela sua restrição ao intervalo  $]-\pi, \pi]$  (valores principais dos pontos da curva de  $\arg(X(e^{j\omega}))$ ), então  $\arg(X(e^{j\omega}))$ , no ponto  $\omega = \omega_0$ , pode ser representado por

$$\arg(X(e^{j\omega_0})) = \theta(\omega_0) + k.2\pi, \quad k = 0, \pm 1, \pm 2, \dots, \pm\infty \quad (2.36)$$

o que expressa o fato de que valores de argumento de um número complexo distanciados por  $k.2\pi$  representam o mesmo ponto no plano complexo. Ou seja, com base em (2.36), os pontos da curva de  $\arg(X(e^{j\omega}))$  podem ser rearranjados de infinitas maneiras, todas elas, representando uma mesma curva  $X(e^{j\omega})$ .

Aplicando a função log-gen a (2.35) obtém-se

$$\ln_\gamma(X(e^{j\omega})) = \begin{cases} \frac{[X(e^{j\omega})]^\gamma - 1}{\gamma}, & 0 < |\gamma| \leq 1 \\ \ln(X(e^{j\omega})) & \gamma = 0 \end{cases} \quad (2.37)$$

No caso em que  $0 < |\gamma| \leq 1$ , substituindo a expressão (2.35) em (2.37), obtém-se

$$\ln_\gamma(X(e^{j\omega})) = \frac{|X(e^{j\omega})|^\gamma . e^{j.\gamma.\arg(X(e^{j\omega}))} - 1}{\gamma} \quad (2.38)$$

---

<sup>§</sup> É necessário que  $\theta(\omega)$  esteja unicamente definida, porém sua escolha é arbitrária (desde que feita sob o critério  $e^{j.\theta(\omega)} = e^{j.\arg(X(e^{j\omega}))}$ ). A definição que restringe os pontos de sua curva aos valores principais de  $\arg(X(e^{j\omega}))$  é aqui utilizada por ser de uso habitual.

Se considerarmos o argumento do termo  $[X(e^{j\omega})]^\gamma = |X(e^{j\omega})|^\gamma \cdot e^{j\gamma \cdot \arg(X(e^{j\omega}))}$  de  $\ln_\gamma(X(e^{j\omega}))$ , e  $\gamma = n/m$ , com  $n$  inteiro,  $m$  natural, ambos não-nulos e  $|n| < m$ , temos, no ponto  $\omega = \omega_0$ ,

$$\arg([X(e^{j\omega_0})]^\gamma) = \gamma \cdot \arg(X(e^{j\omega_0})) = \frac{n}{m} \theta(\omega_0) + k \cdot 2 \frac{n}{m} \pi, \quad (2.39)$$

$$k = 0, \pm 1, \pm 2, \dots, \pm \infty$$

onde se vê que valores antes distanciados por  $k \cdot 2\pi$  em  $\arg(X(e^{j\omega}))$ , representando um mesmo ponto na curva de  $X(e^{j\omega})$ , agora correspondem a valores distanciados por um múltiplo de uma fração de  $2\pi$  ( $k \cdot 2 \frac{n}{m} \pi$ ), representando pontos distintos em  $[X(e^{j\omega})]^\gamma$  e também em  $\ln_\gamma(X(e^{j\omega}))$  (com exceção aos casos em que  $k$  é múltiplo de  $m$ ). Desta maneira, a multiplicidade de representações para  $\arg(X(e^{j\omega}))$ , leva à multivalência de  $\ln_\gamma(X(z))$ . No caso em que  $\gamma = 0$ , substituindo a expressão (2.35) em (2.37), obtém-se

$$\ln(X(e^{j\omega})) = \ln(|X(e^{j\omega})|) + j \cdot \arg(X(e^{j\omega})) \quad (2.40)$$

Substituindo, no ponto  $\omega = \omega_0$ , a expressão (2.36) em (2.40), obtém-se

$$\ln(X(e^{j\omega_0})) = \ln(|X(e^{j\omega_0})|) + j \cdot (\theta(\omega_0) + k \cdot 2\pi), \quad (2.41)$$

$$k = 0, \pm 1, \pm 2, \dots, \pm \infty$$

onde se observa que, com o argumento de  $X(e^{j\omega})$  agora representando a parte imaginária de  $\ln(X(e^{j\omega}))$ , valores distanciados de  $k \cdot 2\pi$  em  $\arg(X(e^{j\omega}))$  correspondem a valores distanciados de  $k \cdot 2\pi$  em  $\text{Im}[\ln(X(e^{j\omega}))]$  e conduzem, portanto, a pontos distintos em  $\ln(X(e^{j\omega}))$ . Assim se manifesta a multivalência de  $\ln_\gamma(X(z))$  para  $\gamma = 0$ .

## 2.5.2 Critério para Unicidade de $\ln_\gamma(X(z))$

A admissão de saltos de valores por  $k \cdot 2\pi$  (com  $k$  podendo assumir qualquer valor inteiro) para cada ponto de  $\arg(X(e^{j\omega}))$ , faz com que sua curva admita uma infinidade

de representações descontínuas, além de uma representação contínua (deslocável por múltiplos inteiros de  $2\pi$  no eixo das ordenadas). As descontinuidades em  $\arg(X(e^{j\omega}))$ , por representarem saltos de  $k \cdot \gamma \cdot 2\pi$  no argumento de  $[X(e^{j\omega})]^\gamma$  (para  $0 < |\gamma| \leq 1$ ) ou saltos de  $k \cdot 2\pi$  na parte imaginária de  $\ln(X(e^{j\omega}))$ , representam descontinuidades em  $\ln_\gamma(X(z))$ . Se, por exemplo, for utilizada como critério de desambiguação a habitual restrição de  $\arg(X(e^{j\omega}))$  ao seu valor principal  $-\pi < \arg(X(e^{j\omega})) \leq \pi$ , descontinuidades ocorrem quando sua curva *cruza* os limites  $\pi$  ou  $-\pi$  do eixo das ordenadas.

A descontinuidade de  $\ln_\gamma(X(z))$  em um ponto  $z = z_0$  do domínio de convergência implica em que não existe a derivada  $\frac{d}{dz} [\ln_\gamma(X(z))]_{z=z_0}$  portanto  $\ln_\gamma(X(z))$  não é analítica no domínio de convergência, o que é exigido para que  $\ln_\gamma(X(z))$  seja uma transformada  $z$  válida. Assim, a exigência de analiticidade recai sobre a exigência de continuidade para  $\arg(X(z))$ . A partir da expressão para a derivada  $\frac{d}{dz} [\ln_\gamma(X(z))]$  dada por (A.9) (Apêndice A), pode-se estabelecer a relação

$$\frac{1}{1 + \gamma \cdot \ln_\gamma(X(z))} \cdot \frac{d}{dz} [\ln_\gamma(X(z))] = \frac{1}{X(z)} \cdot \frac{dX(z)}{dz} \quad (2.42)$$

Expressando  $X(z)$  por

$$X(z) = |X(z)| \cdot e^{j \cdot \arg(X(z))} \quad (2.43)$$

o lado direito da igualdade em (2.42) é dado por

$$\frac{1}{X(z)} \cdot \frac{dX(z)}{dz} = \frac{1}{|X(z)|} \cdot \frac{d|X(z)|}{dz} + j \cdot \frac{d}{dz} [\arg(X(z))] \quad (2.44)$$

onde a presença do termo  $j \cdot \frac{d}{dz} [\arg(X(z))]$  mostra que a analiticidade de  $X(z)$  se expressa para  $\arg(X(z))$  contínua em todo o domínio de convergência de  $X(z)$ . Assim, a relação entre  $\ln_\gamma(X(z))$  e  $X(z)$  dada por (2.42), representa um critério para a analiticidade de  $\ln_\gamma(X(z))$  em seu domínio de convergência.

Das condições estabelecidas no início da seção para  $x(n)$ , sua estabilidade implica na existência de  $\ln_\gamma(X(e^{j\omega}))$ , o que leva à estabilidade de  $x_\gamma(n)$ . Já o fato de

$x(n)$  ser real implica em que a representação contínua de  $\arg(X(e^{j\omega}))$  tenha simetria ímpar sob a condição  $\arg(X(e^{j\omega}))_{\omega=0} = 0$ . Esta última condição é satisfeita pela exigência de que a amostra inicial  $x_\gamma(0)$  do cepstrum-gen seja real. Uma vez que a relação (2.42) é uma equação diferencial, não contempla a amostra  $x_\gamma(0)$ , já que esta representa o termo constante de  $\ln_\gamma(X(z))$ . A obtenção de  $x_\gamma(0)$  será abordada na seção 2.7.

O desenvolvimento apresentado até o momento apontou os critérios para a unicidade da função  $\ln_\gamma(X(z))$ , de maneira que ela conduzisse a uma seqüência  $x_\gamma(n)$  real e estável, assumindo conhecida a seqüência  $x(n)$  real e estável. Se for considerada a situação inversa, em que se parte de uma seqüência  $x_\gamma(n)$  conhecida, argumentos simétricos aos apresentados aqui valem para a unicidade de  $X(z)$ , já que, para  $0 < |\gamma| \leq 1$ ,

$$X(z) = \exp_\gamma(X_\gamma(z)) = \left(1 + \gamma \cdot X_\gamma(z)\right)^{1/\gamma} \quad (2.45)$$

onde  $X_\gamma(z)$  representa a transformada  $z$  de  $x_\gamma(n)$  (definida em (2.6)). Em (2.45), quando o expoente  $1/\gamma$  assume valores não inteiros, a função exponencial generalizada é multivalente. Assim, reescrevendo a equação (2.42) como

$$\frac{1}{1 + \gamma \cdot X_\gamma(z)} \cdot \frac{dX_\gamma(z)}{dz} = \frac{1}{\exp_\gamma(X_\gamma(z))} \cdot \frac{d}{dz} \left[ \exp_\gamma(X_\gamma(z)) \right] \quad (2.46)$$

por argumentos semelhantes àqueles dados para o lado direito da igualdade em (2.42), o lado esquerdo de (2.46) impõe analiticidade a  $\exp_\gamma(X_\gamma(z))$  em seu domínio de convergência. Da mesma maneira que a equação (2.42) não contempla a amostra  $x_\gamma(0)$ , a equação (2.46) não contempla a amostra  $x(0)$ , cuja obtenção será abordada na seção 2.7. Dado o caráter simétrico das relações entre  $\ln_\gamma(X(z))$  e  $X(z)$  e entre  $\exp_\gamma(X_\gamma(z))$  e  $X_\gamma(z)$ , as equações (2.42) e (2.46) ficam melhor representadas por

$$\frac{1}{1 + \gamma \cdot X_\gamma(z)} \cdot \frac{dX_\gamma(z)}{dz} = \frac{1}{X(z)} \cdot \frac{dX(z)}{dz} \quad (2.47)$$

de onde é possível a derivação de relações implícitas entre  $x(n)$  e  $x_\gamma(n)$  e entre diferentes seqüências  $x_\gamma(n)$ .

### 2.5.3 Relação Implícita entre $x_\gamma(n)$ e $x(n)$

Rearranjando a expressão (2.47) obtém-se

$$z \cdot \frac{dX_\gamma(z)}{dz} \cdot X(z) = z \cdot \frac{dX(z)}{dz} \cdot [1 + \gamma \cdot X_\gamma(z)] \quad (2.48)$$

Considerando os pares de transformada  $z$

$$-n \cdot x(n) \leftrightarrow z \cdot \frac{dX(z)}{dz} \quad \text{e} \quad -n \cdot x_\gamma(n) \leftrightarrow z \cdot \frac{dX_\gamma(z)}{dz} \quad (2.49)$$

a partir da aplicação da transformada  $z$  inversa à equação (2.48) obtém-se

$$\sum_{k=-\infty}^{\infty} \left(\frac{k}{n}\right) \cdot x_\gamma(k) \cdot x(n-k) = x(n) + \gamma \cdot \sum_{k=-\infty}^{\infty} \left(\frac{k}{n}\right) \cdot x(k) \cdot x_\gamma(n-k), \quad (2.50)$$

$$n \neq 0$$

que representa uma relação implícita entre a seqüência  $x(n)$  e o cepstrum generalizado  $x_\gamma(n)$ . Sob certas condições, tal relação pode ser rearranjada em fórmulas recursivas para a computação de  $x_\gamma(n)$  a partir de  $x(n)$  e vice-versa.

### 2.5.4 Relação Implícita entre $x_{\gamma_2}(n)$ e $x_{\gamma_1}(n)$

A equação (2.47) é válida para qualquer valor de  $\gamma$ , assim,

$$\frac{1}{1 + \gamma_1 \cdot X_{\gamma_1}(z)} \cdot \frac{dX_{\gamma_1}(z)}{dz} = \frac{1}{1 + \gamma_2 \cdot X_{\gamma_2}(z)} \cdot \frac{dX_{\gamma_2}(z)}{dz} \quad (2.51)$$

com  $\gamma_1 \neq \gamma_2$ . A equação (2.51), rearranjada, fornece

$$z \cdot \frac{dX_{\gamma_1}(z)}{dz} \cdot [1 + \gamma_2 \cdot X_{\gamma_2}(z)] = z \cdot \frac{dX_{\gamma_2}(z)}{dz} \cdot [1 + \gamma_1 \cdot X_{\gamma_1}(z)] \quad (2.52)$$

A partir da aplicação da transformada  $z$  inversa à equação (2.52), obtém-se

$$x_{\gamma_1}(n) + \gamma_2 \cdot \sum_{k=-\infty}^{\infty} \left(\frac{k}{n}\right) \cdot x_{\gamma_1}(k) \cdot x_{\gamma_2}(n-k)$$

$$= x_{\gamma_2}(n) + \gamma_1 \cdot \sum_{k=-\infty}^{\infty} \left(\frac{k}{n}\right) \cdot x_{\gamma_2}(k) \cdot x_{\gamma_1}(n-k), \quad n \neq 0 \quad (2.53)$$

que representa uma relação implícita entre as seqüências  $x_{\gamma_2}(n)$  e  $x_{\gamma_1}(k)$ , ou seja, entre dois cepstrums generalizados associados a valores distintos de  $\gamma$ . Sob certas condições, tal relação pode ser rearranjada em uma fórmula recursiva para a computação de  $x_{\gamma_2}(n)$  a partir de  $x_{\gamma_1}(k)$ .

## 2.6 Critério de Fase Mínima para $x(n)$

Para fins de generalização, até o momento foram assumidas seqüências  $x(n)$  e  $x_\gamma(n)$  sem restrições quanto à causalidade (com exceção da seção 2.4.1, onde se assumiu causalidade para a formulação apresentada). Porém, para os propósitos do presente trabalho, o cepstrum generalizado deverá estar associado a uma seqüência  $x(n)$  de fase mínima, o que o restringe à causalidade. Como em geral dispomos de uma seqüência de fase arbitrária, sua versão de fase mínima deve ser obtida. Na presente seção, deduz-se uma condição que permite estabelecer um critério para a seqüência de fase mínima. Tal dedução permite, ainda, concluir a causalidade do cepstrum-gen sob a condição de fase mínima.

Uma seqüência  $x(n)$  de fase mínima é caracterizada por ter todos os pólos e zeros de sua transformada  $z$ ,  $X(z)$ , no interior da CRU. Tal condição implica em que, se é conhecida a magnitude de  $X(e^{j\omega})$ , sua fase fica unicamente determinada. Isto procede, uma vez que a diversidade de fases admitidas para uma mesma magnitude espectral decorre da ambigüidade no posicionamento de pólos e zeros, que podem, individualmente, estar no interior ou no exterior da CRU, em posições mutuamente inter-relacionadas. Naturalmente que a estabilidade de uma seqüência de fase mínima está diretamente vinculada à causalidade. Isto é justificado pelo fato de que, uma vez que todos os pólos se encontram no interior da CRU, a região de convergência deve corresponder à região externa à circunferência delimitada pelo pólo mais afastado da origem para que a CRU esteja nela inclusa (condição para estabilidade). Uma importante implicação da característica de fase mínima de  $X(z)$  é que, sua inversa  $1/X(z)$  também é de fase mínima. Assim, ambos  $X(z)$  e  $1/X(z)$  estão associados a seqüências causais e estáveis.

A relação entre o conjunto de pólos e zeros de  $X(z)$  de fase mínima e os pontos singulares de  $\ln_\gamma(X(z))$  permite que se deduza, como já citado, a causalidade do

cepstrum generalizado, e fornece a condição para que, a partir de uma seqüência conhecida de fase genérica, seja possível a obtenção de sua versão de fase mínima.

### 2.6.1 PSDs de $\ln_\gamma(X(z))$ e Condição de Fase Mínima

A abordagem a seguir leva à condição de fase mínima que é aqui tratada no contexto amplo do cepstrum generalizado e permite concluir que o cepstrum generalizado associado a uma seqüência de fase mínima é causal.

Com  $F(z)$  representando uma função racional de pólos e zeros genérica, das singularidades de  $\ln_\gamma(F(z))$  são de interesse para este tópico somente aquelas com comportamento divergente, ou seja, se

$$\lim_{z \rightarrow z_0} |\ln_\gamma(F(z))| = \infty \quad (2.54)$$

então  $z_0$  é um ponto singular divergente de  $\ln_\gamma(F(z))$ . Os pontos singulares de  $\ln_\gamma(F(z))$  são denominados de pontos de ramificação, uma vez que  $\ln_\gamma(F(z))$  é uma função multivalente e portanto ramificada em torno de tais pontos (o valor principal de uma função é um de seus ramos) [19]. Impondo analiticidade à função, as características de tais pontos restringem-se àquelas de interesse para as análises a seguir. De importância, serão os pontos de ramificação de  $\ln_\gamma(X(z))$  com comportamento de pólo no que se refere à sua característica divergente. Tais pontos serão aqui denominados de pontos singulares divergentes (PSDs) de  $\ln_\gamma(X(z))$ . Se, por exemplo, assumirmos que  $F(z)$  é dada por

$$F(z) = \frac{1 - bz^{-1}}{1 - az^{-1}} \quad (2.55)$$

O ponto  $z = a$  é um pólo de  $F(z)$  e é um PSD de  $[F(z)]^{|\gamma|}$ , uma vez que  $\lim_{z \rightarrow a} |[F(z)]^{|\gamma|} = \infty$ . O ponto  $z = b$ , por sua vez, é um zero de  $F(z)$  e é um ponto singular sem comportamento divergente em  $[F(z)]^{|\gamma|}$ , já que  $\lim_{z \rightarrow b} |[F(z)]^{|\gamma|} = 0$ . No caso de  $\ln(F(z))$ , ambos os limites  $z \rightarrow a$  e  $z \rightarrow b$  aplicados a  $|\ln(F(z))|$  divergem, então, ambos o pólo  $z = a$  e zero  $z = b$  de  $F(z)$  são PSDs em  $\ln(F(z))$ .

Será assumido aqui, que a função  $\ln_\gamma(X(z))$  é analítica no seu domínio de convergência.  $\ln_\gamma(X(z))$  pode ser explicitada da seguinte maneira

$$\ln_\gamma(X(z)) = \begin{cases} \frac{1}{\gamma} \{[X(z)]^\gamma - 1\}, & 0 < \gamma \leq 1 \\ \ln(X(z)), & \gamma = 0 \\ \frac{1}{\gamma} \left\{ \left[ \frac{1}{X(z)} \right]^{|\gamma|} - 1 \right\}, & -1 \leq \gamma < 0 \end{cases} \quad (2.56)$$

Cada um dos casos acima será avaliado, considerando que  $X(z)$  é de fase mínima, ou seja, que todos os seus pólos e zeros se encontram no interior da CRU.

Para  $0 < \gamma \leq 1$ : como os PSDs de  $[X(z)]^\gamma$  são os pólos de  $X(z)$ , os PSDs de  $\ln_\gamma(X(z))$  também são os pólos de  $X(z)$ . Como os pólos de  $X(z)$  estão no interior da CRU, os PSDs de  $\ln_\gamma(X(z))$  também se encontram no interior da CRU.

Para  $-1 \leq \gamma < 0$ : os zeros de  $X(z)$  se transformam em pólos de  $1/X(z)$  e portanto em PSDs de  $[1/X(z)]^{|\gamma|}$ . Assim, os PSDs de  $\ln_\gamma(X(z))$  são os zeros de  $X(z)$ . Como os zeros de  $X(z)$  se encontram no interior da CRU, os PSDs de  $\ln_\gamma(X(z))$  também se encontram no interior da CRU.

Para  $\gamma = 0$ , os pólos de  $X(z)$  são PSDs em  $\ln(X(z))$ . O mesmo vale para os zeros de  $X(z)$ , uma vez que  $\ln(v)$  tem um PSD em  $v = 0$ . Assim, ambos os pólos e os zeros de  $X(z)$  são os PSDs de  $\ln(X(z))$ . Como os pólos e zeros de  $X(z)$  se encontram no interior da CRU, os PSDs de  $\ln_\gamma(X(z))$  também se encontram no interior da CRU. A recíproca é verdadeira, ou seja, a constatação de que todos os PSDs da função  $\ln(X(z))$ , onde  $X(z)$  é racional, se encontram no interior da CRU é suficiente para que se afirme que  $X(z)$  é de fase mínima. Assim se estabelece a condição de fase mínima com base nos PSDs de  $\ln(X(z))$ . Essa propriedade da função logarítmica será utilizada para que, a partir de uma seqüência conhecida de fase genérica, seja obtida a sua versão de fase mínima.

A avaliação acima mostra que a propriedade de fase mínima de  $X(z)$  implica em que, para qualquer valor que  $\gamma$  possa assumir, os PSDs de  $\ln_\gamma(X(z))$  se encontrarão no interior da CRU. Com base no comportamento de pólo de um PSD, considera-se aqui que, da mesma maneira como a presença dos pólos de uma função racional estável no interior da CRU justifica a causalidade, a presença dos PSDs de  $\ln_\gamma(X(z))$  no interior da CRU implica na causalidade do cepstrum generalizado. Com base em tais considerações e naquelas para os PSDs de  $\ln(X(z))$  ( $\gamma = 0$ ), a constatação de que a



seqüência  $x_0(n)$  (associada a  $\ln(X(z))$ ) é causal e estável é suficiente para que se afirme que  $X(z)$  é de fase mínima. A condição de fase mínima, antes estabelecida com base nos PSDs de  $\ln(X(z))$ , é, portanto, aqui reafirmada com base em  $x_0(n)$ , seqüência denominada *cepstrum complexo*.

## 2.6.2 Cepstrum Complexo: Critério de Fase Mínima

De posse de uma seqüência  $\bar{x}(n)$  estável e de fase genérica, cuja transformada  $z$  é denotada por  $\bar{X}(z)$ , a constatação de que todos os seus pólos se encontram no interior da CRU, levaria à conclusão de que  $\bar{x}(n)$  é causal e estável, porém, não necessariamente de fase mínima. Por outro lado, dado que os pólos e zeros de  $\bar{X}(z)$  são todos convertidos em PSDs em  $\ln(\bar{X}(z))$ , a constatação de que  $\ln(\bar{X}(z))$  tem todos os seus PSDs no interior da CRU levaria à conclusão de que  $\bar{x}(n)$  é de fase mínima. Considerando que  $\bar{x}(n)$  tem fase genérica, um critério para a obtenção de sua versão de fase mínima deve conduzir os PSDs de  $\ln(\bar{X}(z))$  ao interior da CRU. Isso seria equivalente a exigir que  $\ln(\bar{X}(z))$  tivesse sua versão modificada  $\ln(X(z))$  representada pela transformada  $z$  de uma seqüência causal e estável. Com base nessas condições determinaremos uma seqüência de fase mínima  $x(n)$  a partir da seqüência de fase genérica  $\bar{x}(n)$ . Com  $\bar{X}(e^{j\omega})$  e  $X(e^{j\omega})$  denotando, respectivamente, as transformadas de Fourier de  $\bar{x}(n)$  e  $x(n)$ , temos

$$|X(e^{j\omega})| = |\bar{X}(e^{j\omega})| \quad (2.57)$$

e, com base na condição de fase mínima, deseja-se determinar  $\arg(X(e^{j\omega}))$  e portanto  $X(e^{j\omega})$ . Aplicando a função logarítmica a  $\bar{X}(e^{j\omega})$ , obtém-se

$$\ln(\bar{X}(e^{j\omega})) = \ln(|\bar{X}(e^{j\omega})|) + j \cdot \arg(\bar{X}(e^{j\omega})) \quad (2.58)$$

Estamos, porém, interessados em

$$\ln(X(e^{j\omega})) = \ln(|\bar{X}(e^{j\omega})|) + j \cdot \arg(X(e^{j\omega})) \quad (2.59)$$

onde  $\ln(|\bar{X}(e^{j\omega})|)$  é conhecido e devemos determinar  $\arg(X(e^{j\omega}))$  e  $\ln(X(e^{j\omega}))$ . A condição de fase mínima corresponde à exigência de que a seqüência  $x_0(n)$ , associada a  $\ln(X(e^{j\omega}))$ , seja causal e estável. A restrição de causalidade a  $x_0(n)$ , traz a

propriedade de que, uma vez conhecido  $\text{Re} \left[ \ln \left( X(e^{j\omega}) \right) \right]$ , determina-se completamente  $\text{Im} \left[ \ln \left( X(e^{j\omega}) \right) \right]$  e  $\ln \left( X(e^{j\omega}) \right)$ . Além da restrição de causalidade, a exigência de que  $x_0(n)$  seja real permite que a propriedade mencionada possa ser aplicada com correspondência direta no domínio do tempo. Nesse caso, o problema é reduzido àquele de se determinar a seqüência  $x_0(n)$ . Tal seqüência pode ser representada por

$$x_0(n) = x_{0p}(n) + x_{0i}(n) \quad (2.60)$$

onde  $x_{0p}(n)$  e  $x_{0i}(n)$  são as partes par e ímpar de  $x_0(n)$ , respectivamente dadas por

$$x_{0p}(n) = \frac{1}{2} [x_0(n) + x_0(-n)] \quad (2.61)$$

$$x_{0i}(n) = \frac{1}{2} [x_0(n) - x_0(-n)] \quad (2.62)$$

Quando  $x_0(n)$  é real, são válidos os seguintes pares de transformada de Fourier

$$x_{0p}(n) \leftrightarrow \text{Re} \left[ \ln \left( X(e^{j\omega}) \right) \right] \quad \text{e} \quad x_{0i}(n) \leftrightarrow j \cdot \text{Im} \left[ \ln \left( X(e^{j\omega}) \right) \right] \quad (2.63)$$

ou seja

$$x_{0p}(n) \leftrightarrow \ln(|\bar{X}(e^{j\omega})|) \quad \text{e} \quad x_{0i}(n) \leftrightarrow j \cdot \arg \left( X(e^{j\omega}) \right) \quad (2.64)$$

Com base em (2.61) e (2.62), pode-se deduzir que, para  $x_0(n)$  causal (condição de fase mínima), as seqüências  $x_{0i}(n)$  e  $x_0(n)$  são completamente determinadas a partir de  $x_{0p}(n)$  [17]. Uma vez que se conhece  $\ln(|\bar{X}(e^{j\omega})|)$ , é possível a obtenção de  $x_{0p}(n)$  por

$$x_{0p}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(|\bar{X}(e^{j\omega})|) e^{j\omega n} d\omega \quad (2.65)$$

A exigência de causalidade para a seqüência  $x_0(n)$  implica em que ela possa ser determinada, com base em (2.61), por

$$x_0(n) = \begin{cases} 2 \cdot x_{0p}(n), & n > 0 \\ x_{0p}(n), & n = 0 \\ 0, & n < 0 \end{cases} \quad (2.66)$$

e  $\ln(X(e^{j\omega}))$  estaria determinada pela transformada de Fourier de  $x_0(n)$ . A seqüência  $x_0(n)$  pode, porém, conduzir diretamente à seqüência de fase mínima  $x(n)$ , dada a propriedade mútua de causalidade de  $x_0(n)$  e  $x(n)$ . Isso será abordado na próxima seção. As expressões (2.65) e (2.66) definem o critério de fase mínima, que deverá estar satisfeito para que se prossigam os cálculos do cepstrum-gen.

A seqüência  $x_{0p}(n)$ , definida por (2.65), é denominada *cepstrum real*, nome dado pelo fato dela ser obtida através da transformada inversa de Fourier do logaritmo da magnitude espectral, que é real.

A seqüência  $x_0(n)$  é denominada *cepstrum complexo*, uma vez que ela é definida por

$$x_0(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(X(e^{j\omega})) e^{j\omega n} d\omega \quad (2.67)$$

(expressão válida com  $\ln(X(z))$  analítica no domínio de convergência), representando a transformada inversa de Fourier do logaritmo de uma função complexa. A computação de integrais de funções complexas multivalentes como (2.67), exige desempacotamento de fase para que sejam removidas as descontinuidades de  $\arg(X(e^{j\omega}))$ , o que é uma tarefa custosa. Como os cálculos estarão restritos a seqüências  $x(n)$  de fase mínima, as transformações entre  $x(n)$  e  $x_\gamma(n)$  e entre  $x_{\gamma_1}(n)$  e  $x_{\gamma_2}(n)$  (com  $\gamma_2 \neq \gamma_1$ ) são efetuadas através de fórmulas recursivas, o que será abordado a seguir.

## 2.7 Cálculo do Cepstrum Generalizado

Uma vez obtido o cepstrum complexo  $x_0(n)$  pelo critério de fase mínima (contemplado por (2.66)), é possível obter a seqüência de fase mínima  $x(n)$ . Uma seqüência de fase mínima é causal assim como o cepstrum generalizado a ela associado (como visto na seção 2.6.1). A causalidade de  $x(n)$  e  $x_\gamma(n)$  levada em conta na relação implícita (2.50) leva a fórmulas recursivas que permitem a obtenção de  $x_\gamma(n)$  a partir de  $x(n)$  e vice-versa (para  $n \geq 1$ ). O mesmo vale para dois cepstrums-gen  $x_{\gamma_2}(n)$  e  $x_{\gamma_1}(n)$  (com  $\gamma_2 \neq \gamma_1$ ) que, quando causais, conduzem a relação implícita (2.53) a uma fórmula recursiva para o cálculo de  $x_{\gamma_2}(n)$  a partir de  $x_{\gamma_1}(n)$  para  $n \geq 1$ . Já as amostras  $x(0)$  e  $x_\gamma(0)$  podem ser obtidas com base na relação (2.23), válida para  $x(n)$  e  $x_\gamma(n)$

causais. O mesmo se aplica a  $x_{\gamma_2}(0)$  e  $x_{\gamma_1}(0)$ , obtidas por uma extensão da relação (2.23). Os cálculos devem se restringir a cepstrums-gen vinculados a uma seqüência de fase mínima. Isso garante que as seqüências envolvidas no cálculo sejam mutuamente causais, condição para sua inter-relação recursiva.

### 2.7.1 Cálculo de $x(n)$ a partir de $x_\gamma(n)$

Com  $x'_\gamma(n)$  representando o cepstrum generalizado normalizado, relacionado a  $x_\gamma(n)$  por (2.34), a seqüência de fase mínima normalizada  $x'(n)$  é calculada, a partir de  $x'_\gamma(n)$ , através da fórmula recursiva

$$x'(n) = x'_\gamma(n) + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \cdot x'_\gamma(k) \cdot x'(n-k) - \gamma \cdot \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \cdot x'(k) \cdot x'_\gamma(n-k), \quad (2.68)$$

$$n > 0$$

A dedução de tal fórmula encontra-se no final deste tópico. A amostra inicial  $x(0)$  da seqüência de fase mínima é obtida, a partir de  $x_\gamma(0)$ , por

$$x(0) = \exp_\gamma \left( x_\gamma(0) \right) \quad (\text{expressão (2.23)})$$

e, com base em (2.28), a seqüência de fase mínima  $x(n)$ , para  $n > 0$ , é obtida, a partir de  $x'(n)$  por

$$x(n) = x(0) \cdot x'(n) \quad (2.69)$$

A seqüência de fase mínima é calculada a partir do cepstrum complexo  $x_0(n)$  pela fórmula

$$x'(n) = x_0(n) + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \cdot x_0(k) \cdot x'(n-k), \quad (2.70)$$

$$n > 0$$

obtida adotando  $\gamma = 0$  na fórmula (2.68).

#### Dedução da fórmula recursiva (2.68)

Assumindo  $x(n) = 0$  e  $x_\gamma(n) = 0$  para  $n < 0$  na relação implícita (2.50) e rearranjando seus termos, obtém-se

$$x(n) = \sum_{k=1}^n \binom{k}{n} \cdot x_\gamma(k) \cdot x(n-k) - \gamma \cdot \sum_{k=1}^n \binom{k}{n} \cdot x(k) \cdot x_\gamma(n-k) \quad (2.71)$$

Colocando os termos  $k = n$  para fora dos somatórios, e rearranjando-os, obtém-se

$$\begin{aligned} x(n) + \gamma \cdot x(n) \cdot x_\gamma(0) &= \\ &= x_\gamma(n) \cdot x(0) + \sum_{k=1}^{n-1} \binom{k}{n} \cdot x_\gamma(k) \cdot x(n-k) - \gamma \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot x(k) \cdot x_\gamma(n-k) \end{aligned} \quad (2.72)$$

Dividindo a expressão por  $[(1 + \gamma \cdot x_\gamma(0)) \cdot x(0)]$ , obtém-se

$$\frac{x(n)}{x(0)} = \frac{x_\gamma(n)}{1 + \gamma \cdot x_\gamma(0)} + \sum_{k=1}^{n-1} \binom{k}{n} \cdot \frac{x_\gamma(k)}{1 + \gamma \cdot x_\gamma(0)} \cdot \frac{x(n-k)}{x(0)} - \gamma \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot \frac{x(k)}{x(0)} \cdot \frac{x_\gamma(n-k)}{1 + \gamma \cdot x_\gamma(0)} \quad (2.73)$$

e, com base em (2.28) e (2.34), a fórmula (2.68) é obtida.

## 2.7.2 Cálculo de $x_\gamma(n)$ a partir de $x(n)$

O cepstrum generalizado normalizado  $x'_\gamma(n)$  é calculado a partir da seqüência de fase mínima normalizada  $x'(n)$  através da fórmula recursiva

$$\begin{aligned} x'_\gamma(n) &= x'(n) + \gamma \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot x'(k) \cdot x'_\gamma(n-k) - \sum_{k=1}^{n-1} \binom{k}{n} \cdot x'_\gamma(k) \cdot x'(n-k), \\ &n > 0 \end{aligned} \quad (2.74)$$

Tal fórmula é obtida por um simples rearranjo de termos em (2.68). A amostra inicial do cepstrum generalizado é obtida, a partir de  $x(0)$ , por

$$x_\gamma(0) = \ln_\gamma(x(0)) \quad (2.75)$$

relação obtida aplicando-se a função log-gen à expressão (2.23). Com base em (2.34), o cepstrum generalizado  $x_\gamma(n)$ , para  $n > 0$ , é obtido, a partir de  $x'_\gamma(n)$  por

$$x_\gamma(n) = (1 + \gamma \cdot x_\gamma(0)) \cdot x'_\gamma(n) \quad (2.76)$$

O cepstrum complexo  $x_0(n)$  é calculado a partir da seqüência de fase mínima pela fórmula

$$x_0(n) = x'(n) - \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \cdot x_0(k) \cdot x'(n-k), \quad (2.77)$$

$$n > 0$$

obtida adotando  $\gamma = 0$  na fórmula (2.74).

### 2.7.3 Cálculo de $x_{\gamma_2}(n)$ a partir de $x_{\gamma_1}(n)$

Cepstrums generalizados com diferentes valores de  $\gamma$ , associados a uma mesma seqüência de fase mínima, podem ser obtidos diretamente um a partir do outro. Com  $\gamma_2 \neq \gamma_1$ , o cepstrum-gen normalizado  $x'_{\gamma_2}(n)$  pode ser obtido a partir do cepstrum-gen normalizado  $x'_{\gamma_1}(n)$  através da fórmula recursiva

$$x'_{\gamma_2}(n) = x'_{\gamma_1}(n) + \gamma_2 \cdot \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \cdot x'_{\gamma_1}(k) \cdot x'_{\gamma_2}(n-k) - \gamma_1 \cdot \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) \cdot x'_{\gamma_2}(k) \cdot x'_{\gamma_1}(n-k) \quad (2.78)$$

$$n > 0$$

A dedução de tal fórmula encontra-se no final deste tópico. A amostra inicial  $x_{\gamma_2}(0)$  é obtida a partir de  $x_{\gamma_1}(0)$  por

$$x_{\gamma_2}(0) = \ln_{\gamma_2} \left( \exp_{\gamma_1} \left( x_{\gamma_1}(0) \right) \right) \quad (2.79)$$

relação obtida substituindo  $x(0) = \exp_{\gamma_1} \left( x_{\gamma_1}(0) \right)$  (relação (2.23)) em  $x_{\gamma_2}(0) = \ln_{\gamma_2} \left( x(0) \right)$  (relação (2.75)). Com base em (2.34), o cepstrum generalizado  $x_{\gamma_2}(n)$ , para  $n > 0$ , é obtido, a partir de  $x'_{\gamma_2}(n)$  por

$$x_{\gamma_2}(n) = \left( 1 + \gamma_2 \cdot x_{\gamma_2}(0) \right) \cdot x'_{\gamma_2}(n) \quad (2.80)$$

A fórmula recursiva (2.78) será representada como uma transformação. Algumas de suas propriedades serão abordadas no próximo tópico.

#### Dedução da fórmula recursiva (2.78)

Assumindo  $x_{\gamma_2}(n) = 0$  e  $x_{\gamma_1}(n) = 0$  para  $n < 0$  na relação implícita (2.53) e rearranjando seus termos, obtém-se

$$x_{\gamma_2}(n) = x_{\gamma_1}(n) + \gamma_2 \cdot \sum_{k=1}^n \left(\frac{k}{n}\right) \cdot x_{\gamma_1}(k) \cdot x_{\gamma_2}(n-k) - \gamma_1 \cdot \sum_{k=1}^n \left(\frac{k}{n}\right) \cdot x_{\gamma_2}(k) \cdot x_{\gamma_1}(n-k) \quad (2.81)$$

Colocando os termos  $k = n$  para fora dos somatórios, e rearranjando-os, obtém-se

$$\begin{aligned}
& x_{\gamma_2}(n) + \gamma_1 \cdot x_{\gamma_2}(n) \cdot x_{\gamma_1}(0) = \\
& = x_{\gamma_1}(n) + \gamma_2 \cdot x_{\gamma_1}(n) \cdot x_{\gamma_2}(0) + \gamma_2 \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot x_{\gamma_1}(k) \cdot x_{\gamma_2}(n-k) - \gamma_1 \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot x_{\gamma_2}(k) \cdot x_{\gamma_1}(n-k)
\end{aligned} \tag{2.82}$$

Fatorando os termos de fora dos somatórios, obtém-se

$$\begin{aligned}
& x_{\gamma_2}(n) \cdot (1 + \gamma_1 \cdot x_{\gamma_1}(0)) = \\
& = x_{\gamma_1}(n) \cdot (1 + \gamma_2 \cdot x_{\gamma_2}(0)) + \gamma_2 \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot x_{\gamma_1}(k) \cdot x_{\gamma_2}(n-k) - \gamma_1 \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot x_{\gamma_2}(k) \cdot x_{\gamma_1}(n-k)
\end{aligned} \tag{2.83}$$

Dividindo a expressão por  $[(1 + \gamma_1 \cdot x_{\gamma_1}(0)) \cdot (1 + \gamma_2 \cdot x_{\gamma_2}(0))]$ , obtém-se

$$\begin{aligned}
\frac{x_{\gamma_2}(n)}{1 + \gamma_2 \cdot x_{\gamma_2}(0)} &= \frac{x_{\gamma_1}(n)}{1 + \gamma_1 \cdot x_{\gamma_1}(0)} \\
&+ \gamma_2 \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot \frac{x_{\gamma_1}(k)}{1 + \gamma_1 \cdot x_{\gamma_1}(0)} \cdot \frac{x_{\gamma_2}(n-k)}{1 + \gamma_2 \cdot x_{\gamma_2}(0)} \\
&- \gamma_1 \cdot \sum_{k=1}^{n-1} \binom{k}{n} \cdot \frac{x_{\gamma_2}(k)}{1 + \gamma_2 \cdot x_{\gamma_2}(0)} \cdot \frac{x_{\gamma_1}(n-k)}{1 + \gamma_1 \cdot x_{\gamma_1}(0)}
\end{aligned} \tag{2.84}$$

e, com base em (2.34), a fórmula (2.78) é obtida.

## 2.7.4 Transformação Logarítmica Generalizada e Suavização Espectral

A fórmula recursiva (2.78) que representa o cálculo do cepstrum generalizado  $x'_{\gamma_2}(n)$  com parâmetro  $\gamma_2$  a partir de outro  $x'_{\gamma_1}(n)$  com parâmetro  $\gamma_1$ , satisfaz  $\exp_{\gamma_2}(X_{\gamma_2}(z)) = \exp_{\gamma_1}(X_{\gamma_1}(z)) = X(z)$  assumindo, a priori, que  $x'_{\gamma_2}(n)$  e  $x'_{\gamma_1}(n)$  são seqüências de comprimento infinito. Uma vez que na prática não se dispõe de seqüências infinitas, serão assumidos comprimentos finitos para a discussão a seguir.

A fórmula recursiva (2.78) corresponde a uma transformação que leva o cepstrum-gen normalizado  $x'_{\gamma_1}(n)$  ao cepstrum-gen normalizado  $x'_{\gamma_2}(n)$  e será expressa por

$$x'_{\gamma_2, N_2} = T_{cg}(x'_{\gamma_1, N_1}, N_1, N_2, \gamma_1, \gamma_2) \tag{2.85}$$

onde  $\mathbf{x}'_{\gamma_1, N_1} = [x'_{\gamma_1}(1), x'_{\gamma_1}(2), \dots, x'_{\gamma_1}(N_1)]^T$ ,  $\mathbf{x}'_{\gamma_2, N_2} = [x'_{\gamma_2}(1), x'_{\gamma_2}(2), \dots, x'_{\gamma_2}(N_2)]^T$ .  $N_1$  e  $N_2$  são números inteiros positivos que representam os comprimentos assumidos para os cepstrums generalizados normalizados  $x'_{\gamma_1}(n)$  e  $x'_{\gamma_2}(n)$ . A transformação se restringe a seqüências de fase mínima reais, e cepstrums-gen obtidos sob o critério de fase mínima (ver seqüências 2.6.2 e 2.7).

Algumas propriedades relacionadas aos comprimentos das seqüências envolvidas na transformação (2.85) serão enunciadas com base em seu objetivo:

- **Suavização Espectral:** Quando se visa uma representação suavizada do espectro da seqüência de entrada, restringe-se o comprimento da seqüência transformada a um valor, em geral, baixo em relação àquele da seqüência original (isso corresponde a uma truncagem da seqüência de saída). Porém para o cálculo do cepstrum-gen são exigidas para a entrada somente o número de amostras da saída. Assim, desejando-se obter uma seqüência transformada de comprimento  $N_2 = M$  a partir de uma seqüência de comprimento  $N > M$ , são necessárias somente  $N_1 = M$  amostras da seqüência de entrada na transformação (2.85).
- **Representação Espectral Completa:** Quando se deseja uma representação completa do espectro da seqüência original através da seqüência transformada, a última deve ser extensa o suficiente. Considerando uma seqüência de entrada com comprimento  $N_1 = M$ , devem ser calculadas  $N_2 \gg M$  amostras da seqüência de saída na transformação (2.85).

A seqüência 2.8 aborda a suavização espectral com base no cepstrum generalizado.

### 2.7.5 Computação do Cepstrum Generalizado

Esta seqüência aborda a computação do cepstrum generalizado associado à seqüência de fase mínima  $x(n)$ . Normalmente os cálculos do cepstrum-gen são baseados em uma seqüência causal finita e de fase genérica. Denotando tal seqüência por  $\bar{x}(n)$ , a computação de seu espectro é obtida de forma discreta pela DFT



$$\bar{X}(k) = \sum_{n=0}^{N-1} \bar{x}(n) e^{-j\left(\frac{2\pi}{N}\right)kn} \quad (2.86)$$

onde  $N$ , o número de amostras da DFT, deve ser igual ou maior que o número de amostras da seqüência  $\bar{x}(n)$ . Caso seja adotado  $N$  maior do que o comprimento de  $\bar{x}(n)$ , a seqüência deve ser preenchida com amostras nulas de maneira a totalizar  $N$  amostras. Para o cálculo do cepstrum ( $\gamma = 0$ ) associado à seqüência de fase mínima, o cepstrum real  $x_{0p}(n)$ , dado por (2.65), é computado pela DFT inversa

$$x_{0p}(n) = \sum_{k=0}^{N-1} \ln(|\bar{X}(k)|) e^{j\left(\frac{2\pi}{N}\right)kn} \quad (2.87)$$

e o cepstrum complexo  $x_0(n)$ , associado à seqüência de fase mínima  $x(n)$ , é obtido por (2.66). Considerando que o cepstrum  $x_{0p}(n)$  é, em geral, uma seqüência de comprimento infinito e  $\ln(|\bar{X}(k)|)$  corresponde a um período da série de Fourier de uma seqüência dada pela repetição periódica de  $x_{0p}(n)$  com período  $N$ , a fórmula (2.87) fornece uma versão de  $x_{0p}(n)$  com *aliasing*. Entretanto  $|x_{0p}(n)|$  tem decaimento no mínimo tão veloz quanto  $1/n$  [17]. Isso permite que as distorções por *aliasing* temporal sejam minimizadas com  $N$  suficientemente grande.

Uma vez obtido o cepstrum complexo  $x_0(n)$ , qualquer seqüência, seja ela o cepstrum-gen  $x_\gamma(n)$  ou a seqüência de fase mínima  $x(n)$ , pode ser computada com base na transformação (2.85). As representações dos cálculos de  $x(n)$  e  $x_\gamma(n)$  serão feitas com base na notação dada por (2.85). A representação genérica para a computação de uma seqüência a partir de outras é dada por

$$x_{\gamma_2}(n) = \begin{cases} \ln_{\gamma_2} \left( \exp_{\gamma_1} \left( x_{\gamma_1}(0) \right) \right), & n = 0 \\ \left[ 1 + \gamma_2 \ln_{\gamma_2} \left( \exp_{\gamma_1} \left( x_{\gamma_1}(0) \right) \right) \right] \cdot \left( T_{cg} \left( \mathbf{x}'_{\gamma_1, N_1}, N_1, N_2, \gamma_1, \gamma_2 \right) \right)_n & n > 0 \end{cases} \quad (2.88)$$

onde  $\mathbf{x}'_{\gamma_1, N_1} = [x'_{\gamma_1}(1), x'_{\gamma_1}(2), \dots, x'_{\gamma_1}(N_1)]^T$  e  $(\cdot)_n$  fornece a  $n$ -ésima componente de um vetor, de maneira que  $\left( T_{cg} \left( \mathbf{x}'_{\gamma_1, N_1}, N_1, N_2, \gamma_1, \gamma_2 \right) \right)_n = \left( \mathbf{x}'_{\gamma_2, N_2} \right)_n = x'_{\gamma_2}(n)$ . A seqüência de fase mínima  $x(n)$  pode ser calculada a partir do cepstrum-gen, substituindo  $\gamma_1 = \gamma$  e  $\gamma_2 = 1$  em (2.88), assim

$$x(n) = \begin{cases} \exp(x_\gamma(0)) - 1, & n = 0 \\ \exp(x_\gamma(0)) \cdot (T_{cg}(x'_{\gamma, N_1}, N_1, N_2, \gamma, 1))_n & n > 0 \end{cases} \quad (2.89)$$

onde com  $\gamma = 0$ , computa-se a seqüência de fase mínima  $x(n)$  a partir do cepstrum  $x_0(n)$ . Já o cepstrum-gen pode ser calculado a partir da seqüência de fase mínima considerando  $\gamma_1 = 1$  e  $\gamma_2 = 0$  em (2.88), o que fornece

$$x_\gamma(n) = \begin{cases} \ln_\gamma(1 + x(0)), & n = 0 \\ [1 + \ln_\gamma(1 + x(0))] \cdot (T_{cg}(x'_{\gamma, N_1}, N_1, N_2, 1, \gamma))_n & n > 0 \end{cases} \quad (2.90)$$

onde  $x'_{\gamma, N_1} = [x'(1), x'(2), \dots, x'(N_1)]^T$ , cujas componentes são dadas por  $x'(k) = x(k)/x(0)$ .

## 2.8 Suavização Espectral

A aplicação de um *janelamento passa-baixas* ao cepstrum generalizado (filtragem linear passa-baixas no domínio cepstral generalizado) conduz a versões suavizadas do espectro original com características que variam de acordo com  $\gamma$ , parâmetro associado à função logarítmica generalizada (log-gen), que pode assumir valores no intervalo contínuo  $[-1, 1]$ .

A função log-gen pode ser considerada a base de um sistema que unifica o sistema de deconvolução de raiz espectral [16] (SRDS, *Spectral Root Deconvolution System*), que tem base em potências fracionárias do espectro, e o sistema de deconvolução homomórfica logarítmica [17] (LHDS, *Logarithmic Homomorphic Deconvolution System*) que, por sua vez, tem base na função logarítmica. Tal sistema unificado pode ser considerado um *sistema de deconvolução homomórfica logarítmica generalizada* (GLHDS – *Generalized Logarithmic Homomorphic Deconvolution System*), e será assim referido no texto. Uma operação de deconvolução pelo GLHDS produz resultados equivalentes àqueles obtidos na deconvolução pelo SRDS, porém com a característica de que, para  $\gamma = 0$ , uma vez que a função log-gen se reduz à função logarítmica e, por extensão, o cepstrum-gen se reduz ao cepstrum, o GLHDS se reduz ao LHDS. O GLHDS, portanto, é um sistema que unifica o LHDS e o SRDS. Levando em conta a

conexão entre o GLHDS e LHDS, serão utilizados para o cepstrum-gen, alguns dos termos normalmente associados ao cepstrum. O domínio cepstral generalizado será denominado domínio da *quefrência* (que assume unidades de tempo). Um janelamento do cepstrum-gen corresponderá a uma operação de *liftragem* do espectro, que é efetuada no domínio da *quefrência* e é assim denominada por sua analogia à filtragem de um sinal (efetuada no domínio da frequência).

### 2.8.1 Suavização do Espectro log-gen

O espectro logarítmico generalizado (log-gen) corresponde ao espectro tomado na escala não-linear definida pela função log-gen. A função log-gen de uma variável real, conforme o decréscimo do valor de  $\gamma$ , varia produzindo desde uma curva linear ( $\gamma = 1$ ), passando por uma curva logarítmica ( $\gamma = 0$ ), até a curva log-gen associada a  $\gamma = -1$  (ver Figura 2.1) que, no intervalo de interesse, tem as características da anterior mais acentuadas. Como resultado da aplicação da função log-gen ao espectro, o intervalo de variação da amplitude espectral fica tão menor quanto maior a amplitude do espectro regular, característica que se acentua com o decréscimo de  $\gamma$  (exemplos na Figura 2.3). Isso acarreta uma série de efeitos sobre diferentes elementos do espectro regular, merecendo atenção o efeito de *compressão* vertical relativa dos picos de ressonância, o que os leva a se manifestarem em variações mais amenas na curva do espectro log-gen do que na curva original. Tal efeito pode também ser interpretado como um aumento da largura de banda dos picos. Esse conjunto de características, já esperado quando se visualiza uma curva de espectro na escala logarítmica, tem aqui o diferencial de ter intensidade variável com  $\gamma$ . A aplicação de uma *liftragem* passa-baixas ao cepstrum-gen (suavização do espectro log-gen), ao selecionar somente parciais (senóides) de oscilação lenta, tende a manter bem descritas variações lentas da curva do espectro log-gen, não acompanhando variações bruscas e oscilações mais velozes. Tal propriedade corresponde a uma redução na resolução espectral, já que o corte de amostras do cepstrum-gen via *liftragem* corresponde a um aumento na largura do lóbulo central do núcleo de convolução associado ao comprimento da seqüência. Assim, após a *liftragem*, além da esperada remoção da maior parte da estrutura fina do espectro, os picos de ressonância ficam tão melhor descritos quanto menor o valor de  $\gamma$ . Isso, porém, ganha restrições uma vez que a *liftragem* faz com que elementos como vales e estrutura fina, por dominarem o espectro log-gen em amplitude para valores mais baixos de  $\gamma$ ,

intensifiquem efeitos de vazamento espectral, impactando negativamente a descrição de picos de ressonância. Assim, a escolha do valor de  $\gamma$  para uma melhor descrição de picos deve ser ponderada com base em tal compromisso.

A mesma consideração feita para os picos de ressonância pode ser feita para os vales espectrais. Na escala log-gen, a *compressão* vertical dos picos equivale a uma *expansão* vertical relativa dos vales, efeito reduzido com o aumento de  $\gamma$ . Como resultado de um incremento em  $\gamma$ , os vales da curva do espectro log-gen se manifestam em variações mais lentas, enquanto os picos de ressonância se tornam mais estreitos e proeminentes. Quando o parâmetro  $\gamma$  atinge seu limite superior ( $\gamma = 1$ ), obtém-se o espectro na escala linear. Assim, analogamente à boa descrição esperada para os picos ante a *lifragem*, os vales ficam tão melhor descritos quanto maior o valor de  $\gamma$ . Em contrapartida, com o aumento do valor de  $\gamma$ , o aumento na proeminência dos picos de ressonância de maior amplitude no espectro log-gen faz com que, na *lifragem*, sejam intensificados efeitos de vazamento espectral, o que leva a distorções na descrição dos vales. A Figura 2.4 (a) (c) e (e) mostra exemplos das curvas do espectro log-gen suavizado. A seção a seguir aborda a suavização do espectro log-gen considerando o domínio da *quefrência*.

## 2.8.2 Deconvolução Homomórfica log-gen

No contexto do espectro logarítmico ( $\gamma = 0$ ), para um sinal de fala vozeado, a estrutura fina do espectro decorre basicamente da periodicidade dos pulsos de ar (provenientes da glote) que excitam o trato vocal, e sua contribuição ao cepstrum ocorre na forma de impulsos de mesmo período. Já as características de ressonância do trato, por se manifestarem em variações lentas no espectro, têm sua contribuição concentrada em baixas *quefrências*, decaindo rapidamente. Assim, se o período dos pulsos for grande o bastante, não ocorre sobreposição entre as contribuições, o que permite que elas sejam separadas via *lifragem*. Essa é a base do LHDS. A diferença do cepstrum-gen ( $\gamma \neq 0$ ) para o cepstrum ( $\gamma = 0$ ), é que, no cepstrum-gen, a contribuição do trato aparece convoluída com a contribuição dos pulsos, sendo periodicamente replicada. Isso faz com que o GLHDS (com  $\gamma \neq 0$ ) não seja um sistema adequado para recuperar a contribuição dos pulsos. A contribuição do trato, por sua vez, pode ser recuperada na condição de que tenha decaimento veloz o suficiente para não se sobrepor às suas

réplicas. A velocidade do decaimento depende da escolha do valor de  $\gamma$ . Quanto mais veloz o decaimento, mais concentrada em *quefrência* está a informação do trato e melhor *sintonizado* está o valor de  $\gamma$  ao espectro em análise [16], já que a informação pode ser bem selecionada pela *lifragem*. Essa condição equivale àquela que minimiza os efeitos de vazamento espectral. O vazamento espectral é conseqüência da queda brusca de valores do cepstrum-gen em decorrência do corte efetuado na *lifragem*. Tal efeito torna-se reduzido quando o cepstrum-gen, na *quefrência* de corte, já se encontra com valores baixos em função de um decaimento rápido. As ressonâncias de sinais de fala não-vozeados, por terem banda larga, têm sua melhor especificação na operação de *lifragem* do cepstrum ( $\gamma = 0$ ), já que este não privilegia picos ou vales do espectro na escala logarítmica (dB). O cepstrum-gen tem seu melhor uso com base em sinais que possam ser representados pela convolução de um trem de impulsos periódico com a resposta de um sistema ressonante.

### **2.8.3 Envoltória Espectral Característica e Remoção de Viés Estatístico**

A aplicação da função log-gen inversa (função exponencial generalizada) ao espectro log-gen suavizado o reconduz à escala linear, fornecendo a envoltória do espectro original, cuja curva tem perfil característico do valor de  $\gamma$  e, portanto, é aqui denominada *envoltória espectral característica*. As considerações feitas permitem concluir que uma envoltória de qualidade aceitável só é obtida para um intervalo de  $\gamma$  que a previna das distorções mais intensas, decorrentes de vazamento espectral. Dentro desse intervalo, o critério para o valor de  $\gamma$  pode provir, por exemplo, da escolha por uma representação que privilegie picos em detrimento de vales, ou vice-versa. O espectro representado pela exponencial generalizada, conduz a uma expressão cuja forma varia continuamente com  $\gamma$  e, para alguns de seus valores, remetem a modelos conhecidos, dentre eles, o modelo de zeros (*all-zero*) ( $\gamma = 1$ ), o modelo de pólos (*all-pole*) ( $\gamma = -1$ ) e o modelo exponencial ( $\gamma = 0$ ), que pode ser considerado um modelo de pólos e zeros (*pole-zero model*) (ver seção 2.4). A envoltória espectral característica, portanto, assume perfis equivalentes àqueles associados a tais modelos. Nesse âmbito, o cepstrum-gen truncado (*liftrado*) é um conjunto de coeficientes de um sistema linear cuja reposta em frequência é dada pela exponencial generalizada do espectro log-gen

suavizado, podendo ser usada para uma ressíntese do sinal original [20]. A Figura 4.3 (b) (d) e (f) mostra as envoltórias espectrais características associadas, respectivamente, às curvas suavizadas (a) (c) e (e) (da mesma figura).

A *liftragem* do cepstrum generalizado é uma técnica linear que, independentemente de uma escolha adequada para o valor de  $\gamma$ , produz efeitos de vazamento espectral. Tais distorções levam a perturbações quadro a quadro decorrentes, principalmente, das flutuações das quedas profundas da estrutura fina, e da variação na frequência fundamental. Esta última implica em variações na distância em *quefrência* entre a informação do trato e a primeira réplica dela (ou o primeiro *pulso*, no caso  $\gamma = 0$ ) e, como as contribuições não são idealmente separadas entre si, a *liftragem* seleciona informação indesejável em maior ou menor grau quadro a quadro, o que contribui para as perturbações. O problema do vazamento espectral é traduzido no âmbito da estimação estatística de parâmetros, como um problema de viés (*bias*) estatístico. A aplicação de uma condição que remove o viés do espectro logarítmico suavizado leva a uma envoltória espectral estável e satisfatória, podendo proporcionar uma ressíntese de qualidade superior. Tal técnica corresponde a uma suavização não-linear do espectro de fala. No contexto do presente capítulo, sua aplicação é interpretada como um ajuste de valores do cepstrum-gen suavizado para a remoção do viés. Entretanto, para valores de  $\gamma$  no intervalo  $[-1,0]$ , o método tem solução única e independe dos valores do cepstrum-gen linearmente suavizado. O lugar deste é assumido por incógnitas a serem determinadas pelo método (o uso do cepstrum-gen como valor inicial reduz o esforço computacional para a solução). Nesse contexto, a forma funcional obtida pela exponencial generalizada do espectro log-gen com  $N$  coeficientes arbitrários, corresponde à expressão para um modelo paramétrico denominado modelo cepstral generalizado (GC, *generalized cepstral*) de ordem  $N$ . Os parâmetros a serem estimados pela técnica correspondem aos coeficientes cepstrais generalizados (GC). Um caso especial do modelo GC é o modelo de pólos (*all-pole*), obtido com  $\gamma = -1$ , quando a solução para o método leva aos coeficientes LPC (*Linear Predictive Coding*). O capítulo 4 trará uma abordagem detalhada do modelo cepstral generalizado e do modelo mel-cepstral generalizado (MGC, *mel-generalized cepstral*).

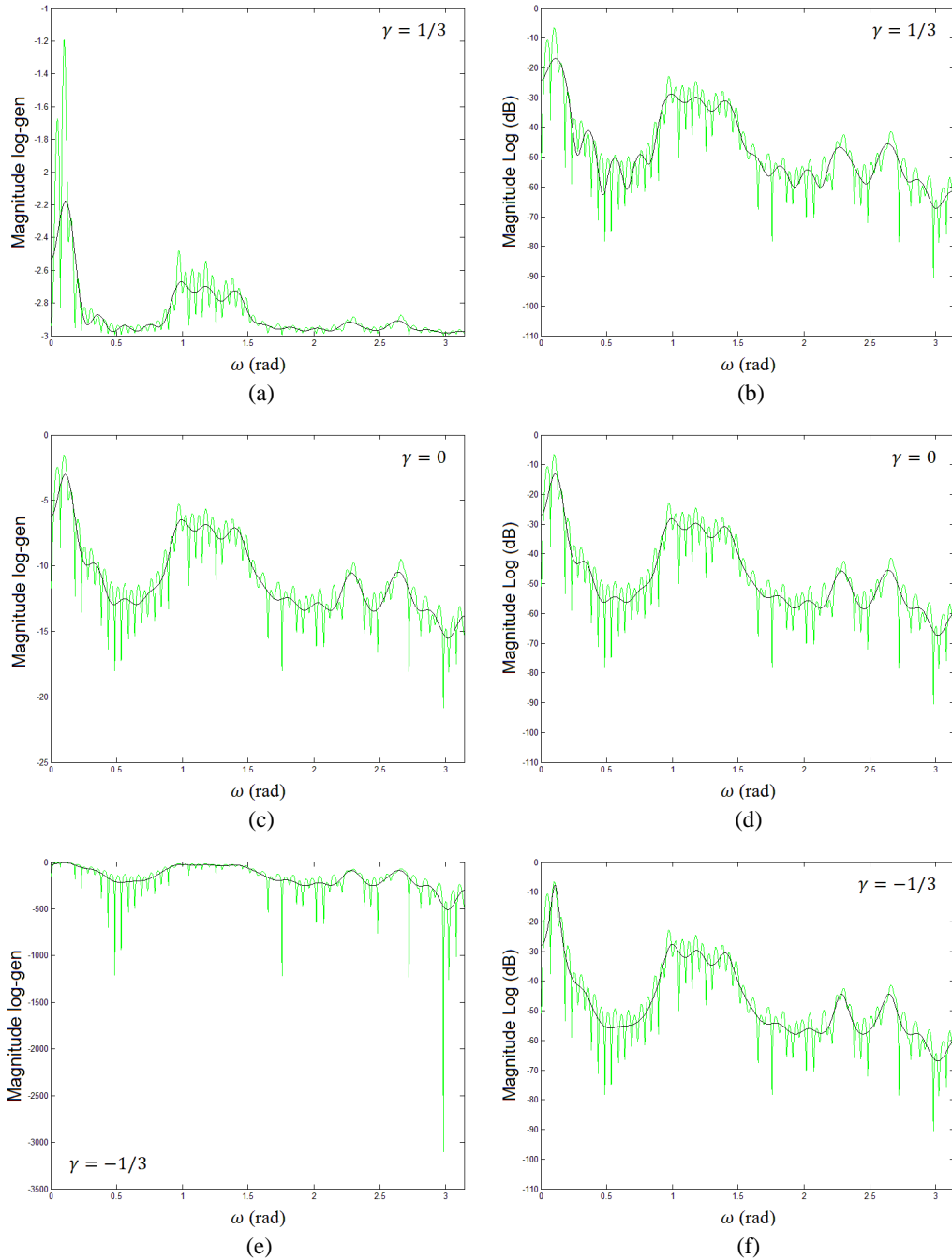


Figura 2.4: (a) (c) (e) Espectros log-gen suavizados (ordem 30) com base nos espectros log-gen associados a  $\gamma = 1/3$ ,  $\gamma = 0$  e  $\gamma = -1/3$  (em verde). (b) (d) (f) Magnitudes (dB) das envoltórias espectrais características associadas a  $\gamma = 1/3$ ,  $\gamma = 0$  e  $\gamma = -1/3$  sobrepostas ao espectro original (em verde). *Observação: a magnitude log-gen é expressa por  $\ln_\gamma(|\cdot|^2)$ .*

## Capítulo 3

# Seqüências Distorcidas em Frequência e o Mel-Cepstrum Generalizado

### 3.1 Introdução

O capítulo anterior abordou o cepstrum generalizado, seqüência representada no espectro tomado na escala logarítmica generalizada de amplitudes. O presente capítulo trata de seqüências representadas em espectros tomados na escala não-linear de frequências definida por uma transformação bilinear (ou *passa-tudo*) de primeira ordem. Tal transformação, que depende de um parâmetro  $\alpha$  que assume valores no intervalo contínuo  $] -1, 1[$ , produz uma gama contínua de diferentes curvas (ver exemplos na Figura 3.1). Seu uso permite a derivação de fórmulas recursivas para a transformação frequencial de seqüências, que é aqui abordada, por estender o cepstrum generalizado ao mel-cepstrum generalizado, seqüência representada no espectro tomado mutuamente nas escalas log-gen de amplitude e na escala de frequências, na qual se baseia a análise Mel-Cepstral Generalizada.

Com base na transformação bilinear de primeira ordem, adotada para a distorção do eixo de frequências, deriva-se uma fórmula recursiva para o cálculo de seqüências distorcidas em frequência [21]. Sua aplicação corresponde à realização de uma filtragem em cascata que produz, a cada etapa, uma amostra da seqüência distorcida em



freqüência. Tal cálculo corresponde a uma transformação em freqüência e suas propriedades são tratadas ante o objetivo de suavização e representação espectral.

A suavização do espectro distorcido em freqüência condiz com uma análise espectral com resolução não-uniforme e pode ser efetuada através de uma truncagem da seqüência distorcida em freqüência. Isso equivale a uma análise efetuada sobre a seqüência original por uma janela temporal variável conforme a freqüência. O cálculo de um valor interpretado como um *comprimento* aproximado de tal janela permite estabelecer critérios para os comprimentos das seqüências envolvidas na transformação em freqüência. Uma atribuição adequada ao valor do parâmetro  $\alpha$  aproxima a escala mel de freqüência perceptual, levando à incorporação, pelo espectro, da resolução frequencial auditiva aproximada.

Aborda-se o cálculo do mel-cepstrum generalizado, efetuado por uma combinação das transformações frequencial e log-gen e a suavização do espectro mel-log-gen que é tratada como uma extensão à abordagem dada ao cepstrum na seção 2.8.

## 3.2 Espectro Distorcido em Freqüência

### 3.2.1 Seqüências Distorcidas em Freqüência

Seja  $x(n)$  uma seqüência conhecida real e estável e,  $\tilde{x}(n)$ , uma seqüência real e estável relacionada a  $x(n)$  com base no que será exposto a seguir.

Considere as transformadas  $z$  de  $x(n)$  e  $\tilde{x}(n)$ , dadas respectivamente por

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (\text{expressão (2.5)})$$

e

$$\tilde{X}(\tilde{z}) = \sum_{n=-\infty}^{\infty} \tilde{x}(n)\tilde{z}^{-n} \quad (3.1)$$

relacionadas por

$$\tilde{X}(\tilde{z}) = X(z(\tilde{z})) \quad (3.2)$$

onde

$$z(\tilde{z}) = \frac{\tilde{z} + \alpha}{1 + \alpha\tilde{z}} \quad (3.3)$$

Tal relação corresponde a uma transformação bilinear de primeira ordem com parâmetro  $\alpha \in ]-1,1[$  real, e tem as seguintes propriedades:

- Com  $z = |z|e^{j\omega}$  e  $\tilde{z} = |\tilde{z}|e^{j\tilde{\omega}}$ , a transformação (3.3) mapeia a CRU  $\tilde{z} = e^{j\tilde{\omega}}$  na CRU  $z = e^{j\omega}$ , e vice-versa (transformação (3.9)), onde é estabelecida uma relação não-linear entre  $\tilde{\omega}$  e  $\omega$ .
- O intervalo  $-\pi < \tilde{\omega} \leq \pi$  é mapeado de um para um no intervalo  $-\pi < \omega \leq \pi$  com  $\omega$  relacionado a  $\tilde{\omega}$  por uma função ímpar, monotonicamente crescente, com o vínculo de que  $\tilde{\omega} = \pi$  é mapeado em  $\omega = \pi$ .
- A região interior  $|\tilde{z}| < 1$  e a região exterior  $|\tilde{z}| > 1$  à CRU do plano  $\tilde{z}$  são, respectivamente, mapeadas na região interior  $|z| < 1$  e na região exterior  $|z| > 1$  à CRU do plano  $z$ .

Avaliada na CRU, a expressão (3.2) tem a forma

$$\tilde{X}(e^{j\tilde{\omega}}) = X(e^{j\omega(\tilde{\omega})}) \quad (3.4)$$

que corresponde à versão distorcida em frequência do espectro  $X(e^{j\omega})$ , com base na relação não-linear entre  $\tilde{\omega}$  e  $\omega$ . A primeira propriedade associada à relação (3.3) implica em que, se  $x(n)$  é estável,  $\tilde{x}(n)$  deve ser estável e vice-versa. A segunda propriedade permite concluir que a simetria de  $X(e^{j\omega})$  é preservada em  $\tilde{X}(e^{j\tilde{\omega}})$  e, portanto, se  $x(n)$  é real,  $\tilde{x}(n)$  também é real e vice-versa.

A seqüência  $\tilde{x}(n)$  corresponde aos coeficientes de Fourier do espectro distorcido em frequência  $\tilde{X}(e^{j\tilde{\omega}})$  e é obtida pela transformada  $z$  inversa de (3.2), dada por

$$\tilde{x}(n) = \frac{1}{2\pi j} \oint_C \tilde{X}(\tilde{z}) \tilde{z}^{n-1} d\tilde{z} \quad (3.5)$$

onde  $C$  é uma curva fechada, tomada no sentido anti-horário, que inclui a origem e se encontra inteiramente na região de convergência (RC) estável do plano  $\tilde{z}$  (ou seja, a RC que inclui a circunferência de raio unitário).

A relação de  $e^{j\omega(\tilde{\omega})}$  com  $e^{j\tilde{\omega}}$  é dada por

$$e^{j\omega(\tilde{\omega})} = \frac{e^{j\tilde{\omega}} + \alpha}{1 + \alpha e^{j\tilde{\omega}}} \quad (3.6)$$

de onde se pode deduzir a relação não-linear entre  $\tilde{\omega}$  e  $\omega$  dada por

$$\omega(\tilde{\omega}) = \tan^{-1} \left[ \frac{(1 - \alpha^2) \sin \tilde{\omega}}{(1 + \alpha^2) \cos \tilde{\omega} + 2\alpha} \right] \quad (3.7)$$

ou, alternativamente, por

$$\omega(\tilde{\omega}) = \tilde{\omega} - 2 \tan^{-1} \left[ \frac{\alpha \sin \tilde{\omega}}{1 + \alpha \cos \tilde{\omega}} \right] \quad (3.8)$$

A Figura 3.1 mostra curvas de  $\omega(\tilde{\omega})$  para diferentes valores de  $\alpha$ .

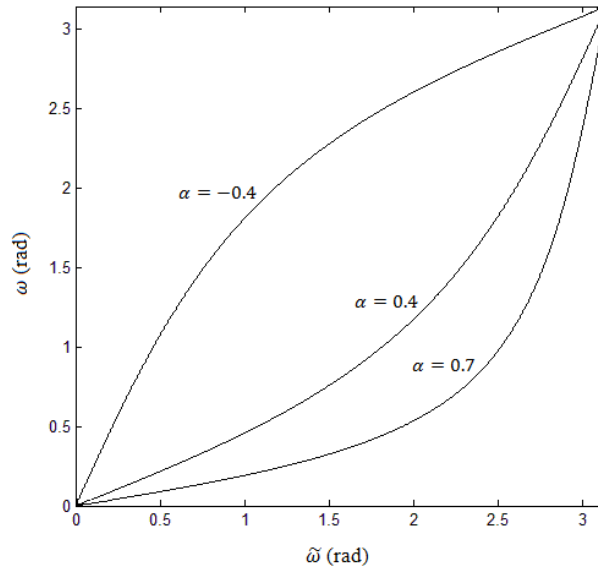


Figura 3.1: Curvas de mapeamento do eixo de frequências para os valores  $\alpha = 0.7$ ,  $\alpha = 0.4$  e  $\alpha = -0.4$ .

A Figura 3.2 (a) mostra o espectro de fala em dB da emissão da vogal “i” por uma voz masculina<sup>§</sup>. A adoção dos eixos não-lineares de frequências (substituição  $\omega = \omega(\tilde{\omega})$ ) com parâmetros  $\alpha = 0.5$  e  $\alpha = -0.5$  (Figura 3.2 (b)) produz as magnitudes espectrais distorcidas em frequência observadas em (c) e (d).

<sup>§</sup> O espectro foi obtido a partir de um segmento de 24ms (384 amostras) de um sinal fala com frequência de amostragem de 16kHz. Adotou-se janela de Blackman.

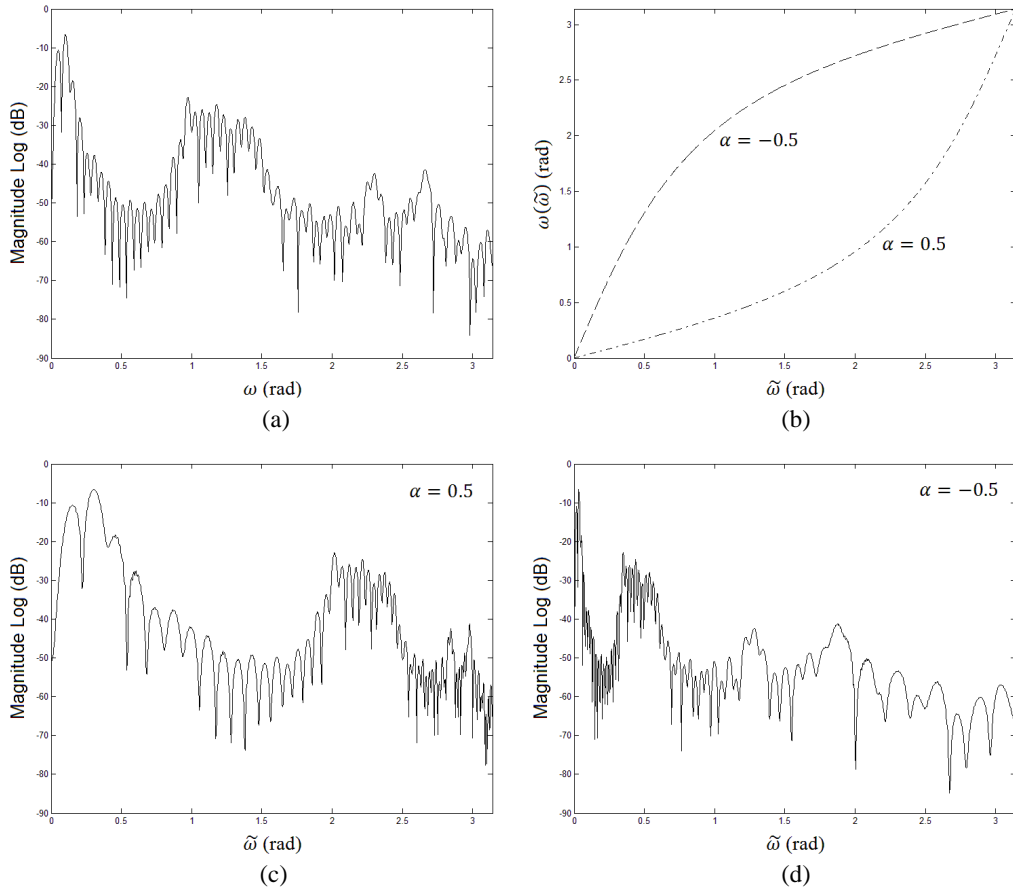


Figura 3.2: (a) Magnitude espectral (dB) de um sinal de fala. (b) Curvas de mapeamento do eixo de frequências com  $\alpha = 0.5$  e  $\alpha = -0.5$ . (c) Magnitude do espectro distorcido em frequência com parâmetro  $\alpha = 0.5$ . (d) Magnitude do espectro distorcido em frequência com parâmetro  $\alpha = -0.5$ .

### 3.3 Representação do Espectro através da Seqüência Distorcida em Frequência

O espectro distorcido em frequência pode ser reconduzido à escala linear de frequências utilizando-se a inversão da expressão (3.3)<sup>§</sup>, dada por

$$\tilde{z}(z) = \frac{z - \alpha}{1 - \alpha z} \quad (3.9)$$

no lugar da variável independente  $\tilde{z}$  na expressão (3.1), assim

<sup>§</sup> Na inversão de (3.3), dada em (3.9),  $z(\tilde{z})$  passa a ter papel de variável independente e é, portanto, expressa por  $z$ .  $\tilde{z}$ , por sua vez, passa a ter papel de variável dependente de  $z$  e é, então, expressa por  $\tilde{z}(z)$ .

$$X(z) = \tilde{X}(\tilde{z}(z)) \quad (3.10)$$

portanto

$$X(z) = \sum_{n=-\infty}^{\infty} \tilde{x}(n)[\tilde{z}(z)]^{-n} \quad (3.11)$$

que corresponde a  $X(z)$  representado através da seqüência distorcida em frequência. Vale observar que a expressão (3.9) é obtida trocando  $\alpha$  por  $-\alpha$  e intercambiando  $z$  e  $\tilde{z}$  na equação (3.3). Avaliada na CRU, a expressão (3.10) é dada por

$$X(e^{j\omega}) = \tilde{X}(e^{j\tilde{\omega}(\omega)}) \quad (3.12)$$

A relação de  $e^{j\tilde{\omega}(\omega)}$  com  $e^{j\omega}$  é dada por

$$e^{j\tilde{\omega}(\omega)} = \frac{e^{j\omega} - \alpha}{1 - \alpha e^{j\omega}} \quad (3.13)$$

de onde se pode extrair a relação não-linear entre  $\tilde{\omega}$  e  $\omega$  dada por

$$\tilde{\omega}(\omega) = \tan^{-1} \left[ \frac{(1 - \alpha^2) \sin \omega}{(1 + \alpha^2) \cos \omega - 2\alpha} \right] \quad (3.14)$$

ou, alternativamente, por

$$\tilde{\omega}(\omega) = \omega + 2 \tan^{-1} \left[ \frac{\alpha \sin \omega}{1 - \alpha \cos \omega} \right] \quad (3.15)$$

A Figura 3.3 mostra curvas de  $\tilde{\omega}(\omega)$  para diferentes valores de  $\alpha$ .

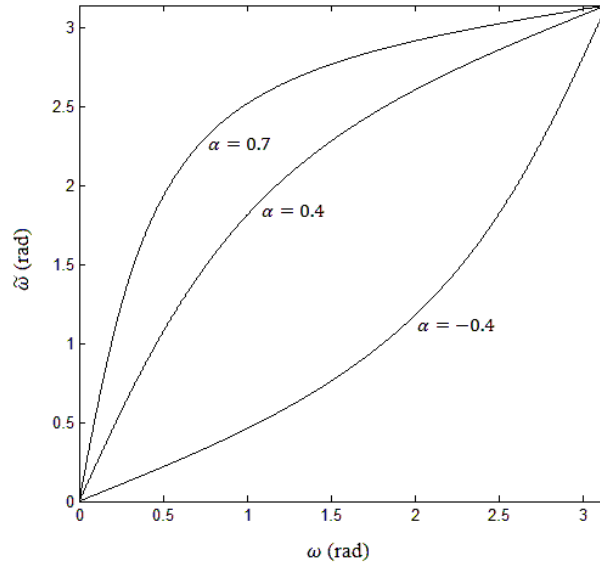


Figura 3.3: Curvas de mapeamento do eixo de frequências para os valores  $\alpha = 0.7$ ,  $\alpha = 0.4$  e  $\alpha = -0.4$ .

## 3.4 Cálculo da Seqüência Distorcida em Frequência

### 3.4.1 Cálculo de $\tilde{x}(n)$ a partir de $x(n)$

Para fins de generalização, até o momento foram assumidas seqüências  $x(n)$  e  $\tilde{x}(n)$  sem restrições quanto à causalidade. Porém, para os propósitos do presente trabalho, a seqüência distorcida em frequência  $\tilde{x}(n)$  deverá estar associada a uma seqüência  $x(n)$  causal. A causalidade de  $x(n)$  implica na causalidade de  $\tilde{x}(n)$ . Uma seqüência causal e estável tem seus pólos todos no interior da CRU. Pode-se deduzir da expressão (3.9), que um ponto no plano  $z$  de módulo  $|z| < 1$  é mapeado no plano  $\tilde{z}$  em um ponto de módulo  $|\tilde{z}| < 1$ . Assim, como  $x(n)$  é causal e tem, portanto, todos os seus pólos no interior da CRU,  $\tilde{x}(n)$  também os tem e é, portanto, causal.

Tomando a expressão (3.5) na CRU, obtém-se

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{X}(e^{j\tilde{\omega}}) e^{j\tilde{\omega}n} d\tilde{\omega} \quad (3.16)$$

Como, em geral não se conhece  $\tilde{X}(e^{j\tilde{\omega}})$ , a seqüência  $\tilde{x}(n)$  não pode ser diretamente obtida por (3.16). Porém, aplicando a substituição de variáveis  $\tilde{\omega} = \tilde{\omega}(\omega)$ , obtém-se

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{X}(e^{j\tilde{\omega}(\omega)}) e^{j\tilde{\omega}(\omega)n} \left( \frac{d\tilde{\omega}(\omega)}{d\omega} \right) d\omega \quad (3.17)$$

Derivando (3.14) em relação a  $\omega$ , obtém-se

$$\frac{d\tilde{\omega}(\omega)}{d\omega} = \frac{(1 - \alpha^2)e^{j\omega}}{(e^{j\omega} - \alpha)(1 - \alpha e^{j\omega})} \quad (3.18)$$

que é real e é expressa aqui em termos de  $e^{j\omega}$ . A substituição de (3.12), (3.13) e (3.18) em (3.17), resulta em

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \frac{(1 - \alpha^2)e^{j\omega}}{(e^{j\omega} - \alpha)(1 - \alpha e^{j\omega})} \left( \frac{e^{j\omega} - \alpha}{1 - \alpha e^{j\omega}} \right)^n d\omega \quad (3.19)$$

Considerando que  $x(n)$  é uma seqüência real e causal, (3.19) pode ser expressa por

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \Lambda(e^{j\omega}) \Phi_n(e^{j\omega}) d\omega \quad (3.20)$$

onde

$$\Lambda(e^{j\omega}) = \frac{1}{1 - \alpha e^{-j\omega}} \quad (3.21)$$

e

$$\Phi_n(e^{j\omega}) = \begin{cases} 1, & n = 0 \\ \frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} [\Psi(e^{j\omega})]^{n-1}, & n > 0 \end{cases} \quad (3.22)$$

com

$$\Psi(e^{j\omega}) = e^{-j\tilde{\omega}(\omega)} = \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \quad (3.23)$$

A expressão (3.20) é demonstrada no final deste tópico. A seqüência  $\tilde{x}(n)$ , expressa por (3.20), pode ser obtida considerando-se  $i = 0$  na seguinte transformada inversa de Fourier

$$\tilde{x}(n, i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \Lambda(e^{j\omega}) \Phi_n(e^{j\omega}) e^{j\omega i} d\omega \quad (3.24)$$

de maneira que

$$\tilde{x}(n, 0) = \tilde{x}(n) \quad (3.25)$$

A transformada (3.24) expressa uma operação de filtragem em que  $\tilde{x}(n, i)$  é uma saída do banco de filtros em cascata  $\Lambda(z)\Phi_n(z)$  com entrada  $x(-i)$ . A Figura 3.4 mostra o diagrama da filtragem em cascata para a obtenção de  $\tilde{x}(n)$ .

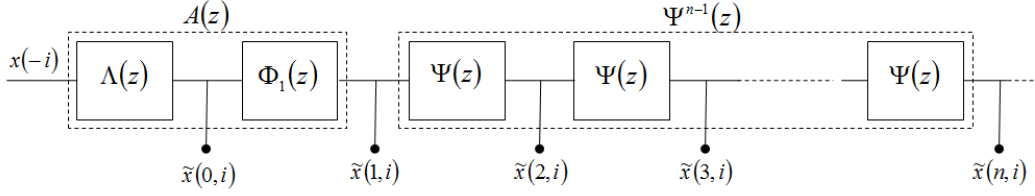


Figura 3.4: Banco de filtros em cascata com entrada  $x(-i)$  e saída  $\tilde{x}(n, i)$  em sua  $(n+1)$ -ésima seção. Os valores  $\tilde{x}(n, 0)$  das seqüências de saída constituem a seqüência  $\tilde{x}(n)$ .

Adotando

$$\tilde{X}_n(z) \leftrightarrow \tilde{x}(n, i) \quad (3.26)$$

a filtragem em cascata da Figura 3.4 contempla as seguintes etapas

$$\begin{aligned} \tilde{X}_0(z) &= \Lambda(z)X^*(z) \\ \tilde{X}_1(z) &= \Phi_1(z)\tilde{X}_0(z) \\ \tilde{X}_n(z) &= \Psi(z)\tilde{X}_{n-1}(z), \quad n \geq 2 \end{aligned} \quad (3.27)$$

onde, explicitando  $\Lambda(z)$ ,  $\Phi_1(z)$  e  $\Psi(z)$ , e multiplicando  $1 - \alpha z^{-1}$ , obtém-se o conjunto de equações

$$\begin{aligned} \tilde{X}_0(z) - \alpha\tilde{X}_0(z)z^{-1} &= X^*(z) \\ \tilde{X}_1(z) - \alpha\tilde{X}_1(z)z^{-1} &= (1 - \alpha^2)\tilde{X}_0(z)z^{-1} \\ \tilde{X}_n(z) - \alpha\tilde{X}_n(z)z^{-1} &= \tilde{X}_{n-1}(z)z^{-1} - \alpha\tilde{X}_{n-1}(z), \quad n \geq 2 \end{aligned} \quad (3.28)$$

Aplicando a transformada  $z$  inversa a (3.28) e rearranjando seus termos, chega-se a

$$\tilde{x}(n, i) = \begin{cases} x(-i) + \alpha\tilde{x}(0, i-1), & n = 0 \\ (1 - \alpha^2)\tilde{x}(0, i-1) + \alpha\tilde{x}(1, i-1), & n = 1 \\ \tilde{x}(n-1, i-1) + \alpha(\tilde{x}(n, i-1) - \tilde{x}(n-1, i)), & n = 2, 3, \dots, \infty \end{cases} \quad (3.29)$$

$$i = -\infty, \dots, -2, -1, 0$$

$$\tilde{x}(n) = \tilde{x}(n, 0)$$

que representa a fórmula recursiva para o cálculo de  $\tilde{x}(n)$  a partir de  $x(n)$ .



### Demonstração de (3.20)

A expressão (3.19), rearranjada, pode ser expressa por

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \frac{(1-\alpha^2)e^{j\omega}}{(1-\alpha e^{j\omega})^2} \left( \frac{e^{j\omega} - \alpha}{1 - \alpha e^{j\omega}} \right)^{n-1} d\omega \quad (3.30)$$

que, para  $n = 0$ , é dada por

$$\tilde{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \frac{(1-\alpha^2)e^{j\omega}}{(e^{j\omega} - \alpha)(1 - \alpha e^{j\omega})} d\omega \quad (3.31)$$

Expandindo o segundo fator do integrando em frações parciais, obtém-se

$$\tilde{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \frac{\alpha e^{-j\omega}}{1 - \alpha e^{-j\omega}} d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \frac{1}{1 - \alpha e^{j\omega}} d\omega \quad (3.32)$$

A primeira integral de (3.32) é nula. Isso é justificado considerando que ela corresponde à zero-ésima amostra da convolução entre as seqüências causais associadas a  $X(e^{j\omega})$  e  $\frac{\alpha e^{-j\omega}}{1 - \alpha e^{-j\omega}}$ . Como ambas as seqüências são causais, isso equivale ao produto entre suas *zero-ésimas* amostras. Como o fator  $\frac{\alpha e^{-j\omega}}{1 - \alpha e^{-j\omega}}$  tem um atraso e, portanto a seqüência a ele associada tem amostra zero nula, tal produto é nulo. Assim, somente o segundo termo de (3.32) é mantido e  $\tilde{x}(0)$  passa a ser expresso por

$$\tilde{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \frac{1}{1 - \alpha e^{j\omega}} d\omega \quad (3.33)$$

A causalidade da seqüência  $x(n)$  implica então, que  $\tilde{x}(0)$  pode ser obtido por (3.33) no lugar de (3.31).

Como  $\tilde{x}(0)$  é real, conjugando (3.33) obtém-se

$$\tilde{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \frac{1}{1 - \alpha e^{-j\omega}} d\omega \quad (3.34)$$

Destaca-se que, para a validade de (3.34), a condição de que  $\tilde{x}(0)$  seja real equivale à condição de que  $x(n)$  e, portanto,  $\tilde{x}(n)$  sejam reais. Isso procede, para  $\tilde{x}(0)$ , na expressão (3.33), já que ela representa uma convolução (amostra zero) entre a seqüência associada ao fator  $\frac{1}{1 - \alpha e^{j\omega}}$ , que é real, e a seqüência  $x(n)$ ; de maneira que só se garantem resultados reais para  $x(n)$  real. A condição é generalizada para  $\tilde{x}(n)$  estendendo-se tal raciocínio à expressão (3.19), ou simplesmente por considerar que a simetria do espectro é preservada no mapeamento de freqüências (ver seção 3.2.1).

Assumindo-se  $n = 1$  na expressão (3.30) conjugada, obtém-se

$$\tilde{x}(1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \left[ \frac{1}{1 - \alpha e^{-j\omega}} \right] \left[ \frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} \right] d\omega \quad (3.35)$$

Assumindo  $n \geq 2$  na expressão (3.30) conjugada

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \left[ \frac{1}{1 - \alpha e^{-j\omega}} \right] \left[ \frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} \right] \left[ \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \right]^{n-1} d\omega \quad (3.36)$$

As equações (3.34), (3.35) e (3.36), representadas com base em  $\Lambda(e^{j\omega})$ ,  $\Phi_1(e^{j\omega})$  e  $\Psi(e^{j\omega})$  ((3.21), (3.22) e (3.23)) são dadas respectivamente por

$$\tilde{x}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \Lambda(e^{j\omega}) d\omega \quad (3.37)$$

$$\tilde{x}(1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \Lambda(e^{j\omega}) \Phi_1(e^{j\omega}) d\omega \quad (3.38)$$

$$\tilde{x}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X^*(e^{j\omega}) \Lambda(e^{j\omega}) \Phi_1(e^{j\omega}) [\Psi(e^{j\omega})]^{n-1} d\omega, \quad n \geq 2 \quad (3.39)$$

As expressões (3.34), (3.38) e (3.39) são contempladas pela expressão (3.20).

### 3.4.2 Cálculo de $\tilde{x}_{\alpha_2}(n)$ a partir de $\tilde{x}_{\alpha_1}(n)$

É possível generalizar (3.29) para permitir o trânsito entre seqüências associadas a diferentes distorções espectrais. Seja  $\tilde{x}_{\alpha_1}(n)$  uma seqüência distorcida em freqüência conhecida, associada ao espectro no eixo de freqüências  $\tilde{\omega}_1$  relacionado ao eixo linear  $\omega$  por

$$e^{j\tilde{\omega}_1(\omega)} = \frac{e^{j\omega} - \alpha_1}{1 - \alpha_1 e^{j\omega}} \quad (3.40)$$

Uma segunda seqüência distorcida em freqüência  $\tilde{x}_{\alpha_2}(n)$ , por sua vez, tem espectro no eixo  $\tilde{\omega}_2$  relacionado ao eixo linear  $\omega$  por

$$e^{j\tilde{\omega}_2(\omega)} = \frac{e^{j\omega} - \alpha_2}{1 - \alpha_2 e^{j\omega}} \quad (3.41)$$

Considerando que o eixo linear  $\omega$  está relacionado a  $\tilde{\omega}_1$  por

$$e^{j\omega(\tilde{\omega}_1)} = \frac{e^{j\tilde{\omega}_1} + \alpha_1}{1 + \alpha_1 e^{j\tilde{\omega}_1}} \quad (3.42)$$

Substituindo (3.42) em (3.41) com  $\omega = \omega(\tilde{\omega}_1)$ , após alguns passos de desenvolvimento, obtém-se a relação

$$e^{j\tilde{\omega}_2(\omega(\tilde{\omega}_1))} = \frac{e^{j\tilde{\omega}_1} - \bar{\alpha}}{1 - \bar{\alpha}e^{j\tilde{\omega}_1}} \quad (3.43)$$

com

$$\bar{\alpha} = \frac{\alpha_2 - \alpha_1}{1 - \alpha_1\alpha_2} \quad (3.44)$$

Assim, se o desenvolvimento da seção 3.4.1 for refeito considerando  $\tilde{x}_{\alpha_2}(n)$  e  $\tilde{x}_{\alpha_1}(n)$  nos lugares de  $\tilde{x}(n)$  e  $x(n)$ , respectivamente, obtém-se a fórmula recursiva para o cálculo de  $\tilde{x}_{\alpha_2}(n)$  a partir de  $\tilde{x}_{\alpha_1}(n)$ , dada por

$$\tilde{x}_{\alpha_2}(n, i) = \begin{cases} \tilde{x}_{\alpha_1}(-i) + \bar{\alpha}\tilde{x}_{\alpha_2}(0, i-1), & n = 0 \\ (1 - \bar{\alpha}^2)\tilde{x}_{\alpha_2}(0, i-1) + \bar{\alpha}\tilde{x}_{\alpha_2}(1, i-1), & n = 1 \\ \tilde{x}_{\alpha_2}(n-1, i-1) + \bar{\alpha}(\tilde{x}_{\alpha_2}(n, i-1) - \tilde{x}_{\alpha_2}(n-1, i)), & n = 2, 3, \dots, \infty \end{cases} \quad (3.45)$$

$$i = -\infty, \dots, -2, -1, 0$$

$$\tilde{x}_{\alpha_2}(n) = \tilde{x}_{\alpha_2}(n, 0) \quad \bar{\alpha} = \frac{\alpha_2 - \alpha_1}{1 - \alpha_1\alpha_2}$$

### 3.4.3 Transformação em Freqüência e Suavização Espectral

A fórmula recursiva (3.29) que representa o cálculo de uma seqüência distorcida em freqüência  $\tilde{x}_{\alpha_2}(n)$  com parâmetro  $\alpha_2$ , a partir de outra  $\tilde{x}_{\alpha_1}(n)$  distorcida em freqüência com parâmetro  $\alpha_1$ , satisfaz  $\tilde{X}_{\alpha_2}(e^{j\tilde{\omega}_2(\omega)}) = \tilde{X}_{\alpha_1}(e^{j\tilde{\omega}_1(\omega)}) = X(e^{j\omega})$  assumindo, a priori, que  $\tilde{x}_{\alpha_1}(n)$  e  $\tilde{x}_{\alpha_2}(n)$  são seqüências de comprimento infinito. Uma vez que na prática não se dispõe de seqüências infinitas, a discussão a seguir assumirá para elas, comprimentos finitos e fornecerá, posteriormente, critérios para a escolha de seus valores. Assumindo que  $\tilde{\mathbf{x}}_{\alpha_1, N_1} = [\tilde{x}_{\alpha_1}(0) \tilde{x}_{\alpha_1}(1) \tilde{x}_{\alpha_1}(2) \dots \tilde{x}_{\alpha_1}(N_1)]^T$  e  $\tilde{\mathbf{x}}_{\alpha_2, N_2} = [\tilde{x}_{\alpha_2}(0) \tilde{x}_{\alpha_2}(1) \tilde{x}_{\alpha_2}(2) \dots \tilde{x}_{\alpha_2}(N_2)]^T$  (onde  $N_1$  e  $N_2$  assumem valores inteiros positivos), a transformação em freqüência que leva  $\tilde{\mathbf{x}}_{\alpha_1, N_1}$  em  $\tilde{\mathbf{x}}_{\alpha_2, N_2}$  através de

$$\tilde{x}_{\alpha_2}(n, i) = \begin{cases} \tilde{x}_{\alpha_1}(-i) + \bar{\alpha}\tilde{x}_{\alpha_2}(0, i-1), & n = 0 \\ (1 - \bar{\alpha}^2)\tilde{x}_{\alpha_2}(0, i-1) + \bar{\alpha}\tilde{x}_{\alpha_2}(1, i-1), & n = 1 \\ \tilde{x}_{\alpha_2}(n-1, i-1) + \bar{\alpha}(\tilde{x}_{\alpha_2}(n, i-1) - \tilde{x}_{\alpha_2}(n-1, i)), & n = 2, 3, \dots, N_1 \end{cases} \quad (3.46)$$

$$i = -N_2, \dots, -2, -1, 0$$

$$\tilde{x}_{\alpha_2}(n) = \tilde{x}_{\alpha_2}(n, 0) \quad \bar{\alpha} = \frac{\alpha_2 - \alpha_1}{1 - \alpha_1 \alpha_2}$$

será denotada por

$$\tilde{\mathbf{x}}_{\alpha_2, N_2} = T_{\Lambda}(\tilde{\mathbf{x}}_{\alpha_1, N_1}, N_1, N_2, \bar{\alpha}) \quad (3.47)$$

que representa a forma geral da transformação em frequência. Tal transformação se restringe a seqüências reais causais e estáveis, restrições impostas no seu desenvolvimento (seção 3.4.1).

Alguns critérios podem ser utilizados para se determinar os comprimentos das seqüências envolvidas na transformação em frequência, com base em seu objetivo:

- **Suavização Espectral:** Quando se visa uma representação suavizada do espectro da seqüência de entrada restringe-se o comprimento da seqüência transformada a um valor, em geral, baixo em relação àquele da seqüência de entrada. Isso corresponde a uma truncagem da seqüência transformada. A suavização espectral característica de uma seqüência distorcida em frequência produz, no eixo linear de frequências, uma curva espectral com resolução não-uniforme (o que será abordado na seção 3.4.4). Para que a suavização seja efetiva não é necessário que todas as amostras da seqüência de entrada sejam utilizadas, porém ela deve ter um comprimento mínimo. Se deseja-se obter uma seqüência transformada de comprimento  $N_2 + 1 = M$  a partir de uma seqüência de comprimento  $N \gg M$ , devem ser utilizadas no mínimo  $N_1 + 1 \approx \frac{1+|\bar{\alpha}|}{1-|\bar{\alpha}|}M$  amostras da seqüência de entrada na transformação (3.47).
- **Representação Espectral Completa:** Quando se deseja uma representação completa do espectro da seqüência original através da seqüência transformada, a última deve ser extensa o suficiente. Considerando uma seqüência de entrada com comprimento  $N_1 + 1 = M$ , devem ser calculadas no mínimo  $N_2 + 1 \approx \frac{1+|\bar{\alpha}|}{1-|\bar{\alpha}|}M$  amostras da seqüência de saída na transformação (3.47).

Tais critérios são adotados com base na discussão da seção 3.4.4.

A seguir são exemplificadas transformações entre seqüências distorcidas e lineares em freqüência.

Uma transformação que leva uma seqüência na escala linear de freqüências em uma seqüência distorcida em freqüência com parâmetro  $\alpha$  é obtida assumindo  $\alpha_1 = 0$  e  $\alpha_2 = \alpha$ , de maneira que  $\bar{\alpha} = \alpha$  em (3.47). Adotando  $N_1 = J$  e  $N_2 = L$  e expressando  $\tilde{x}_{\alpha_1, N_1}$  por  $\mathbf{x}_J = [x(0) x(1) \dots x(J)]^T$  e  $\tilde{x}_{\alpha_2, N_2}$  por  $\tilde{\mathbf{x}}_L = [\tilde{x}(0) \tilde{x}(1) \dots \tilde{x}(L)]^T$ , a transformação em freqüência que leva  $\mathbf{x}_J$  em  $\tilde{\mathbf{x}}_L$  é expressa por

$$\tilde{\mathbf{x}}_L = T_\Lambda(\mathbf{x}_J, J, L, \alpha) \quad (3.48)$$

caso demonstrado na seção 3.4.1 (considerando seqüências infinitas).

Uma transformação inversa pode ser obtida intercambiando as seqüências e parâmetros atribuídos a  $\tilde{x}_{\alpha_1, N_1}$  e  $\tilde{x}_{\alpha_2, N_2}$  na transformação direta. Assim  $\tilde{x}_{\alpha_1, N_1} = \tilde{\mathbf{x}}_L$ ,  $\tilde{x}_{\alpha_2, N_2} = \mathbf{x}_J$ ,  $N_1 = L$ ,  $N_2 = J$ ,  $\alpha_1 = \alpha$  e  $\alpha_2 = 0$ , de maneira que  $\bar{\alpha} = -\alpha$ . Tal transformação leva  $\tilde{\mathbf{x}}_L$  em  $\mathbf{x}_J$  e é expressa por

$$\mathbf{x}_J = T_\Lambda(\tilde{\mathbf{x}}_L, L, J, -\alpha) \quad (3.49)$$

A seção a seguir mostra que a truncagem de uma seqüência distorcida em freqüência produz, na escala linear de freqüências, uma versão do espectro da seqüência original suavizada com resolução não-uniforme em freqüência.

### 3.4.4 Análise Espectral com Resolução Não-Uniforme

De interesse para o presente trabalho é, não só a representação de  $x(n)$  através de uma seqüência  $\tilde{x}(n)$  distorcida em freqüência, mas também o resultado produzido ao se truncar (janelar)  $\tilde{x}(n)$ . Considerando  $x(n)$  uma seqüência estacionária, quando truncada, seu espectro é suavizado, o que equivale a uma redução em sua resolução em freqüência. Se, porém, uma truncagem for efetuada em  $\tilde{x}(n)$ , seu espectro, ao ser reconduzido à escala linear, terá resolução em freqüência não-uniforme [21]. Nesta seção demonstra-se que tal espectro corresponde àquele que seria obtido a partir de  $x(n)$  por um janelamento de comprimento variável conforme a freqüência de análise. Tal abordagem permite justificar os critérios enunciados na seção anterior para os comprimentos das seqüências envolvidas em uma transformação em freqüência. Atribuindo-se o valor adequado ao parâmetro  $\alpha$ , a curva de mapeamento  $\tilde{\omega}(\omega)$ ,

aproxima a escala mel de frequência perceptual e permite que o espectro suavizado tenha resolução em frequência próxima àquela da audição humana.

Considerando  $x(n)$  estacionária e relacionada a  $\tilde{x}(n)$  por  $\tilde{X}(e^{j\tilde{\omega}(\omega)}) = X(e^{j\omega})$  onde  $x(n) \leftrightarrow X(e^{j\omega})$ ,  $\tilde{x}(n) \leftrightarrow \tilde{X}(e^{j\tilde{\omega}})$  e  $\tilde{\omega}(\omega)$  é dado por (3.13), será demonstrado que se a seqüência  $\tilde{x}(n)$  for truncada em  $M$  de maneira a produzir  $\tilde{x}_M(n) = r_M(n)\tilde{x}(n)$ , onde  $r_M(n)$  é uma janela retangular de comprimento  $M$ , sua transformada de Fourier, dada por  $\tilde{X}_M(e^{j\tilde{\omega}})$ , ao ser reconduzida à escala linear através de  $\tilde{\omega} = \tilde{\omega}(\omega)$  produz

$$\tilde{X}_M(e^{j\tilde{\omega}(\omega)}) = X_{nu}(e^{j\omega}) \quad (3.50)$$

onde  $X_{nu}(e^{j\omega})$  corresponde a uma versão de  $X(e^{j\omega})$  com resolução não uniforme.

Adotando inicialmente uma seqüência comum  $x(n)$  estacionária e janelando-a com uma janela retangular  $r_L(n)$  de comprimento  $L$ , obtém-se  $x_L(n) = x(n)r_L(n)$ . Seu espectro pode ser expresso pela convolução

$$X_L(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega'}) R_L(e^{j(\omega-\omega')}) d\omega' \quad (3.51)$$

onde  $R_L(e^{j\omega})$  é a transformada de Fourier de  $r_L(n)$ . O fator  $R_L(e^{j(\omega-\omega')})$  do integrando da expressão é dado por

$$R_L(e^{j(\omega-\omega')}) \approx \frac{\text{sen}\left(\frac{\omega-\omega'}{2}L\right)}{\frac{\omega-\omega'}{2}} e^{j\frac{\omega-\omega'}{2}L\left(\frac{L-1}{L}\right)} \quad (3.52)$$

aproximação válida para  $|\omega - \omega'| \ll 1$ . Em tal expressão fica explícito que o comprimento  $L$  da janela  $r_L(n)$  determina a largura do lóbulo central de  $R_L(e^{j(\omega-\omega')})$  e que, portanto,  $x_L(n) = x(n)r_L(n)$  fixa uma resolução em frequência para o espectro de  $x(n)$ . Quanto maior o valor de  $L$ , mais estreito é o lóbulo central de  $R_L(e^{j(\omega-\omega')})$  e maior é a resolução de  $X_L(e^{j\omega})$ .

Da mesma maneira que para o espectro de  $x(n)$ , que tem o espectro de sua versão truncada  $x_L(n)$  dado em (3.51), o espectro de  $\tilde{x}_M(n)$ , versão truncada de  $\tilde{x}(n)$  é expresso por

$$\tilde{X}_M(e^{j\tilde{\omega}}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{X}(e^{j\tilde{\omega}'}) R_M(e^{j(\tilde{\omega}-\tilde{\omega}')}) d\tilde{\omega}' \quad (3.53)$$

Reconduzindo (3.53) à escala linear de frequências, através de  $\tilde{\omega} = \tilde{\omega}(\omega)$ , obtém-se

$$\tilde{X}_M(e^{j\tilde{\omega}(\omega)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{X}(e^{j\tilde{\omega}(\omega')}) R_M(e^{j(\tilde{\omega}(\omega) - \tilde{\omega}(\omega'))}) \left( \frac{d\tilde{\omega}(\omega')}{d\omega'} \right) d\omega' \quad (3.54)$$

Substituindo  $\tilde{X}(e^{j\tilde{\omega}(\omega')}) = X(e^{j\omega'})$  e (3.50) em (3.54), obtém-se

$$X_{nu}(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega'}) R_M(e^{j(\tilde{\omega}(\omega) - \tilde{\omega}(\omega'))}) \left( \frac{d\tilde{\omega}(\omega')}{d\omega'} \right) d\omega' \quad (3.55)$$

que corresponde à versão com resolução não-uniforme do espectro original  $X(e^{j\omega})$ . Comparando a expressão (3.55) à expressão (3.51), observa-se que no primeiro caso, quando  $X(e^{j\omega})$  é convoluído com  $R_N(e^{j\omega})$ , o fator  $R_N(e^{j(\omega - \omega')})$  do integrando é caracterizado, naturalmente, por ter largura de lóbulo central constante e ser simétrico no eixo  $\omega'$ , em relação a  $\omega' = \omega$ . Já no integrando de (3.55), o fator que desempenha papel equivalente é dado por  $R_M(e^{j(\tilde{\omega}(\omega) - \tilde{\omega}(\omega'))}) \left( \frac{d\tilde{\omega}(\omega')}{d\omega'} \right)$  onde, devido à relação não-linear entre  $\tilde{\omega}$  e  $\omega'$ , a largura do lóbulo central de  $R_M(e^{j(\tilde{\omega}(\omega) - \tilde{\omega}(\omega'))})$  varia com  $\omega$  e sua curva se *deforma* de maneira que não mais se constata simetria no eixo  $\omega'$  em relação a  $\omega' = \omega$ . A variação da largura do lóbulo central com  $\omega$  propicia a não-uniformidade do espectro resultante.

A dependência da largura do lóbulo central de  $R_M(e^{j(\tilde{\omega}(\omega) - \tilde{\omega}(\omega'))})$  com  $\omega$ , o que especifica a não-uniformidade da resolução espectral obtida em  $X_{nu}(e^{j\omega})$ , pode ser compreendida sob a perspectiva a seguir: Uma vez que o lóbulo central prepondera sobre os lóbulos laterais, é razoável aproximar-se o módulo da expressão (3.55) limitando sua integral ao intervalo  $[\omega - \epsilon, \omega + \epsilon]$  do lóbulo central. Considerando  $\epsilon$  pequeno, pode-se assumir o fator  $\frac{d\tilde{\omega}(\omega')}{d\omega'}$  do integrando de (3.55) aproximadamente constante com valor  $\left. \frac{d\tilde{\omega}(\omega')}{d\omega'} \right|_{\omega'=\omega}$  e  $\tilde{\omega}(\omega) - \tilde{\omega}(\omega')$  pode ser substituído por sua aproximação linear  $(\omega - \omega') \left. \frac{d\tilde{\omega}(\omega')}{d\omega'} \right|_{\omega'=\omega}$ . Assim, a expressão (3.55), em módulo, assume a forma

$$|X_{nu}(e^{j\omega})| \approx \frac{1}{2\pi} \left| \int_{\omega-\epsilon}^{\omega+\epsilon} X(e^{j\omega'}) R_M \left( e^{j(\omega - \omega') \left. \frac{d\tilde{\omega}(\omega')}{d\omega'} \right|_{\omega'=\omega}} \right) \left( \left. \frac{d\tilde{\omega}(\omega')}{d\omega'} \right|_{\omega'=\omega} \right) d\omega' \right| \quad (3.56)$$

onde

$$\begin{aligned}
R_M \left( e^{j \left( (\omega - \omega') \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega} \right)} \right) \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega} \\
= \frac{\text{sen} \left( \frac{(\omega - \omega')}{2} \left( \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega} \right) M \right)}{\frac{(\omega - \omega')}{2}} e^{j \frac{(\omega - \omega')}{2} \left( \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega} \right) M \left( \frac{M-1}{M} \right)} \quad (3.57)
\end{aligned}$$

Na expressão (3.51) o fator associado à janela é uma função de  $(\omega - \omega')$ , enquanto em (3.56), tal fator é função de  $(\omega - \omega') \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega}$ . Em (3.57) observa-se que  $\frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega}$  atua contraindo ou dilatando a curva da função e, assim, provocando diminuição ou aumento na largura de seu lóbulo central. A dependência de  $\frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega}$  com  $\omega$  (mostrada a Figura 3.5(b)) implica, portanto, na dependência da resolução espectral com o valor da frequência  $\omega$  em que a análise é efetuada. Além disso,  $M \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega}$  desempenha o papel de comprimento da janela de análise da seqüência  $x(n)$ . Isso pode ser constatado, assumindo que  $M \frac{d\tilde{\omega}(\omega')}{d\omega'} \Big|_{\omega'=\omega} = M'(\omega)$  em (3.57), o que resulta em

$$R_M \left( e^{j \left( (\omega - \omega') \frac{M'(\omega)}{M} \right)} \right) \frac{M'(\omega)}{M} = \frac{\text{sen} \left( \frac{(\omega - \omega')}{2} M'(\omega) \right)}{\frac{(\omega - \omega')}{2}} e^{j \frac{(\omega - \omega')}{2} M'(\omega) \left( \frac{M-1}{M} \right)} \quad (3.58)$$

expressão onde  $M'(\omega)$  tem o mesmo papel de  $L$  em (3.52). Desta maneira, quando se limita a seqüência  $\tilde{x}(n)$  por uma janela de comprimento  $M$  ( $\tilde{x}_M(n) = \tilde{x}(n)r_M(n)$ ), seu espectro, ao ser conduzido à escala linear de frequências (eixo  $\omega$ ), se aproxima àquele que resultaria de uma análise sobre  $x(n)$  efetuada com uma janela temporal de comprimento variável com a frequência de análise, cujo valor aproximado é  $M'(\omega)$ .  $M'(\omega)$  será denominado de *comprimento equivalente* associado ao janelamento de comprimento  $M$  da seqüência  $\tilde{x}(n)$ .

Será útil comparar as características do espectro que resultaria de uma análise de  $x(n)$  por um janelamento de comprimento  $M$  fixo, e aquele obtido por um janelamento de comprimento variável  $M'(\omega)$ . Derivando (3.14) (ou (3.15)), obtém-se

$$\frac{d\tilde{\omega}(\omega)}{d\omega} = \frac{M'(\omega)}{M} = \frac{1 - \alpha^2}{1 + \alpha^2 - 2\alpha \cos \omega} \quad (3.59)$$



cujas curvas para três valores de  $\alpha$  estão representadas na Figura 3.5(b). Da expressão (3.59) conclui-se que, dado um valor de  $\alpha$ , o comprimento equivalente da janela de análise varia de  $M'(0) = \frac{1+\alpha}{1-\alpha}M$  a  $M'(\pi) = \frac{1-\alpha}{1+\alpha}M$  passando por  $M'(\cos^{-1} \alpha) = M$ . Quando, por exemplo,  $\alpha > 0$ , observa-se, em  $\omega = 0$ , que  $M'(0) > M$ , o que se reflete em uma resolução espectral superior àquela que seria obtida por um janelamento de comprimento  $M$  em  $x(n)$ . Já em  $\omega = \pi$ , observa-se comprimento equivalente  $M'(\pi) < M$  o que se reflete em resolução inferior àquela obtida por um janelamento de comprimento  $M$  em  $x(n)$ . Para  $\alpha < 0$  a situação inversa é observada. Para  $\alpha = 0$ , o eixo de freqüências se torna linear de maneira que  $M'(\omega) = M$ , o que fixa uma resolução para todo o intervalo de freqüências.

Para que o intervalo de variação da resolução do espectro produzido pela análise de  $x(n)$  através da janela de comprimento  $M'(\omega)$  seja efetivo, o valor do comprimento  $N$  de  $x(n)$  deve ser igual ou maior do que o valor máximo do comprimento  $M'(\omega)$ . Caso  $N$  seja inferior a este valor,  $M'(\omega)$  ficará limitado superiormente por  $N$  e a resolução ficará saturada a partir de  $\omega = \omega_s$ , onde  $\omega_s$  satisfaz  $M'(\omega_s) = N$  no sentido de crescimento da resolução no eixo  $\omega$ . Tomando  $\alpha = 0.5$  como exemplo, valor cuja curva de  $\frac{d\tilde{\omega}(\omega)}{d\omega}$  aparece na Figura 3.5 (b), observa-se resolução máxima em  $\omega = 0$ , quando  $\frac{d\tilde{\omega}(\omega)}{d\omega}|_{\omega=0} = 3$ , valor para o qual a análise espectral é efetuada por uma janela de comprimento equivalente  $M'(0) = 3M$ . A resolução mínima ocorre em  $\omega = \pi$ , quando  $\frac{d\tilde{\omega}(\omega)}{d\omega}|_{\omega=\pi} = \frac{1}{3}$  e, portanto a análise espectral é efetuada por uma janela de comprimento equivalente  $M'(\pi) = \frac{M}{3}$ .

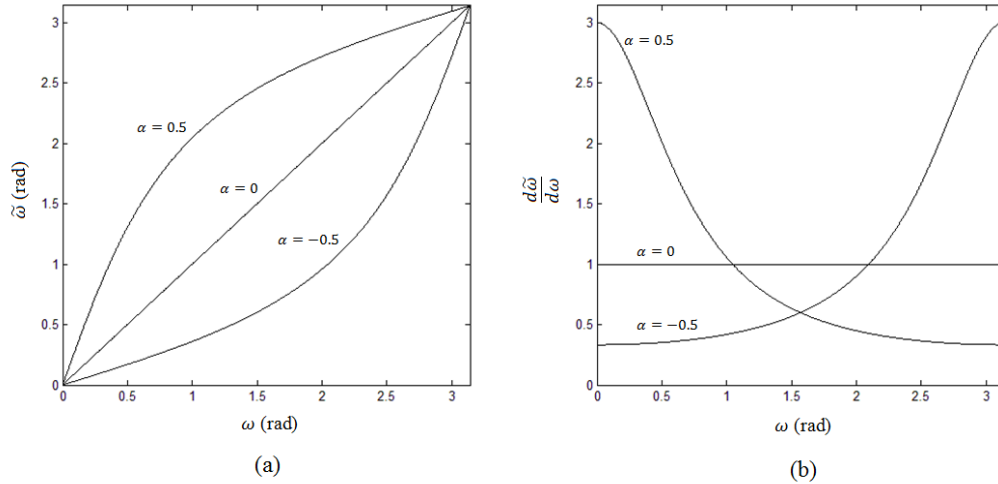


Figura 3.5: (a): Curvas de mapeamento de frequências  $\tilde{\omega}(\omega)$  para  $\alpha = 0.5$ ,  $\alpha = 0$  e  $\alpha = -0.5$ . (b): Respectivas curvas de  $\frac{d\tilde{\omega}(\omega)}{d\omega}$ .

Com base em (3.59) verifica-se que a janela variável assume comprimento equivalente máximo de valor  $\frac{1+|\alpha|}{1-|\alpha|}M$  e mínimo de valor  $\frac{1-|\alpha|}{1+|\alpha|}M$ . A discussão acima nos leva a concluir que, no caso de uma suavização espectral com base na truncagem de  $\tilde{x}(n)$  em  $M$  amostras, para que o espectro resultante tenha resolução variável efetiva ao longo de todo eixo  $\omega$ ,  $\tilde{x}(n)$  deve ter sido transformada a partir de no mínimo  $N \approx \frac{1+|\alpha|}{1-|\alpha|}M$  amostras de  $x(n)$ . Se ao contrário disso deseja-se contemplar por completo o espectro de  $x(n)$ , através de  $\tilde{x}(n)$ , esta última deve ter comprimento equivalente mínimo superior ao comprimento  $N$  da seqüência  $x(n)$ , assim, deve-se exigir que  $\frac{1-|\alpha|}{1+|\alpha|}M > N$ , o que equivale a exigir que  $\tilde{x}(n)$  tenha comprimento  $M > \frac{1+|\alpha|}{1-|\alpha|}N$ . Implicitamente, isso garante que o comprimento equivalente  $M'(\omega)$  da *janela de análise* de  $x(n)$  seja grande o suficiente para variar em um intervalo onde não mais ocorram amostras não-nulas.

A Figura 3.6 mostra as etapas de um processo para a obtenção de um espectro suavizado com resolução não-uniforme. As curvas (d) e (e) permitem a comparação dos espectros suavizados (no eixo  $\omega$ ) associados às versões truncadas em  $M$  amostras das seqüências  $\tilde{x}(n)$  e  $x(n)$ . Comparada à curva (e), a curva (d) apresenta maior resolução nas baixas frequências o que se inverte para as frequências altas.

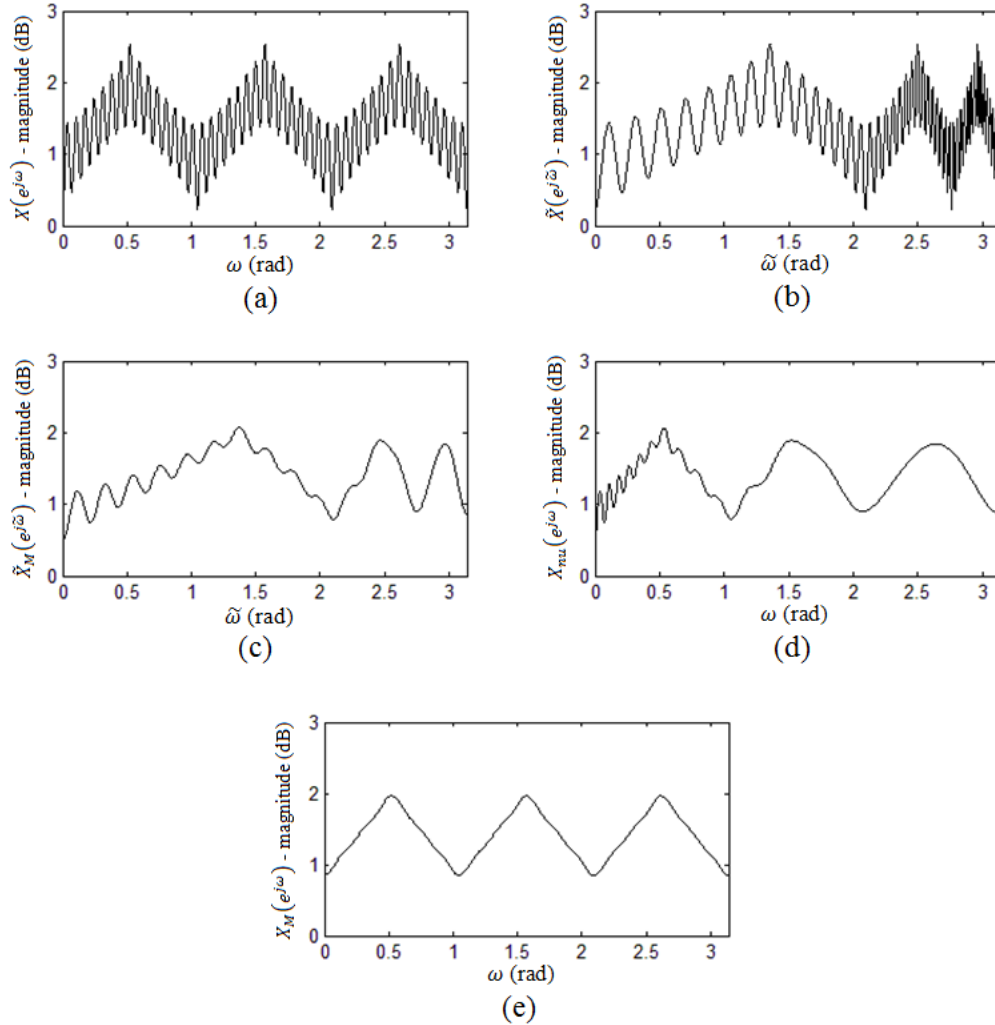


Figura 3.6: Exemplo de obtenção de espectro com resolução não-uniforme (com  $\alpha = 0.5$ ). (a): Magnitude (dB) do espectro  $X(e^{j\omega})$  de uma seqüência  $x(n)$  de comprimento 96. (b): Magnitude (dB) do espectro  $\tilde{X}(e^{j\tilde{\omega}})$  da seqüência transformada  $\tilde{x}(n)$  de comprimento infinito. (c): Magnitude (dB) do espectro  $\tilde{X}_M(e^{j\tilde{\omega}})$  da seqüência  $\tilde{x}_M(n)$  de comprimento 32 (versão truncada de  $\tilde{x}(n)$ ). (d): Magnitude (dB) de  $X_{nu}(e^{j\omega}) = \tilde{X}_M(e^{j\tilde{\omega}(\omega)})$ , uma versão suavizada com resolução espectral não-uniforme de  $X(e^{j\omega})$ . (e) Magnitude (dB) do espectro  $X_M(e^{j\omega})$  da seqüência  $x(n)$  truncada com comprimento 32: resolução uniforme.

A escolha correta do parâmetro  $\alpha$  pode levar o eixo  $\tilde{\omega}$  a aproximar a escala mel de frequência perceptual. Isso aproxima a resolução em frequência do espectro àquela da audição humana. Sob a adoção de um número fixo de parâmetros para a representação do espectro, aqueles que incorporam características perceptuais se mostram mais eficientes em aplicações envolvendo síntese e reconhecimento de fala. A Tabela 3.1 mostra, para diversas taxas de amostragem, os valores de  $\alpha$  que aproximam a escala

mel. A Figura 3.7 compara a curva da escala mel com a curva de aproximação obtida com  $\alpha = 0.42$  (para 16kHz).

Freq. de amostragem	8kHz	10kHz	12kHz	16kHz	20kHz	22.05kHz
Parâmetro $\alpha$ (escala mel)	0.31	0.35	0.37	0.42	0.44	0.45

Tabela 3.1: exemplos de valores de  $\alpha$  que aproximam a escala mel de frequências para diferentes frequências de amostragem.

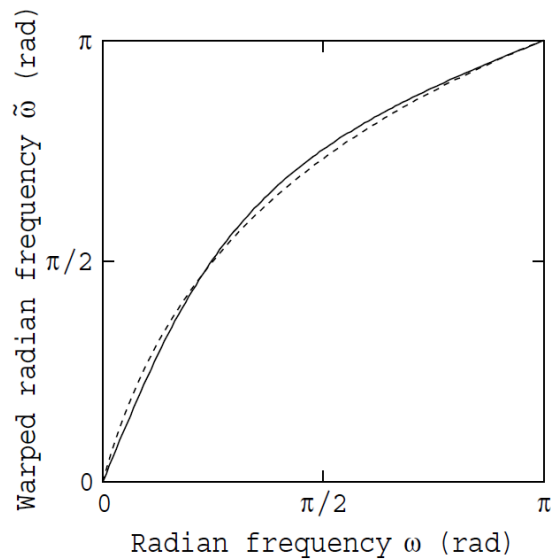


Figura 3.7: Curva de mapeamento de frequências  $\tilde{\omega}(\omega)$  para  $\alpha = 0.42$  e curva da escala mel de frequência perceptual (linha pontilhada). Frequência de amostragem de 16kHz. (figura extraída de [22])

### 3.4.5 DFT e Amostragem do Espectro Distorcido em Frequência

Esta seção aborda o cálculo do espectro distorcido em frequência via DFT e demonstra que o critério para o comprimento de uma seqüência distorcida em frequência  $\tilde{x}(n)$  com base no comprimento da seqüência original  $x(n)$  (critério para representação completa enunciado na seção 3.4.3 e demonstrado na seção anterior) é

também um critério para o número de pontos de uma amostragem do espectro distorcido em frequência  $\tilde{X}(e^{j\tilde{\omega}})$ , para a prevenção de *aliasing* temporal.

O cálculo do espectro distorcido em frequência pode ser obtido via DFT da seguinte maneira: calculam-se, no mínimo  $N_\alpha \approx \frac{1+|\alpha|}{1-|\alpha|}N$  amostras da seqüência distorcida em frequência  $\tilde{x}(n)$ , através da transformação em frequência aplicada à seqüência  $x(n)$  de comprimento  $N$ . O cálculo do espectro distorcido em frequência é feito através da DFT

$$\tilde{X}(e^{j\tilde{\omega}_l}) = \sum_{n=0}^{N_\alpha-1} \tilde{x}(n)e^{-j\tilde{\omega}_l n}, \quad \tilde{\omega}_l = \frac{2\pi}{N_\alpha} l \quad (3.60)$$

$$l = 0, 1, \dots, N_\alpha - 1$$

Considerando que não dispomos da seqüência  $\tilde{x}(n)$ , uma maneira de se efetuar o cálculo discreto da versão distorcida em frequência do espectro de  $x(n)$ , seria através de

$$\tilde{X}(e^{j\tilde{\omega}_l}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega(\tilde{\omega}_l)n}, \quad \tilde{\omega}_l = \frac{2\pi}{N_\alpha} l \quad (3.61)$$

$$l = 0, 1, \dots, N_\alpha - 1$$

que corresponde à transformada de Fourier de  $x(n)$  calculada, ponto a ponto, nas frequências  $\omega(\tilde{\omega}_l)$  com  $l = 0, 1, \dots, N_\alpha - 1$  ( $\omega(\tilde{\omega})$  é dada pela expressão (3.8)). Vamos demonstrar que, de posse de uma alternativa como tal para se amostrar o espectro distorcido em frequência da seqüência  $x(n)$  (de comprimento  $N$ ), devem ser calculadas, para prevenção de *aliasing* temporal, no mínimo  $N_\alpha \approx \frac{1+|\alpha|}{1-|\alpha|}N$  amostras  $\tilde{X}(e^{j\tilde{\omega}_l})$  igualmente espaçadas em  $\tilde{\omega}$ . Para que se verifique tal critério, considere que o espectro de  $x(n)$ , dado por  $X(e^{j\omega})$ , deve ser representado na forma discreta com no mínimo  $N$  amostras  $X(e^{j\omega_k})$  sob pena de distorções por *aliasing* temporal. Isso corresponde a se exigir um limite máximo para o espaçamento entre suas amostras adjacentes, para que se evite sub-especificação da curva de  $X(e^{j\omega})$ , quando esta assume suas variações mais rápidas. Considere agora que o espectro distorcido em frequência pode ser representado por amostras desigualmente espaçadas  $\tilde{X}(e^{j\tilde{\omega}_k})$  nas frequências  $\tilde{\omega}_k = \tilde{\omega}(\omega_k)$ , onde  $\omega_k = \frac{2\pi}{N}k$  com  $k = 0, 1, \dots, N - 1$ . Isso equivale a uma redistribuição das amostras  $X(e^{j\omega_k})$  igualmente espaçadas em  $\omega$ , em uma disposição com espaçamentos desiguais

ao longo do eixo  $\tilde{\omega}$ . Dado que os espaçamentos desiguais  $\tilde{\omega}_k - \tilde{\omega}_{k-1}$  respeitam a relação  $(\tilde{\omega}(\omega_{k+1}) - \tilde{\omega}(\omega_k))/(\omega_{k+1} - \omega_k)$  as amostras  $\tilde{X}(e^{j\tilde{\omega}_k})$ , especificam totalmente o espectro  $\tilde{X}(e^{j\tilde{\omega}})$ . Porém, por exigência computacional, suas amostras devem estar igualmente espaçadas em  $\tilde{\omega}$ , o que exige que se adote o valor mínimo dentre os espaçamentos  $\tilde{\omega}(\omega_{k+1}) - \tilde{\omega}(\omega_k)$ , como o valor máximo<sup>§</sup> permitido aos espaçamentos uniformes  $\tilde{\omega}_l - \tilde{\omega}_{l-1}$  a serem adotados para a amostragem no eixo  $\tilde{\omega}$ . A relação  $(\tilde{\omega}(\omega_{k+1}) - \tilde{\omega}(\omega_k))/(\omega_{k+1} - \omega_k)$  entre os espaçamentos desiguais no eixo  $\tilde{\omega}$  e os espaçamentos uniformes do eixo  $\omega$  é dada em aproximação por  $\frac{d\tilde{\omega}(\omega_k)}{d\omega}$ , e assume valor mínimo  $\frac{d\tilde{\omega}(\omega_{min})}{d\omega} = \frac{1-|\alpha|}{1+|\alpha|}$ . Isso equivale a se exigir que  $\tilde{X}(e^{j\tilde{\omega}})$  seja amostrado com, no mínimo,  $N_\alpha \approx \frac{1+|\alpha|}{1-|\alpha|}N$  amostras igualmente espaçadas em  $\tilde{\omega}$ , sob pena de sub-especificação do espectro, o que caracterizaria *aliasing* temporal (exemplo na 2.7.5). Assim, verifica-se a validade para a amostragem do espectro distorcido em frequência do critério adotado para o comprimento mínimo da seqüência distorcida em frequência. Destaca-se que isso previne qualquer espectro  $\tilde{X}(e^{j\tilde{\omega}_l})$  de distorções por *aliasing* temporal; porém, é possível a adoção de valores inferiores para  $N_\alpha$ , com base em informações prévias sobre a velocidade das variações de  $\tilde{X}(e^{j\tilde{\omega}})$  ao longo do eixo de frequências.

Nota-se que o cálculo do espectro distorcido em frequência é custoso tanto através de (3.60), que exige o cálculo de  $\tilde{x}(n)$  por uma cadeia de  $N_\alpha \approx \frac{1+|\alpha|}{1-|\alpha|}N$  filtros em cascata, quanto por (3.61), que não pode ser efetuado via FFT, tendo as amostras calculadas ponto a ponto.

---

<sup>§</sup> Tal exigência é feita dado que não se tem informações a priori a respeito da curva do espectro. Do contrário, poderia ser relaxada.

## 3.5 Espectro Mel-Logarítmico Generalizado

As seções anteriores trataram da derivação de seqüência distorcidas em freqüência com base em uma seqüência conhecida. A presente seção faz o mesmo tratamento com base no cepstrum generalizado. Um mapeamento não-linear de freqüências aplicado ao espectro logarítmico generalizado leva ao espectro mel-logarítmico generalizado. A seqüência associada a tal espectro é denominada mel-cepstrum generalizado, e pode ser calculada a partir do cepstrum generalizado.

### 3.5.1 Mel-Cepstrum Generalizado

Seja  $x_\gamma(n)$  o cepstrum generalizado associado a uma seqüência  $x(n)$  real e estável conhecida e,  $\tilde{x}_\gamma(n)$ , uma seqüência real e estável relacionada a  $x_\gamma(n)$  com base no que será exposto a seguir.

Considere as transformadas  $z$  de  $x_\gamma(n)$  e  $\tilde{x}_\gamma(n)$  dadas respectivamente por

$$X_\gamma(z) = \sum_{n=-\infty}^{\infty} x_\gamma(n)z^{-n} \quad (\text{expressão (2.6)})$$

e

$$\tilde{X}_\gamma(\tilde{z}) = \sum_{n=-\infty}^{\infty} \tilde{x}_\gamma(n)\tilde{z}^{-n} \quad (3.62)$$

onde  $X_\gamma(z)$  e  $\tilde{X}_\gamma(\tilde{z})$  estão relacionados por

$$\tilde{X}_\gamma(\tilde{z}) = X_\gamma(z(\tilde{z})) = \ln_\gamma \left( X(z(\tilde{z})) \right)^\S \quad (3.63)$$

ou seja, com  $|\tilde{z}| = |z| = 1$ , o espectro mel-logarítmico generalizado  $\tilde{X}_\gamma(e^{j\tilde{\omega}})$  representa o espectro logarítmico generalizado tomado numa escala não-linear de freqüências obtida pela substituição  $\omega = \omega(\tilde{\omega})$  (expressão (3.7)). A seqüência  $\tilde{x}_\gamma(n)$  corresponde, portanto, aos coeficientes de Fourier de  $\ln_\gamma \left( X(e^{j\omega(\tilde{\omega})}) \right)$  que, por sua vez, corresponde ao espectro original tomado simultaneamente na escala logarítmica generalizada de amplitudes e na escala não-linear de freqüências definida por (3.14) e

---

<sup>§</sup> A igualdade tem por objetivo mostrar a relação funcional entre  $X_\gamma(z)$  e  $X(z)$ , porém, dada a multivalência da função log-gen complexa, a definição dada por (2.6) é mais restritiva. Assim, a relação (3.63) será considerada válida sob a restrição de que  $\ln_\gamma(X(z))$  seja uma transformada  $z$  válida de uma seqüência real e estável.

define o espectro mel-logarítmico generalizado (mel-log-gen)  $\tilde{X}_\gamma(e^{j\tilde{\omega}})$ . A seqüência  $\tilde{x}_\gamma(n)$  define o mel-cepstrum generalizado e é obtida pela transformada z inversa de (3.62), dada por

$$\tilde{x}_\gamma(n) = \frac{1}{2\pi j} \oint_C \tilde{X}_\gamma(\tilde{z}) \tilde{z}^{n-1} d\tilde{z} \quad (3.64)$$

onde  $C$  é uma curva fechada, tomada no sentido anti-horário, que inclui a origem e se encontra inteiramente na região de convergência (RC) estável do plano  $\tilde{z}$  (ou seja, a RC que inclui a circunferência de raio unitário).

### 3.5.2 Magnitude Mel-Logarítmica Generalizada

O espectro mel-log-gen  $\tilde{X}_\gamma(e^{j\tilde{\omega}})$  está relacionado ao espectro regular  $X(e^{j\omega(\tilde{\omega})})$  por

$$\tilde{X}_\gamma(e^{j\tilde{\omega}}) = \ln_\gamma \left( X(e^{j\omega(\tilde{\omega})}) \right) \quad (3.65)$$

Pelas mesmas considerações feitas na seção 2.3.2, adota-se  $\ln_\gamma \left( |X(e^{j\omega(\tilde{\omega})})|^2 \right)$  no lugar de  $|\tilde{X}_\gamma(e^{j\tilde{\omega}})|^2$  para representações do espectro mel-log-gen. Da mesma forma que na seção 2.3.2, explicita-se aqui a descrição de  $\ln_\gamma \left( |X(e^{j\omega(\tilde{\omega})})|^2 \right)$  por uma série de Fourier com base no mel-cepstrum generalizado. Utilizando a relação (A.5) do Apêndice A, conclui-se que

$$\ln_\gamma \left( |X(e^{j\omega(\tilde{\omega})})|^2 \right) = 2 \cdot \text{Re}[\tilde{X}_\gamma(e^{j\tilde{\omega}})] + \gamma \cdot |\tilde{X}_\gamma(e^{j\tilde{\omega}})|^2 \quad (3.66)$$

A representação trigonométrica da expressão (3.65) (com base em (3.62)) resulta em

$$\ln_\gamma \left( |X(e^{j\omega(\tilde{\omega})})|^2 \right) = \tilde{d}_\gamma(0) + 2 \cdot \sum_{n=0}^{\infty} \tilde{d}_\gamma(n) \cdot \cos(\tilde{\omega} \cdot n) \quad (3.67)$$

onde



$$\tilde{d}_\gamma(n) = \begin{cases} 2 \cdot \tilde{x}_\gamma(0) + \gamma \cdot \sum_{k=-\infty}^{\infty} \tilde{x}_\gamma^2(k), & n = 0 \\ \tilde{x}_\gamma(n) + \tilde{x}_\gamma(-n) + \gamma \cdot \sum_{k=-\infty}^{\infty} \tilde{x}_\gamma(k+n)\tilde{x}_\gamma(k), & n > 0 \end{cases} \quad (3.68)$$

Em  $\tilde{d}_\gamma(n)$ , as contribuições de  $\tilde{x}_\gamma(m)$  são provenientes de  $\text{Re}[\tilde{X}_\gamma(e^{j\tilde{\omega}})]$  e aquela de sua auto-correlação (termo do somatório) é proveniente de  $|\tilde{X}_\gamma(e^{j\tilde{\omega}})|^2$ . Assim, a curva de  $\ln_\gamma(|X(e^{j\omega(\tilde{\omega})})|^2)$  é diretamente contemplada por uma série de Fourier com coeficientes baseados no mel-cepstrum generalizado  $\tilde{x}_\gamma(n)$ . Quando se assume  $\gamma = 0$  na expressão (3.66), ela se reduz a

$$\ln(|X(e^{j\omega(\tilde{\omega})})|^2) = 2 \cdot \text{Re}[\tilde{X}_0(e^{j\tilde{\omega}})] = 2 \cdot \text{Re}[\ln(\tilde{X}(e^{j\tilde{\omega}}))] \quad (3.69)$$

que representa a magnitude espectral na escala logarítmica. Nesse caso ( $\gamma = 0$ ), a expressão (2.10) assume a forma

$$\ln(|X(e^{j\omega(\tilde{\omega})})|^2) = 2 \cdot \sum_{n=-\infty}^{\infty} \tilde{x}_0(n) \cdot \cos(\tilde{\omega}n) \quad (3.70)$$

Assim, a magnitude espectral mel-logarítmica é representada por uma série trigonométrica ponderada diretamente pelo mel-cepstrum  $\tilde{x}_0(n)$ .

A Figura 3.8 (a) mostra o espectro de fala em dB da emissão da vogal “i” por uma voz masculina<sup>§</sup>. Na Figura 3.8 (b), o espectro se apresenta na escala linear. A magnitude espectral log-gen observada em (d) corresponde à aplicação da função log-gen com parâmetro  $\gamma = -0.2$  (curva da Figura 3.8 (c)) ao espectro na escala linear em (b). As magnitudes espectrais mel-log-gen observadas em (f) e (g) correspondem respectivamente àquelas obtidas pela adoção dos eixos não-lineares de frequências (substituição  $\omega = \omega(\tilde{\omega})$ ) com parâmetros  $\alpha = 0.5$  e  $\alpha = -0.5$  (curvas da Figura 3.8 (e)) para o espectro log-gen da Figura 3.8 (e).

---

<sup>§</sup> O espectro foi obtido a partir de um segmento de 24ms (384 amostras) de um sinal fala com frequência de amostragem de 16kHz. Adotou-se janela de Blackman.

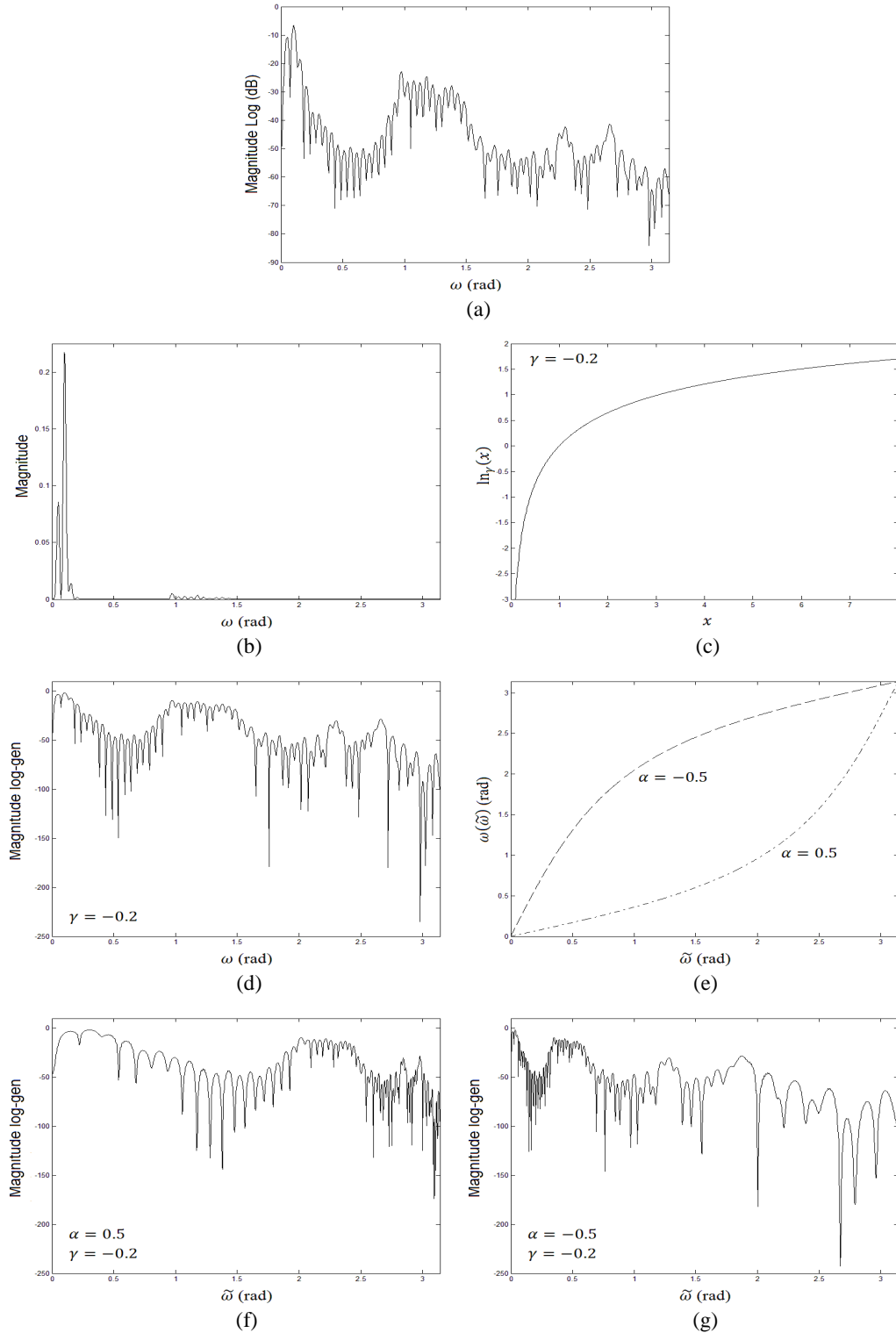


Figura 3.8: (a) Magnitude espectral (dB). (b) Magnitude Espectral (escala linear). (c) Curva da função log-gen de uma variável real com  $\gamma = -0.2$ . (d) Magnitude espectral log-gen com  $\gamma = -0.2$ . (e) Curvas de mapeamento do eixo de freqüências com parâmetros  $\alpha = 0.5$  e  $\alpha = -0.5$ . (f) Magnitude espectral mel-log-gen com  $\alpha = 0.5$  e  $\gamma = -0.2$ . (g) Magnitude espectral mel-log-gen com  $\alpha = -0.5$  e  $\gamma = -0.2$ . Observação: a magnitude log-gen é expressa por  $\ln_\gamma(|\cdot|^2)$ .

### 3.6 Representação Espectral Através do Mel-Cepstrum Generalizado

O espectro mel-log-gen pode ser reconduzido às escalas lineares de magnitude e frequência através da aplicação da função log-gen inversa (exponencial generalizada), e assumindo o mapeamento  $\tilde{z}(z)$  (dado por (3.9)) no lugar da variável independente  $\tilde{z}$ . Assim

$$X(z) = \exp_{\gamma} \left( \tilde{X}_{\gamma}(\tilde{z}(z)) \right)^{\S} \quad (3.71)$$

que é explicitado por

$$X(z) = \exp_{\gamma} \left( \sum_{n=-\infty}^{\infty} \tilde{x}_{\gamma}(n) [\tilde{z}(z)]^{-n} \right) = \begin{cases} \left( 1 + \gamma \cdot \sum_{n=-\infty}^{\infty} \tilde{x}_{\gamma}(n) [\tilde{z}(z)]^{-n} \right)^{1/\gamma}, & 0 < |\gamma| \leq 1 \\ \exp \left( \sum_{n=-\infty}^{\infty} \tilde{x}_{\gamma}(n) [\tilde{z}(z)]^{-n} \right), & \gamma = 0 \end{cases} \quad (3.72)$$

Desta maneira, temos  $X(z)$  representado com base no mel-cepstrum generalizado  $\tilde{x}_{\gamma}(n)$ . Tal representação produz casos especiais cujas formas são conhecidas. Com  $\gamma = -1$ , a expressão (3.15) toma a forma

$$X(z) = \frac{1}{1 - \sum_{n=-\infty}^{\infty} \tilde{x}_{-1}(n) [\tilde{z}(z)]^{-n}} \quad (3.73)$$

que remete ao modelo de pólos (*all-pole*) com resolução não-uniforme baseada na escala mel de frequências (mel-LPC). Com  $\gamma = 1$ , a expressão obtida é

$$X(z) = 1 + \sum_{n=-\infty}^{\infty} \tilde{x}_1(n) [\tilde{z}(z)]^{-n} \quad (3.74)$$

que remete ao modelo de zeros (*all-zero*) com resolução não-uniforme baseada na escala mel de frequências.

---

<sup>\S</sup> Dada a multivalência da função exp-gen complexa, a definição de  $X(z)$  dada por (2.5) é mais restritiva do que aquela dada pela relação (3.71) ou (3.72). Tais relações serão consideradas válidas sob a restrição de que  $\exp_{\gamma} \left( \tilde{X}(\tilde{z}(z)) \right)$  seja uma transformada  $z$  válida de uma seqüência real e estável.

## 3.7 Cálculo do Mel-Cepstrum Generalizado e Suavização Espectral

A presente seção trata do cálculo do mel-cepstrum generalizado e da suavização do espectro logarítmico generalizado. Esta última será abordada por acréscimos ao tratamento dado na seção 2.8 à suavização do cepstrum generalizado. No que se refere ao seu cálculo, o mel-cepstrum-gen não exige acréscimo de condições às seqüências envolvidas além das reunidas no Capítulo 2, para o cepstrum generalizado, e no presente capítulo, para seqüências distorcidas em freqüência. Respeitadas tais condições, o mel-cepstrum-gen pode ser calculado por combinações das transformações log-gen e frequencial. A maneira de combiná-las tem impacto no custo computacional, o que direcionará a abordagem.

Por motivações mais conceituais do que práticas, considerações sobre a aplicação do critério de fase mínima, por meio do mel-cepstrum, são feitas a seguir.

### 3.7.1 Critério de fase mínima com base no Mel-Cepstrum

Considerando  $X(z(\tilde{z})) = \tilde{X}(\tilde{z})$ , em (3.63), o espectro mel-logarítmico ( $\gamma = 0$ ) é dado por

$$\tilde{X}_0(\tilde{z}) = \ln\left(X(z(\tilde{z}))\right) = \ln\left(\tilde{X}(\tilde{z})\right) \quad (3.75)$$

A condição de fase mínima (ver seção 2.6.1) aplicada ao mel-cepstrum  $\tilde{x}_0(n)$ , exige que ele seja causal ou, equivalentemente, que todos os PSDs de  $\ln\left(\tilde{X}(\tilde{z})\right)$  estejam no interior da CRU. Isso garante que  $\tilde{X}(\tilde{z})$  é de fase mínima (pólos e zeros no interior da CRU). Considerando que a região interior à CRU do plano  $\tilde{z}$  é mapeada na região interior à CRU do plano  $z$ , e vice-versa (ver seção 3.2.1), os pólos e zeros de  $\tilde{X}(\tilde{z})$ , no interior da CRU em  $\tilde{z}$ , são pólos e zeros de  $X(z)$  no interior da CRU em  $z$ , portanto a condição que garante que  $\tilde{X}(\tilde{z})$  é de fase mínima garante que  $X(z)$  é de fase mínima. Com isso, o critério de fase mínima (seção (2.6.2)) pode ser aplicado por meio do mel-cepstrum. Nesse caso, a equação (2.66) forneceria o *mel-cepstrum complexo*, se aplicada ao *mel-cepstrum real*, seqüência obtida pela integral em (2.65) considerando seu integrando dado pelo logaritmo do módulo de um espectro distorcido em freqüência.

### 3.7.2 Cálculo de $\tilde{x}_{\gamma_2, \alpha_2}(n)$ a partir de $\tilde{x}_{\gamma_1, \alpha_1}(n)$

Na seção 3.4.3, foi visto que a transformação em frequência tem sua aplicação restrita a seqüências reais causais e estáveis. Já a aplicação da transformação logarítmica generalizada (seção 2.7.4) restringe-se a seqüências reais de fase mínima e cepstrums-gen obtidos sob o critério de fase mínima. Feitas tais restrições, as transformações podem ser utilizadas em qualquer ordem para fornecer o mel-cepstrum-gen. Restrições ocorrem do ponto de vista de complexidade computacional. A presente seção aborda, adotando a via computacional mais econômica, o cálculo do mel-cepstrum-gen [23].

Considerando  $N \gg M$ , onde  $N$  e  $M$  são números inteiros positivos, deseja-se calcular  $M$  amostras do mel-cepstrum generalizado  $\tilde{x}_\gamma(n)$  a partir de  $N$  amostras do cepstrum complexo  $x_0(n)$  (para suavização espectral). Uma vez computado  $x_0(0)$  (ver seção 2.7.5), duas alternativas são possíveis para o cálculo. A primeira consiste em se calcular antes o cepstrum generalizado  $x_\gamma(n)$ , através da transformação logarítmica generalizada (ver seções 2.7.4 e 2.7.5), e depois as amostras do mel-cepstrum generalizado, através da transformação em frequência (ver seção 3.4.3). Considerando que o cálculo de  $M$  amostras transformadas em frequência  $\tilde{x}_\gamma(n)$  exige que se disponha de aproximadamente  $3M$  (para  $\alpha \cong 0.5$ ) amostras de entrada (pelo critério para suavização da seção 3.4.3), é necessário que se compute  $3M$  amostras do cepstrum-gen  $x_\gamma(n)$  através da transformação logarítmica generalizada. A segunda alternativa para o cálculo consiste em calcular antes o mel-cepstrum, através da transformação em frequência, e depois o mel-cepstrum-gen  $\tilde{x}_\gamma(n)$ , através da transformação logarítmica generalizada. Nesse caso, pelas propriedades enunciadas na seção 2.7.4 (para suavização), o cálculo das  $M$  amostras do mel-cepstrum-gen  $\tilde{x}_\gamma(n)$  pela transformação logarítmica exige somente  $M$  amostras do mel-cepstrum  $\tilde{x}_0(n)$  como entrada, o que torna a segunda alternativa para o cálculo, mais econômica computacionalmente. Tais considerações, feitas aqui para o cálculo do mel-cepstrum-gen a partir do cepstrum, estendem-se, naturalmente, para o caso geral do cálculo do mel-cepstrum-gen  $\tilde{x}_{\gamma_2, \alpha_2}(n)$  a partir do mel-cepstrum-gen  $\tilde{x}_{\gamma_1, \alpha_1}(n)$ , com  $(\gamma_2, \alpha_2) \neq (\gamma_1, \alpha_1)$ , o que será abordado pela via de computação mais econômica.

Dada a seqüência  $\tilde{\mathbf{x}}_{\gamma_1, \alpha_1} = [\tilde{x}_{\gamma_1, \alpha_1}(0) \tilde{x}_{\gamma_1, \alpha_1}(1) \tilde{x}_{\gamma_1, \alpha_1}(2) \dots \tilde{x}_{\gamma_1, \alpha_1}(N)]^T$ , obtém-se  $\tilde{\mathbf{x}}_{\gamma_1, \alpha_2} = [\tilde{x}_{\gamma_1, \alpha_2}(0) \tilde{x}_{\gamma_1, \alpha_2}(1) \tilde{x}_{\gamma_1, \alpha_2}(2) \dots \tilde{x}_{\gamma_1, \alpha_2}(M)]^T$  através da transformação

$$\tilde{\mathbf{x}}_{\alpha_2, N_2} = T_\Lambda(\tilde{\mathbf{x}}_{\alpha_1, N_1}, N_1, N_2, \bar{\alpha}) \quad (3.76)$$

Em seguida, determina-se a seqüência normalizada  $\tilde{\mathbf{x}}'_{\gamma_1, \alpha_2} = [\tilde{x}'_{\gamma_1, \alpha_2}(1), \tilde{x}'_{\gamma_1, \alpha_2}(2), \dots, \tilde{x}'_{\gamma_1, \alpha_2}(M)]^T$  onde

$$\tilde{x}'_{\gamma_1, \alpha_2}(n) = \frac{\tilde{x}_{\gamma_1, \alpha_2}(n)}{1 + \gamma \tilde{x}_{\gamma_1, \alpha_2}(0)} \quad (3.77)$$

e, a partir dela, calcula-se  $\tilde{\mathbf{x}}_{\gamma_2, \alpha_2} = [\tilde{x}_{\gamma_2, \alpha_2}(0) \tilde{x}_{\gamma_2, \alpha_2}(1) \tilde{x}_{\gamma_2, \alpha_2}(2) \dots \tilde{x}_{\gamma_2, \alpha_2}(M)]^T$  onde  $\tilde{x}_{\gamma_2, \alpha_2}(n)$  é dado por

$$\tilde{x}_{\gamma_2, \alpha_2}(n) = \begin{cases} \ln_{\gamma_2} \left( \exp_{\gamma_1} \left( \tilde{x}_{\gamma_1, \alpha_2}(0) \right) \right), & n = 0 \\ \left( \left( 1 + \gamma_2 \ln_{\gamma_2} \left( \exp_{\gamma_1} \left( \tilde{x}_{\gamma_1, \alpha_2}(0) \right) \right) \right) \cdot \left( T_{cg}(\tilde{\mathbf{x}}'_{\gamma_1, \alpha_2}, M, M, \gamma_1, \gamma_2) \right)_n \right), & n > 0 \end{cases} \quad (3.78)$$

e  $(\cdot)_n$  indica a  $n$ -ésima componente de um vetor, de maneira que  $\left( T_{cg}(\tilde{\mathbf{x}}'_{\gamma_1, \alpha_2}, M, M, \gamma_1, \gamma_2) \right)_n = (\tilde{\mathbf{x}}'_{\gamma_2, \alpha_2})_n = \tilde{x}'_{\gamma_2, \alpha_2}(n)$ .

### 3.7.3 Suavização Espectral

Nesta seção, como já referido, são feitos acréscimos à abordagem dada à suavização do espectro log-gen na seção 2.8. A suavização é tratada do ponto de vista da descrição da curva do espectro mel-log-gen, e da *liftragem* no domínio da *quefrência*. A envoltória espectral que se obtém é característica de uma análise não-uniforme (seção 3.4.4) aplicada ao espectro mel-log-gen.

Na Figura 3.8, o espectro log-gen ( $\gamma = -0.2$ ) mostrado em (d) é conduzido às curvas de (e) e (f) ao ser tomado nos eixos de freqüências distorcidos com base nos parâmetros  $\alpha = 0.5$  e  $\alpha = -0.5$ . Em (e) ( $\alpha = 0.5$ ), observa-se uma *compressão horizontal* do espectro em freqüências altas, acompanhada por um *alargamento* em freqüências baixas. O contrário ocorre em (f) ( $\alpha = -0.5$ ). A *compressão* do espectro leva sua curva a variar mais rapidamente, enquanto o alargamento leva a variações mais lentas. A aplicação de uma *liftragem* passa-baixas ao mel-cepstrum generalizado (suavização do espectro mel-log-gen), ao selecionar somente parciais de oscilação lenta,

tende a manter bem descritas variações lentas da curva do espectro, não acompanhando variações mais rápidas. Assim, com  $\alpha = 0.5$ , a curva suavizada traz uma descrição mais detalhada do espectro em regiões de baixas frequências, em comparação à sua descrição em regiões de altas frequências. O contrário ocorre para  $\alpha = -0.5$ . A Figura 3.9 (a) e (c) mostra as curvas do espectro mel-log-gen suavizadas associadas aos parâmetros  $\alpha = 0.5$  e  $\alpha = -0.5$ .

Uma vez que a distorção em frequência interfere no caráter periódico da seqüência assim descrita, e o tratamento do SRDS e do LHDS tem base na periodicidade do sinal, é interessante que se explore a *lifragem* no domínio da *mel-quefrência* (domínio mel-cepstral generalizado) observando sua atuação sobre o domínio da *quefrência* (domínio cepstral generalizado). Como foi abordado na seção 3.4.4, a truncagem de uma seqüência distorcida em frequência tem o espectro equivalente àquele obtido pela aplicação de um janelamento de comprimento variável com a frequência sobre a seqüência original. Assim, uma *lifragem* no domínio da *mel-quefrência* equivale a uma *lifragem* em *quefrência* com *quefrência* de corte variável com a frequência de análise. Com base no que foi visto na seção 3.4.4, um janelamento de comprimento  $M$  no mel-cepstrum-gen corresponde, em aproximação, a um janelamento cujo *comprimento equivalente* varia de  $M_- \approx \frac{1-|\alpha|}{1+|\alpha|}M$  a  $M_+ \approx \frac{1+|\alpha|}{1-|\alpha|}M$  (onde  $M_-$  e  $M_+$  assumem valores naturais não-nulos). Assim, o critério para a escolha da *mel-quefrência* de corte  $M$ , na direção de se minimizarem efeitos de vazamento espectral (ver seções 2.8.2 e 2.8.3) na suavização do espectro mel-log-gen, deve ser feita considerando tal intervalo de variação na *quefrência* de corte (domínio do cepstrum-gen).

O retorno do espectro mel-log-gen suavizado às escalas lineares de amplitude e frequência, efetuado pela aplicação mútua da exponencial generalizada e da substituição  $\tilde{\omega} = \tilde{\omega}(\omega)$  ao espectro mel-log-gen, conduz a uma *envoltória espectral característica* (cujo perfil é associado aos valores de  $\alpha$  e  $\gamma$ ). A Figura 3.9 (b) e (d) mostra as *envoltórias espectrais características* obtidas a partir dos espectros mel-log-gen suavizados da Figura 3.9 (a) e (c) (associados aos parâmetros  $(\alpha, \gamma) = (0.5, -0.2)$  e  $(\alpha, \gamma) = (-0.5, -0.2)$ ). A aplicação da condição que remove o viés do espectro logarítmico (ver seção 2.8.3) pode ser estendida ao espectro mel-logarítmico generalizado e leva ao cálculo das  $M$  componentes do mel-cepstrum-gen, de modo a promover um ajuste da envoltória espectral ao espectro original, removendo os efeitos

indesejáveis da *liftragem*. A forma funcional obtida pela recondução do espectro mel-log-gen às escalas lineares de frequência e amplitude leva ao modelo Mel-Cepstral Generalizado (MGC). As componentes do mel-cepstrum-gen assumem o papel de coeficientes de um filtro, sendo denominados de coeficientes Mel-Cepstrais Generalizados (MGC). A análise MGC [2][1], que assume diferentes características de acordo com a escolha dos parâmetros  $\alpha$  e  $\gamma$ , tem as análises Mel-Cepstral[7][6] (também chamada de *ML based Mel-Cepstral Analysis*), Mel-LPC [8], Cepstral [5][4] (UELS) e LPC [3], como casos particulares. O modelo tem solução única no intervalo  $-1 \leq \gamma \leq 0$ , sendo o método para sua solução abordado no capítulo a seguir.

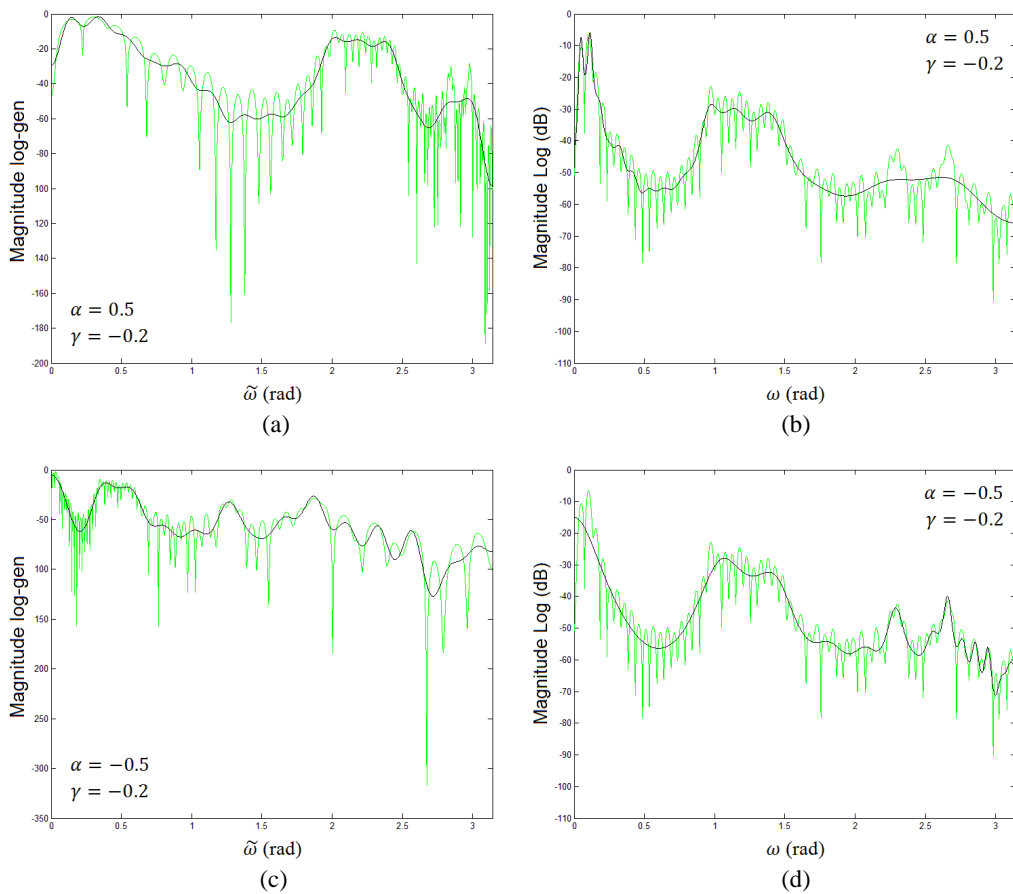


Figura 3.9: (a) (c) Espectros mel-log-gen suavizados (ordem 30) com base nos espectros mel-log-gen associados a  $(\alpha, \gamma) = (0.5, -0.2)$  e  $(\alpha, \gamma) = (-0.5, -0.2)$ . (b) (d) Magnitudes (dB) das envoltórias espectrais características associadas a  $(\alpha, \gamma) = (0.5, -0.2)$  e  $(\alpha, \gamma) = (-0.5, -0.2)$  sobrepostas ao espectro original (em verde). *Observação: a magnitude log-gen é expressa por  $\ln_{\gamma}(|\cdot|^2)$ .*



## Capítulo 4

# Modelo Mel-Cepstral Generalizado (MGC)

### 4.1 Introdução

Os capítulos 2 e 3 trataram do mel-cepstrum generalizado (e o cepstrum generalizado) no contexto do espectro mel-log-gen (e log-gen), abordando os problemas de uma suavização linear para a obtenção de uma envoltória espectral de fala. Nesse contexto, a análise Mel-Cepstral Generalizada (e Cepstral Generalizada) resolve tais problemas ao conduzir uma suavização não-linear, levando a uma descrição superior do espectro pela envoltória. A análise Mel-Cepstral Generalizada é efetuada pelo cálculo da solução do modelo Mel-Cepstral Generalizado (MGC) de ordem  $M$ , para o qual são determinados  $M$  parâmetros, denominados coeficientes MGC.

A análise Mel-Cepstral Generalizada [1][2] (MGC – *Mel-Generalized Cepstral*) é uma abordagem unificada para estimação de envoltória espectral de fala, que tem as análises LPC [3] (predição linear), Cepstral [4][5] (UELS – *Unbiased Estimation of Log-Spectrum*), Mel-Cepstral [6][7] e Mel-LPC [8] como casos particulares. Tais análises são obtidas dentre uma diversidade contínua de modelos pela atribuição de valores aos parâmetros  $\alpha$  e  $\gamma$ . O parâmetro  $\alpha$ , que assume valores no intervalo real  $]-1,1[$ , permite modificar a maneira como a resolução da análise é distribuída ao longo do eixo de frequências do espectro. Já o parâmetro  $\gamma$ , que assume valores no intervalo

real  $[-1,1]$ , proporciona, pelo modelo, diferentes pesos à descrição de picos e vales espectrais. Para  $(\alpha, \gamma) = (0, -1)$ , a análise MGC se reduz à análise LPC (modelo de pólos - *all-pole*). Já para  $(\alpha, \gamma) = (0,0)$ , obtém-se a análise Cepstral (UELS). O modelo restrito a  $\alpha = 0$ , podendo assumir qualquer valor no intervalo  $\gamma = [-1,1]$ <sup>§</sup>, é denominado de modelo Cepstral Generalizado (GC) [9]. Com  $|\alpha| < 1$ , são obtidas as versões distorcidas em frequência dos casos mencionados. Destacam-se a análise Mel-LPC ( $|\alpha| < 1, \gamma = -1$ ) e a análise Mel-Cepstral ( $|\alpha| < 1, \gamma = 0$ ).

A solução para o modelo MGC é única no intervalo  $-1 \leq \gamma \leq 0$ , e é obtida com base na minimização de um critério. Seu cálculo é efetuado por um método iterativo (Newton-Raphson) de convergência quadrática onde, a cada iteração, são calculados o gradiente e a matriz Hessiana do critério espectral. A formulação para os cálculos é tratada para cada caso especial, o que é exigido pela diferenciação de suas metodologias no cálculo computacional. Os casos GC ( $\alpha = 0, \gamma \neq 0$ ) e Cepstral ( $\alpha = 0, \gamma = 0$ ) exigem somente o cálculo de FFTs, a cada iteração do método de Newton-Raphson [9][5]. Já os casos MGC ( $\alpha \neq 0, \gamma \neq 0$ ) e Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ) exigem, além do cálculo de FFTs, o uso de fórmulas recursivas de transformação em frequência [2][7], sendo tratados em uma seção dedicada.

São discutidos os perfis do modelo MGC e o ajuste de suas curvas a diferentes características do espectro de fala. O emprego dos parâmetros MGC em síntese paramétrica estatística, também é abordado.

Um método alternativo para a solução do modelo MGC (e Mel-Cepstral) é proposto. Tal método diferencia-se daquele de [2][7], por contar somente com o cálculo de FFTs a cada iteração do método de Newton-Raphson, sem necessidade do uso de fórmulas recursivas. O método exige que se disponha do periodograma discreto tomado no eixo não-linear de frequências, e faz uso de uma interpolação linear do periodograma regular para o seu cálculo. As análises e experimentos relacionados ao método proposto encontram-se em estágio inicial e devem ser completados no sentido de se identificar a relação *ganho computacional*  $\times$  *qualidade da representação*.

Destaca-se que a formulação apresentada no capítulo, para os cálculos da solução do modelo MGC, é toda demonstrada por meio de Apêndices.

---

<sup>§</sup> Apesar de  $\gamma$  assumir valores no intervalo mencionado, o modelo MGC tem solução única somente para o intervalo  $\gamma = [-1,0]$ .

## 4.2 Preliminares: Análise Cepstral Generalizada

### 4.2.1 Modelo

O espectro log-gen de uma seqüência  $x(n)$  com espectro  $X(e^{j\omega})$  é dado por

$$X_\gamma(e^{j\omega}) = \ln_\gamma(X(e^{j\omega})) = \sum_{n=-\infty}^{\infty} x_\gamma(n)e^{-j\omega n} \quad (4.1)$$

onde  $x_\gamma(n)$  é o cepstrum generalizado. Com base em (2.15), a representação do espectro  $X(e^{j\omega})$  em termos do cepstrum generalizado  $x_\gamma(n)$  é dada por

$$X(e^{j\omega}) = \exp_\gamma\left(\sum_{n=-\infty}^{\infty} x_\gamma(n)e^{-j\omega n}\right) \quad (4.2)$$

O modelo para análise cepstral generalizada de ordem  $M$  é dado por

$$H(z) = \exp_\gamma\left(\sum_{n=0}^M c_\gamma(n)z^{-n}\right) \quad (4.3)$$

onde  $c_\gamma(0), c_\gamma(1), c_\gamma(3), \dots, c_\gamma(M)$  são os coeficientes cepstrais generalizados. Em geral o parâmetro  $\gamma$  pode assumir qualquer valor no intervalo  $[-1, +1]$ . Porém, uma vez que para  $0 < \gamma \leq 1$ , o método aqui descrito não apresenta solução única, serão admitidos valores de  $\gamma$  somente no intervalo  $[-1, 0]$ .

### 4.2.2 Normalização

Será útil que se represente  $H(z)$  pelo produto do filtro normalizado por uma constante de ganho. Assim

$$H(z) = K \cdot D(z) \quad (4.4)$$

onde a constante de ganho é dada por

$$K = \exp_\gamma(c_\gamma(0)) \quad (4.5)$$

e o modelo normalizado, por

$$D(z) = \exp_{\gamma} \left( \sum_{n=1}^M c_{\gamma}'(n) z^{-n} \right) \quad (4.6)$$

onde

$$c_{\gamma}'(n) = \frac{c_{\gamma}(n)}{1 + \gamma c_{\gamma}(0)} \quad (4.7)$$

e correspondem aos coeficientes GC normalizados

### Demonstração

Como  $H(z)$  é causal, os termos com atraso podem ser eliminados do interior de  $\exp_{\gamma}(\cdot)$ , aplicando-se o limite  $\lim_{|z| \rightarrow \infty} H(z)$ . Tal operação fornece a *zero-ésima* amostra da resposta ao impulso de  $H(z)$  que é a constante de ganho  $K$  do modelo. Assim

$$\lim_{|z| \rightarrow \infty} H(z) = K \quad (4.8)$$

Portanto (4.5) é obtida através de

$$K = \lim_{|z| \rightarrow \infty} H(z) = \lim_{|z| \rightarrow \infty} \left( \exp_{\gamma} \left( \sum_{n=0}^M c_{\gamma}(n) z^{-n} \right) \right) = \exp_{\gamma} (c_{\gamma}(0)) \quad (4.9)$$

O filtro normalizado, dado pela expressão (4.6), é obtido por

$$D(z) = \frac{H(z)}{K} = \frac{\exp_{\gamma}(\sum_{n=0}^M c_{\gamma}(n) z^{-n})}{\exp_{\gamma}(c_{\gamma}(0))} \quad (4.10)$$

onde, considerando a relação (A.17) (Apêndice A) para quocientes entre exponenciais generalizadas, temos

$$D(z) = \exp_{\gamma} \left( \sum_{n=1}^M \frac{c_{\gamma}(n)}{1 + \gamma \cdot c_{\gamma}(0)} z^{-n} \right) \quad (4.11)$$

Assim a expressão (4.6) representa a expressão para o filtro normalizado  $D(z)$ .

No lugar dos coeficientes  $c_{\gamma}(n)$ , em geral trabalha-se com a constante  $K$  e os coeficientes normalizados  $c_{\gamma}'(n)$  ( $n \geq 1$ ), sendo  $H(z)$  expresso por

$$H(z) = K \cdot \exp_{\gamma} \left( \sum_{n=1}^M c_{\gamma}'(n) z^{-n} \right) \quad (4.12)$$

Uma vez necessário, os coeficientes  $c_{\gamma}(n)$  são recuperados através de

$$c_{\gamma}(n) = c_{\gamma}'(n) (1 + \gamma c_{\gamma}(0)) \quad (4.13)$$

para  $n \geq 1$ , e o coeficiente  $c_{\gamma}(0)$  é obtido através de

$$c_{\gamma}(0) = \ln_{\gamma}(K) \quad (4.14)$$

As análises GC com  $\gamma = 0$  e  $\gamma = -1$  correspondem, respectivamente às análises Cepstral (UELS) e LPC, casos tratados a seguir.

### 4.2.3 Caso Cepstral ( $\gamma = 0$ )

Para  $\gamma = 0$ , a exponencial generalizada se reduz à exponencial padrão e o modelo  $H(z)$  é então dado pela expressão

$$H(z) = \exp_0 \left( \sum_{n=0}^M c_0(n) z^{-n} \right) = e^{\sum_{n=0}^M c_0(n) z^{-n}} \quad (4.15)$$

que é o modelo para a análise cepstral. Sua representação pelo produto da constante de ganho pelo filtro normalizado é dada por

$$H(z) = K \cdot D(z) = K \cdot e^{\sum_{n=1}^M c_0(n) z^{-n}} \quad (4.16)$$

onde  $K = e^{c_0(0)}$ , fator multiplicativo obtido simplesmente colocando o termo constante para fora do somatório em (4.15) (ou por (4.9)). Observa-se que os coeficientes Cepstrais  $c_0(n)$  ( $n \geq 1$ ) não são afetados na normalização, por não carregarem informação de ganho. Isso pode ser verificado assumindo  $\gamma = 0$  em (4.7), quando vale a igualdade  $c_0'(n) = c_0(n)$  ( $n \geq 1$ ).

### 4.2.4 Caso LPC ( $\gamma = -1$ )

Para  $\gamma = -1$ , a expressão para o filtro normalizado assume a forma

$$D(z) = \exp_{-1} \left( \sum_{n=1}^M c_{-1}'(n) z^{-n} \right) = \frac{1}{1 - \sum_{n=1}^M c_{-1}'(n) z^{-n}} \quad (4.17)$$

que representa um filtro de predição linear de ordem  $M$  com coeficientes LP dados por  $c_{-1}'(1), c_{-1}'(2), \dots, c_{-1}'(M)$ . O filtro  $H(z)$  é portanto dado por

$$H(z) = \frac{K}{1 - \sum_{n=1}^M c_{-1}'(n) z^{-n}} \quad (4.18)$$

onde a constante de ganho é dada por  $K = 1/(1 - c_{-1}(0))$  (com base em (4.5)).

## 4.3 Análise Mel-Cepstral Generalizada

### 4.3.1 Modelo

Representando o espectro mel-log-gen de uma seqüência  $x(n)$  cujo espectro é dado por  $X(e^{j\omega}) = \tilde{X}_\gamma(e^{j\tilde{\omega}(\omega)})$

$$\tilde{X}_\gamma(e^{j\tilde{\omega}}) = \ln_\gamma \left( X(e^{j\omega(\tilde{\omega})}) \right) = \sum_{n=-\infty}^{\infty} \tilde{x}_\gamma(n) e^{-j\tilde{\omega}n} \quad (4.19)$$

onde  $\tilde{x}_\gamma(n)$  é o mel-cepstrum generalizado. Com base em (2.15), a representação do espectro  $X(e^{j\omega})$  em termos do mel-cepstrum generalizado  $\tilde{x}_\gamma(n)$  é dada por

$$X(e^{j\omega}) = \exp_\gamma \left( \sum_{n=-\infty}^{\infty} \tilde{x}_\gamma(n) e^{-j\tilde{\omega}(\omega)n} \right) \quad (4.20)$$

onde, com base na expressão (3.13)

$$e^{-j\tilde{\omega}(\omega)} = \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \quad (4.21)$$

O modelo para análise mel-cepstral generalizada de ordem  $M$  é dado por

$$H(z) = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) [\tilde{z}(z)]^{-n} \right) \quad (4.22)$$

onde  $\tilde{c}_\gamma(0), \tilde{c}_\gamma(1), \tilde{c}_\gamma(3), \dots, \tilde{c}_\gamma(M)$  são os coeficientes mel-cepstrais generalizados. Com base na expressão (3.9),  $[\tilde{z}(z)]^{-1}$  é dado por

$$[\tilde{z}(z)]^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (4.23)$$

onde se vê que, para  $\alpha = 0$ ,  $[\tilde{z}(z)]^{-1} = z^{-1}$  e o modelo MGC se reduz ao modelo GC.

No modelo GC, a constante de ganho dada por (4.5) é função somente do coeficiente  $c_\gamma(0)$ . No caso MGC, a constante de ganho é dada por

$$K = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n \right) \quad (4.24)$$

onde se vê que  $K$  depende de todos os coeficientes  $\tilde{c}_\gamma(n)$ . Para a aplicação do método para o cálculo dos coeficientes MGC, o ganho deve estar associado somente ao zero-ésimo coeficiente do modelo. Uma representação alternativa para o modelo (4.25) faz uso de um conjunto de coeficientes transformados que satisfaz tal exigência. Tais coeficientes são obtidos por uma transformação linear aplicada aos coeficientes MGC. A próxima seção aborda o modelo  $H(z)$  com base nos coeficientes transformados.

#### Demonstração de (4.24)

Como  $H(z)$  é causal, os termos com atraso podem ser eliminados do interior de  $\exp_\gamma(\cdot)$ , aplicando-se o limite  $\lim_{|z| \rightarrow \infty} H(z)$ , o que fornece a constante de ganho  $K$  do modelo. Assim

$$K = \lim_{|z| \rightarrow \infty} H(z) = \lim_{|z| \rightarrow \infty} \left[ \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) [\tilde{z}(z)]^{-n} \right) \right] = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n \right) \quad (4.25)$$

já que

$$\lim_{|z| \rightarrow \infty} [\tilde{z}(z)]^{-1} = \lim_{|z| \rightarrow \infty} \left( \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \right) = -\alpha \quad (4.26)$$

### 4.3.2 Modelo Baseado nos Coeficientes MGC Transformados $b_\gamma(n)$

O modelo para a análise MGC pode ser expresso por

$$H(z) = \exp_\gamma \left( \sum_{n=0}^M b_\gamma(n) \Phi_n(z) \right) \quad (4.27)$$

onde

$$\Phi_n(z) = \begin{cases} 1 & n = 0 \\ \frac{(1 - \alpha^2)z^{-1}}{1 - \alpha z^{-1}} [\tilde{z}(z)]^{-(n-1)} & n > 0 \end{cases} \quad (4.28)$$

O modelo MGC em tal forma tem  $b_\gamma(0)$  como único termo livre de atraso no somatório em (4.27). Da mesma maneira que o coeficiente  $c_\gamma(0)$  no caso cepstral,  $b_\gamma(0)$  é o coeficiente associado ao ganho no modelo. Com  $\mathbf{b}_\gamma = [b_\gamma(0) b_\gamma(1) b_\gamma(2) \dots b_\gamma(M)]^T$  e  $\tilde{\mathbf{c}}_\gamma = [\tilde{c}_\gamma(0) \tilde{c}_\gamma(1) \tilde{c}_\gamma(2) \dots \tilde{c}_\gamma(M)]^T$  os coeficientes  $b_\gamma(n)$  são obtidos, com base nos coeficientes  $\tilde{c}_\gamma(n)$ , pela transformação linear

$$\mathbf{b}_\gamma = \mathbf{A}_{M+1} \tilde{\mathbf{c}}_\gamma \quad (4.29)$$

e os coeficientes  $\tilde{c}_\gamma(n)$  são obtidos a partir dos coeficientes  $b_\gamma(n)$  pela transformação

$$\tilde{\mathbf{c}}_\gamma = \mathbf{A}_{M+1}^{-1} \mathbf{b}_\gamma \quad (4.30)$$

As matrizes  $\mathbf{A}_{M+1}$  e  $\mathbf{A}_{M+1}^{-1}$ , de dimensão  $(M+1) \times (M+1)$ , são dadas por

$$\mathbf{A}_{M+1} = \begin{bmatrix} 1 & (-\alpha)^1 & (-\alpha)^2 & (-\alpha)^3 & \dots & (-\alpha)^M \\ 0 & 1 & (-\alpha)^1 & (-\alpha)^2 & \dots & (-\alpha)^{M-1} \\ 0 & 0 & 1 & (-\alpha)^1 & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \ddots & (-\alpha)^2 \\ \vdots & \vdots & \ddots & \ddots & \ddots & (-\alpha)^1 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \quad (4.31)$$

e



$$\mathbf{A}_{M+1}^{-1} = \begin{bmatrix} 1 & \alpha & 0 & 0 & \dots & 0 \\ 0 & 1 & \alpha & 0 & \dots & 0 \\ 0 & 0 & 1 & \alpha & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \alpha \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \quad (4.32)$$

O tr nsito entre os coeficientes  $b_\gamma(n)$  e  $\tilde{c}_\gamma(n)$  pode tamb m ser efetuado atrav s das f rmulas

$$b_\gamma(n) = \begin{cases} \tilde{c}_\gamma(n) & n = M \\ \tilde{c}_\gamma(n) - \alpha b_\gamma(n+1) & 0 \leq n < M \end{cases} \quad (4.33)$$

$$\tilde{c}_\gamma(n) = \begin{cases} b_\gamma(n) & n = M \\ b_\gamma(n) + \alpha b_\gamma(n+1) & 0 \leq n < M \end{cases} \quad (4.34)$$

O Ap ndice C mostra as dedu es do modelo MGC na forma (4.27) e das transforma es acima.

Da mesma maneira que para  $\tilde{c}_\gamma(n)$ , note pelas express es (4.33) ou (4.34) que, com  $\alpha = 0$ , os coeficientes MGC transformados  $b_\gamma(n)$  se reduzem aos coeficientes GC  $c_\gamma(n)$  e, naturalmente, o modelo  $H(z)$  na forma (4.27) se reduz ao modelo GC.

### 4.3.3 Normaliza o

A constante de ganho do modelo  $H(z)$  na forma (4.27)   dada por

$$K = \exp_\gamma(b_\gamma(0)) \quad (4.35)$$

e o modelo normalizado   expresso por

$$D(z) = \exp_\gamma\left(\sum_{n=1}^M b_\gamma'(n)\Phi_n(z)\right) \quad (4.36)$$

onde

$$b_\gamma'(n) = \frac{b_\gamma(n)}{1 + \gamma b_\gamma(0)} \quad (4.37)$$

e correspondem aos coeficientes MGC transformados normalizados.

### Demonstração

Como visto em (4.8), o limite  $\lim_{|z| \rightarrow \infty} H(z)$  elimina os termos com atraso do interior de  $\exp_\gamma(\cdot)$  e fornece a constante de ganho  $K$  do modelo. Assim, a expressão (4.35) é obtida por

$$K = \lim_{|z| \rightarrow \infty} H(z) = \lim_{|z| \rightarrow \infty} \left( \exp_\gamma \left( \sum_{n=0}^M b_\gamma(n) \Phi_n(z) \right) \right) = \exp_\gamma(b_\gamma(0)) \quad (4.38)$$

válido uma vez que, para  $n \geq 1$

$$\lim_{|z| \rightarrow \infty} \Phi_n(z) = \lim_{|z| \rightarrow \infty} \left( \frac{(1 - \alpha^2)z^{-1}}{1 - \alpha z^{-1}} [\bar{z}(z)]^{-(n-1)} \right) = 0 \quad (4.39)$$

O filtro normalizado, dado pela expressão (4.36) é obtido por

$$D(z) = \frac{H(z)}{K} = \frac{\exp_\gamma(\sum_{n=0}^M b_\gamma(n) \Phi_n(z))}{\exp_\gamma(b_\gamma(0))} \quad (4.40)$$

onde considerando a relação (A.17) (Apêndice A) para quocientes entre exponenciais generalizadas, temos

$$D(z) = \exp_\gamma \left( \sum_{n=1}^M \frac{b_\gamma(n)}{1 + \gamma \cdot b_\gamma(0)} \Phi_n(z) \right) \quad (4.41)$$

e, portanto, (4.36) representa a expressão para o filtro normalizado  $D(z)$ .

No lugar dos coeficientes  $b_\gamma(n)$ , em geral trabalha-se com a constante  $K$  e os coeficientes normalizados  $b_\gamma'(n)$  ( $n \geq 1$ ), sendo  $H(z)$  expresso por

$$H(z) = K \cdot \exp_\gamma \left( \sum_{n=1}^M b_\gamma'(n) \Phi_n(z) \right) \quad (4.42)$$

Uma vez necessário, os coeficientes  $b_\gamma(n)$  são recuperados através de

$$b_\gamma(n) = b_\gamma'(n) (1 + \gamma b_\gamma(0)) \quad (4.43)$$

para  $n \geq 1$ , e o coeficiente  $b_\gamma(0)$  é obtido através de

$$b_\gamma(0) = \ln_\gamma(K) \quad (4.44)$$

As análises MGC com  $\gamma = 0$  e  $\gamma = -1$  correspondem, respectivamente, às análises Mel-Cepstral e Mel-LPC, casos tratados a seguir.

#### 4.3.4 Caso Mel-Cepstral ( $\gamma = 0$ )

Para  $\gamma = 0$ , a exponencial generalizada se reduz à exponencial padrão e o filtro  $H(z)$  é então dado pela expressão

$$H(z) = \exp\left(\sum_{n=0}^M b_0(n)\Phi_n(z)\right) = e^{\sum_{n=0}^M b_0(n)\Phi_n(z)} \quad (4.45)$$

que é o modelo para a análise Mel-Cepstral. Sua representação pelo produto da constante de ganho pelo filtro normalizado é dada por

$$H(z) = K \cdot D(z) = K \cdot e^{\sum_{n=1}^M b_0(n)\Phi_n(z)} \quad (4.46)$$

onde  $K = e^{b_0(0)}$ , fator multiplicativo obtido simplesmente colocando o termo constante para fora do somatório em (4.45) (ou por (4.38)). Observa-se que os coeficientes Mel-Cepstrais transformados  $b_0(n)$  ( $n \geq 1$ ) não são afetados na normalização, por não carregarem informação de ganho. Isso pode ser verificado assumindo  $\gamma = 0$  em (4.37), quando vale a igualdade  $b_0'(n) = b_0(n)$  ( $n \geq 1$ ).

#### 4.3.5 Caso Mel-LPC ( $\gamma = -1$ )

Para  $\gamma = -1$ , a expressão para o modelo normalizado assume a forma

$$D(z) = \exp_{-1}\left(\sum_{n=1}^M b_{-1}'(n)\Phi_n(z)\right) = \frac{1}{1 - \sum_{n=1}^M b_{-1}'(n)\Phi_n(z)} \quad (4.47)$$

que representa um filtro de predição linear distorcido em frequência de ordem  $M$  com coeficientes Mel-LPC transformados dados por  $b_{-1}'(1), b_{-1}'(2), \dots, b_{-1}'(M)$ . O modelo  $H(z)$  é portanto dado por

$$H(z) = \frac{K}{1 - \sum_{n=1}^M b_{-1}'(n)\Phi_n(z)} \quad (4.48)$$

onde a constante de ganho é dada por  $K = 1/(1 - b_{-1}(0))$  (com base em (4.35)).

## 4.4 Critério Espectral

Para que sejam obtidos os coeficientes do Modelo MGC, deve-se minimizar o seguinte critério

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\exp(R(\omega)) - R(\omega) - 1] d\omega \quad (4.49)$$

onde

$$R(\omega) = \ln(I_N(\omega)) - \ln|H(e^{j\omega})|^2 \quad (4.50)$$

com

$$I_N(\omega) = \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-j\omega n} \right|^2 / \sum_{n=0}^{N-1} w^2(n) \quad (4.51)$$

que corresponde ao periodograma modificado, uma estimação da densidade espectral de potência do quadro  $x(n)$  do sinal de fala em análise.  $w(n)$  é uma janela de comprimento  $N$  (ex: janela de Blackman). A minimização de (4.49) fornece os coeficientes do modelo MGC. A solução do modelo cepstral ( $\gamma = 0$ ,  $\alpha = 0$ ) com base em tal critério é denominada UELS [4] (*Unbiased Estimation of Log Spectrum*). No contexto da análise cepstral tradicional, tal método se constitui em uma suavização não-linear, que leva à remoção do viés do espectro logarítmico linearmente suavizado. O critério tem a mesma forma daquele para a estimação da máxima verossimilhança (ML-*Maximum Likelihood*) de um processo AR Gaussiano estacionário [24][25]. O método se estende aos modelos Mel-Cepstral [18], GC [9] e MGC [7][1].

Minimizar  $E$  com respeito aos coeficientes MGC  $[b_\gamma(0) \ b_\gamma(1) \ b_\gamma(2) \ \dots \ b_\gamma(M)]^T$  equivale a minimizar

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (4.52)$$

com respeito a  $\mathbf{b}'_\gamma = [b'_\gamma(1) \ b'_\gamma(2) \ \dots \ b'_\gamma(M)]^T$ , através de

$$\nabla_{\mathbf{b}'}\varepsilon = 0 \quad (4.53)$$

(onde  $\nabla_{\mathbf{b}'} = \left[ \frac{\partial}{\partial b'_\gamma(1)} \quad \frac{\partial}{\partial b'_\gamma(2)} \quad \dots \quad \frac{\partial}{\partial b'_\gamma(M)} \right]^T$ ) e minimizar  $E$  com respeito a  $K$ , através de  $\partial E/\partial K = 0$ , o que fornece

$$K = \sqrt{\varepsilon_{min}} \quad (4.54)$$

onde  $\varepsilon_{min}$  é o valor de  $\varepsilon$  minimizado através de (4.53).

$\varepsilon$ , dado pela expressão (4.52), pode ser interpretado como a média quadrática do erro de predição da saída do filtro inverso mostrado na Figura 4.1. Assim, minimizar  $\varepsilon$  é o mesmo que minimizar  $E[e^2(n)]$  (onde  $E[\cdot]$  é o operador média ou valor esperado) para a determinação dos coeficientes do filtro  $D(z)$ .

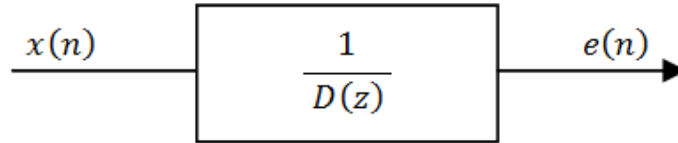


Figura 4.1: Filtro inverso:  $D(z)$  é determinado minimizando-se  $E[e^2(n)]$

Para a solução de (4.53) utiliza-se o método de Newton-Raphson, abordado na seção 4.5.

Observa-se que essa seção tratou o critério de minimização com base nos coeficientes  $b'_\gamma(k)$ , já que estes, por representarem o caso de análise mais geral (MGC), englobam implicitamente os coeficientes  $b'_0(k)$ ,  $c'_\gamma(k)$  e  $c'_0(k)$  associados, respectivamente, aos casos Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ), GC ( $\alpha = 0, \gamma \neq 0$ ) e Cepstral ( $\alpha = 0, \gamma = 0$ ). As próximas seções seguirão esse mesmo padrão, utilizando as notações específicas, quando requerido.

#### 4.4.1 Unicidade e Estabilidade da Solução

A equação (4.53) tem solução única para  $-1 \leq \gamma \leq 0$  já que, neste intervalo,  $\varepsilon$  é convexo com respeito a  $\mathbf{b}'_\gamma$ . A convexidade de  $\varepsilon$  é demonstrada em [2].

Dado que temos solução única no intervalo mencionado, é necessário que se garanta que ela conduz a um modelo estável. Para os casos Cepstral e Mel-Cepstral ( $\gamma = 0$ ), a estabilidade é garantida uma vez que os coeficientes do modelo se restringem a seqüências de Cepstrum (ou Mel-Cepstrum) reais, estáveis e causais. Isso corresponde

à própria condição de fase mínima (ver seção 2.6.1) para  $H(z)$  que, portanto, tem estabilidade garantida. A estabilidade da solução também é garantida para  $-1 \leq \gamma < 0$  (modelos GC e MGC), o que é demonstrado em [2].

#### 4.4.2 Computação do Periodograma

O cálculo do periodograma é efetuado da seguinte forma. Calcula-se a FFT de  $N_F$  pontos

$$P(\omega_k) = \sum_{n=0}^{N_F-1} p(n)e^{-j\omega_k n} \quad \omega_k = \frac{2\pi}{N_F}k \quad k = 0, 1, \dots, N_F - 1 \quad (4.55)$$

onde  $p(n)$  é dado por

$$p(n) = \begin{cases} w(n)x(n), & n = 0, 1, \dots, N - 1 \\ 0, & n = N, N + 1, \dots, N_F - 1 \end{cases} \quad (4.56)$$

que corresponde a um bloco janelado do sinal de fala estendido com amostras nulas para assumir o comprimento  $N_F$ , dado pela menor potência de 2 de valor igual ou superior a  $N$  (requisito para o cálculo de FFTs [17]). O periodograma amostrado é, então, obtido por

$$I_N(\omega_k) = \frac{1}{\sum_{n=0}^{N-1} w^2(n)} |P(\omega_k)|^2, \quad \omega_k = \frac{2\pi}{N_F}k \quad k = 0, 1, \dots, N_F - 1 \quad (4.57)$$

Nota-se que, para  $w(n)$ , é aconselhável a adoção de uma janela de transição suave, com as janelas de Blackman e Hanning.

Nas próximas seções, a formulação, mesmo no âmbito de seu tratamento computacional, manterá a notação com base no eixo contínuo de frequências. Quando forem requeridos cálculos de transformadas via FFT com base no periodograma, considera-se implícita a FFT de  $N_F$  pontos calculada a partir de  $I_N(\omega_k)$ , dado por (4.57).

## 4.5 Método de Newton-Raphson para a Solução do Modelo

Esta seção descreve o método de Newton-Raphson, utilizado para a solução de (4.53), e a formulação para o cálculo do gradiente e da Hessiana de  $\varepsilon$ , exigido a cada iteração do método. Uma única formulação, aquela relativa à análise MGC, poderia ser abordada já que, com as devidas atribuições de valores a  $\alpha$  e  $\gamma$ , ela se reduz aos casos Mel-Cepstral ( $\gamma = 0$ ), GC ( $\alpha = 0$ ) e Cepstral ( $\alpha = 0, \gamma = 0$ ). Uma vez que tais reduções conduzem a importantes particularidades no cálculo computacional de cada caso, suas formulações serão abordadas individualmente. Serão tratadas, portanto, as análises: MGC, representando o caso em que  $\alpha \neq 0$  e  $\gamma \neq 0$ , Mel-Cepstral, em que  $\alpha \neq 0$  e  $\gamma = 0$ , GC, em que  $\alpha = 0$  e  $\gamma \neq 0$ , e Cepstral, em que  $\alpha = 0$  e  $\gamma = 0$  (os parâmetros se restringem aos domínios  $\alpha = ]-1,1[$  e  $\gamma = [-1,0]$ ). As formulações apresentadas nas seções 4.5.2 - 4.5.5 são baseadas nas demonstrações do Apêndice D.

O cálculo computacional da solução será abordado na presente seção somente para os casos em que  $\alpha = 0$  (GC e Cepstral). Por serem mais complexas, as computações para os casos MGC e Mel-Cepstral serão tratadas na seção 4.6 e também na seção 4.8, onde uma abordagem alternativa é apresentada.

### 4.5.1 Método de Newton-Raphson

Uma vez que  $\varepsilon$  tem um único mínimo global, a aplicação do método de Newton-Raphson para a solução de  $\nabla_{\mathbf{b}'_{\gamma}} \varepsilon = 0$  tem convergência garantida. Através do método, em sua  $i$ -ésima iteração, encontra-se a raiz da equação que corresponde à aproximação de primeira ordem (linear) de  $\nabla_{\mathbf{b}'_{\gamma}} \varepsilon$  no ponto  $\mathbf{b}'_{\gamma} = \mathbf{b}'_{\gamma}{}^{(i)}$ , o que determina um novo ponto  $\mathbf{b}'_{\gamma} = \mathbf{b}'_{\gamma}{}^{(i+1)}$  mais próximo da solução que satisfaz  $\nabla_{\mathbf{b}'_{\gamma}} \varepsilon = 0$ . Sucessivas aplicações de tal operação fornecem  $\mathbf{b}'_{\gamma}$  tão próximo quanto se queira de uma solução exata. A aproximação de primeira ordem de  $\nabla_{\mathbf{b}'_{\gamma}} \varepsilon$  é dada por

$$\nabla_{\mathbf{b}'_{\gamma}} \varepsilon \approx \nabla_{\mathbf{b}'_{\gamma}} \varepsilon \Big|_{\mathbf{b}'_{\gamma} = \mathbf{b}'_{\gamma}{}^{(i)}} + \mathbf{H} \Big|_{\mathbf{b}'_{\gamma} = \mathbf{b}'_{\gamma}{}^{(i)}} \Delta \mathbf{b}'_{\gamma}{}^{(i)} \quad (4.58)$$

A raiz de (4.58) é obtida resolvendo-se, para  $\Delta \mathbf{b}'_{\gamma}{}^{(i)}$ , o sistema linear

$$\mathbf{H}|_{\mathbf{b}'_\gamma = \mathbf{b}'_\gamma(i)} \Delta \mathbf{b}'_\gamma(i) = -\nabla_{\mathbf{b}'_\gamma} \varepsilon |_{\mathbf{b}'_\gamma = \mathbf{b}'_\gamma(i)} \quad (4.59)$$

e  $\mathbf{b}'_\gamma(i)$  é atualizado, fornecendo  $\mathbf{b}'_\gamma(i+1)$ , através de

$$\mathbf{b}'_\gamma(i+1) = \mathbf{b}'_\gamma(i) + \Delta \mathbf{b}'_\gamma(i) \quad (4.60)$$

Nas próximas seções, para  $\nabla_{\mathbf{b}'_\gamma} \varepsilon |_{\mathbf{b}'_\gamma = \mathbf{b}'_\gamma(i)}$  e  $\mathbf{H}|_{\mathbf{b}'_\gamma = \mathbf{b}'_\gamma(i)}$  serão utilizadas as notações reduzidas  $\nabla_{\mathbf{b}'_\gamma} \varepsilon$  e  $\mathbf{H}$ .

#### 4.5.2 Caso Mel-Cepstral Generalizado ( $\alpha \neq 0, \gamma \neq 0$ )

Para o caso MGC ( $\alpha \neq 0, \gamma \neq 0$ ), na iteração  $i$  do método de Newton, de posse do vetor de coeficientes  $\mathbf{b}'_\gamma(i) = [b'_\gamma(i)(1) \ b'_\gamma(i)(2) \ \dots \ b'_\gamma(i)(M)]^T$ , o vetor  $\nabla_{\mathbf{b}'_\gamma} \varepsilon$  e a Matriz  $\mathbf{H}$  são dados respectivamente por

$$\nabla_{\mathbf{b}'_\gamma} \varepsilon = -2 \cdot \mathbf{r} \quad (4.61)$$

e

$$\mathbf{H} = 2 \cdot [\mathbf{P} + (1 + \gamma)\mathbf{Q}] \quad (4.62)$$

onde

$$\mathbf{r} = [\tilde{r}_\gamma(1) \ \tilde{r}_\gamma(2) \ \tilde{r}_\gamma(3) \ \dots \ \tilde{r}_\gamma(M)]^T \quad (4.63)$$

e

$$\mathbf{P} = \begin{bmatrix} \tilde{p}_\gamma(0) & \tilde{p}_\gamma(1) & \dots & \tilde{p}_\gamma(M-1) \\ \tilde{p}_\gamma(1) & \tilde{p}_\gamma(0) & \dots & \tilde{p}_\gamma(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_\gamma(M-1) & \tilde{p}_\gamma(M-2) & \dots & \tilde{p}_\gamma(0) \end{bmatrix} \quad (4.64)$$

$$\mathbf{Q} = \begin{bmatrix} \tilde{q}_\gamma(2) & \tilde{q}_\gamma(3) & \dots & \tilde{q}_\gamma(M+1) \\ \tilde{q}_\gamma(3) & \tilde{q}_\gamma(4) & \dots & \tilde{q}_\gamma(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}_\gamma(M+1) & \tilde{q}_\gamma(M+2) & \dots & \tilde{q}_\gamma(2M) \end{bmatrix} \quad (4.65)$$

onde  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  são expressos por



$$\tilde{r}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) \Phi_k^*(e^{j\omega}) d\omega \quad (4.66)$$

$$k = 1, 2, \dots, M$$

$$\tilde{p}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_1(e^{j\omega}) \cdot \Phi_{k+1}^*(e^{j\omega}) d\omega, \quad (4.67)$$

$$k = 0, 1, \dots, M - 1$$

e

$$\tilde{q}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_1^*(e^{j\omega}) \cdot \Phi_{k-1}^*(e^{j\omega}) d\omega, \quad (4.68)$$

$$k = 2, 3, \dots, 2M$$

onde

$$D(e^{j\omega}) = \left( 1 + \gamma \sum_{n=1}^M b_\gamma'^{(i)}(n) \Phi_n(e^{j\omega}) \right)^{1/\gamma} \quad (4.69)$$

que corresponde ao caso  $\gamma \neq 0$  ( $-1 \leq \gamma < 0$ ) de  $D(e^{j\omega})$  expresso por (4.36). A computação de  $D(e^{j\omega})$  e das integrais acima é abordada na seção 4.6.1. Uma segunda abordagem também é apresentada na seção 4.8.

Uma vez calculados o gradiente e a Hessiana de  $\varepsilon$ , resolve-se o sistema

$$\mathbf{H} \cdot \Delta \mathbf{b}_\gamma'^{(i)} = -\nabla_{\mathbf{b}'_\gamma} \varepsilon \quad (4.70)$$

e atualiza-se  $\mathbf{b}_\gamma'^{(i)}$  através de

$$\mathbf{b}_\gamma'^{(i+1)} = \mathbf{b}_\gamma'^{(i)} + \Delta \mathbf{b}_\gamma'^{(i)} \quad \text{expressão (4.60)}$$

### 4.5.3 Caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ )

Para o caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ ), na iteração  $i$  do método de Newton, de posse do vetor de coeficientes  $\mathbf{b}_0^{(i)} = [b_0^{(i)}(1) b_0^{(i)}(2) \dots b_0^{(i)}(M)]^T$ , o vetor  $\nabla_{\mathbf{b}'} \varepsilon$  e a Matriz  $\mathbf{H}$  são dados respectivamente por

$$\nabla_{\mathbf{b}'} \varepsilon = -2. \mathbf{r} \quad (4.71)$$

e

$$\mathbf{H} = 2. [\mathbf{P} + \mathbf{Q}] \quad (4.72)$$

onde

$$\mathbf{r} = [\tilde{r}_0(1) \quad \tilde{r}_0(2) \quad \tilde{r}_0(3) \quad \dots \quad \tilde{r}_0(M)]^T \quad (4.73)$$

e

$$\mathbf{P} = \begin{bmatrix} \tilde{p}_0(0) & \tilde{p}_0(1) & \dots & \tilde{p}_0(M-1) \\ \tilde{p}_0(1) & \tilde{p}_0(0) & \dots & \tilde{p}_0(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_0(M-1) & \tilde{p}_0(M-2) & \dots & \tilde{p}_0(0) \end{bmatrix} \quad (4.74)$$

$$\mathbf{Q} = \begin{bmatrix} \tilde{q}_0(2) & \tilde{q}_0(3) & \dots & \tilde{q}_0(M+1) \\ \tilde{q}_0(3) & \tilde{q}_0(4) & \dots & \tilde{q}_0(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}_0(M+1) & \tilde{q}_0(M+2) & \dots & \tilde{q}_0(2M) \end{bmatrix} \quad (4.75)$$

onde  $\tilde{r}_0(k)$ ,  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$  são expressos por

$$\tilde{r}_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_k^*(e^{j\omega}) d\omega, \quad k = 0, 1, \dots, 2M \quad (4.76)$$

$$\tilde{p}_0(k) = \begin{cases} (1 - \alpha^2)\tilde{r}_0(0) + 2\alpha\tilde{r}_0(1), & k = 0 \\ \tilde{r}_0(k) + \alpha\tilde{r}_0(k+1), & k = 1, 2, \dots, M-1 \end{cases} \quad (4.77)$$

e

$$\tilde{q}_0(k) = \tilde{r}_0(k) + \alpha\tilde{r}_0(k-1), \quad k = 2, 3, \dots, 2M \quad (4.78)$$

onde

$$D(e^{j\omega}) = \exp\left(\sum_{n=1}^M b_0^{(i)}(n)\Phi_n(e^{j\omega})\right) \quad (4.79)$$

que corresponde ao caso  $\gamma = 0$  de  $D(e^{j\omega})$  expresso por (4.36). A computação de  $D(e^{j\omega})$  e da integral para  $\tilde{r}_0(k)$  é abordada na seção 4.6.1. Uma segunda abordagem também é apresentada na seção 4.8.

Uma vez calculados o gradiente e a Hessiana de  $\varepsilon$ , resolve-se o sistema

$$\mathbf{H} \cdot \Delta \mathbf{b}_0^{(i)} = -\nabla_{\mathbf{b}'} \varepsilon \quad (4.80)$$

e atualiza-se  $\mathbf{b}_0^{(i)}$  através de

$$\mathbf{b}_0^{(i+1)} = \mathbf{b}_0^{(i)} + \Delta \mathbf{b}_0^{(i)} \quad (4.81)$$

#### 4.5.4 Caso Cepstral Generalizado ( $\alpha = 0, \gamma \neq 0$ )

Para o caso GC ( $\alpha = 0, \gamma \neq 0$ ), na iteração  $i$  do método de Newton, de posse do vetor de coeficientes  $\mathbf{c}_\gamma^{(i)} = [c_\gamma^{(i)}(1) \ c_\gamma^{(i)}(2) \ \dots \ c_\gamma^{(i)}(M)]^T$ , o vetor  $\nabla_{\mathbf{b}'} \varepsilon$  e a Matriz  $\mathbf{H}$  são dados respectivamente por

$$\nabla_{\mathbf{b}'} \varepsilon = -2 \cdot \mathbf{r} \quad (4.82)$$

e

$$\mathbf{H} = 2 \cdot [\mathbf{P} + (1 + \gamma)\mathbf{Q}] \quad (4.83)$$

onde

$$\mathbf{r} = [r_\gamma(1) \ r_\gamma(2) \ r_\gamma(3) \ \dots \ r_\gamma(M)]^T \quad (4.84)$$

e

$$\mathbf{P} = \begin{bmatrix} p_\gamma(0) & p_\gamma(1) & \dots & p_\gamma(M-1) \\ p_\gamma(1) & p_\gamma(0) & \dots & p_\gamma(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ p_\gamma(M-1) & p_\gamma(M-2) & \dots & p_\gamma(0) \end{bmatrix} \quad (4.85)$$

$$\mathbf{Q} = \begin{bmatrix} q_\gamma(2) & q_\gamma(3) & \cdots & q_\gamma(M+1) \\ q_\gamma(3) & q_\gamma(4) & \cdots & q_\gamma(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ q_\gamma(M+1) & q_\gamma(M+2) & \cdots & q_\gamma(2M) \end{bmatrix} \quad (4.86)$$

onde  $r_\gamma(k)$ ,  $p_\gamma(k)$  e  $q_\gamma(k)$  são expressos por

$$r_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) e^{j\omega k} d\omega \quad k = 1, 2, \dots, M \quad (4.87)$$

$$p_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 0, 1, \dots, M-1 \quad (4.88)$$

e

$$q_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) e^{j\omega k} d\omega \quad k = 2, 3, \dots, 2M \quad (4.89)$$

onde

$$D(e^{j\omega}) = \left( 1 + \gamma \sum_{n=1}^M c_\gamma^{(i)}(n) e^{-j\omega n} \right)^{1/\gamma} \quad (4.90)$$

que corresponde ao caso  $\gamma \neq 0$  ( $-1 \leq \gamma < 0$ ) de  $D(e^{j\omega})$  expresso por (4.6). A computação de  $D(e^{j\omega})$  e das integrais acima é abordada no final deste tópico.

Uma vez calculados o gradiente e a Hessiana de  $\varepsilon$ , resolve-se o sistema

$$\mathbf{H} \cdot \Delta \mathbf{c}_\gamma^{(i)} = -\nabla_{\mathbf{b}'} \varepsilon \quad (4.91)$$

e atualiza-se  $\mathbf{c}_\gamma^{(i)}$  através de

$$\mathbf{c}_\gamma^{(i+1)} = \mathbf{c}_\gamma^{(i)} + \Delta \mathbf{c}_\gamma^{(i)} \quad (4.92)$$

### Computação dos Coeficientes GC

Na iteração  $i$ , para o cálculo de  $D(e^{j\omega})$  computa-se a FFT da expressão no interior dos parênteses em (4.90). Com base nela são computados  $r_\gamma(k)$ ,  $p_\gamma(k)$  e  $q_\gamma(k)$  via FFT, e  $\mathbf{r}$ ,  $\mathbf{P}$  e  $\mathbf{Q}$  são determinados. Resolve-se o sistema (4.91) para  $\Delta \mathbf{c}_\gamma^{(i)}$  através de

$$[\mathbf{P} + (1 + \gamma)\mathbf{Q}]\Delta\mathbf{c}'_{\gamma}{}^{(i)} = \mathbf{r} \quad (4.93)$$

Como  $\mathbf{P}$  e  $\mathbf{Q}$  são respectivamente uma matriz de Toeplitz e uma matriz de Hankel, algoritmos eficientes, como os sugeridos em [26] e [27], podem ser utilizados para a computação de sua soma. Calcula-se  $\mathbf{c}'_{\gamma}{}^{(i+1)}$  através de (4.92) e inicia-se uma nova iteração (caso necessário).

Como critério de parada calcula-se a taxa de variação de  $\varepsilon$  a cada iteração. Para isso, antes que se complete a iteração  $i$ , computa-se  $r_{\gamma}(k)$  e

$$\varepsilon^{(i)} = r_{\gamma}(0) + \gamma \mathbf{r}^T \cdot \mathbf{c}'_{\gamma}{}^{(i)} \quad (4.94)$$

com  $\mathbf{r}$  dado por (4.84). A dedução de tal expressão é semelhante àquela feita para (E.34) no Apêndice E. De posse de  $\varepsilon^{(i-1)}$ , calculado na iteração anterior, determina-se a taxa de variação de  $\varepsilon^{(i)}$  para  $\varepsilon^{(i-1)}$ . A convergência pode ser considerada suficiente se a taxa for inferior a 0.1%. Caso contrário, completa-se a iteração para o cálculo de  $\mathbf{c}'_{\gamma}{}^{(i+1)}$ .

Um conjunto de coeficientes iniciais próximo à solução pode ser obtido. Para isso parte-se de  $M$  amostras do cepstrum complexo dadas no vetor

$$\mathbf{c}'_{0,M} = [c_0(1) \ c_0(2) \ c_0(3) \ \dots \ c_0(M)]^T \quad (4.95)$$

(onde não se inclui o coeficiente  $c_0(0)$ ). As amostras de índice  $k > 0$  do cepstrum complexo são dadas por

$$c_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(I_N(\omega)) e^{j\omega k} d\omega \quad k = 1, 2, \dots, M \quad (4.96)$$

que pode ser computado através de uma FFT de  $N_F$  pontos<sup>§</sup> ( $N_F$  é comprimento do periodograma  $I_N(\omega)$  amostrado). A partir de  $\mathbf{c}'_{0,M}$ , calcula-se o vetor de coeficientes iniciais

$$\mathbf{c}'_{\gamma}{}^{(0)} = [c'_{\gamma}{}^{(0)}(1) \ c'_{\gamma}{}^{(0)}(2) \ \dots \ c'_{\gamma}{}^{(0)}(M)]^T \quad (4.97)$$

obtido pela transformação

---

<sup>§</sup> Idealmente, tal cálculo exigiria infinitos pontos para que fossem evitadas distorções por *aliasing* no cepstrum. A distorção pelo cálculo com  $N_F$  pontos é, porém, aceitável para os coeficientes iniciais.

$$\mathbf{c}'_{\gamma}{}^{(0)} = T_{cg}(\mathbf{c}'_{0,M}, M, M, 0, \gamma) \quad (4.98)$$

à qual está associada a fórmula recursiva (2.78) e cujas propriedades são abordadas na seção 2.7.4.

Observa-se que para  $\gamma = -1$  (caso LPC) não é necessário o cálculo de coeficientes iniciais pois, uma vez que  $\nabla_{\mathbf{b}'}\varepsilon$  se torna linear, uma solução exata é obtida em uma única *iteração*.

#### 4.5.5 Caso Cepstral ( $\alpha = 0, \gamma = 0$ )

Para o caso Cepstral ( $\alpha = 0, \gamma \neq 0$ ), na iteração  $i$  do método de Newton, de posse do vetor de coeficientes  $\mathbf{c}'_0{}^{(i)} = [c_0^{(i)}(1) \ c_0^{(i)}(2) \ \dots \ c_0^{(i)}(M)]^T$ , o vetor  $\nabla_{\mathbf{b}'}\varepsilon$  e a Matriz  $\mathbf{H}$  são dados respectivamente por

$$\nabla_{\mathbf{b}'}\varepsilon = -2. \mathbf{r} \quad (4.99)$$

e

$$\mathbf{H} = 2. [\mathbf{P} + \mathbf{Q}] \quad (4.100)$$

onde

$$\mathbf{r} = [r_0(1) \ r_0(2) \ r_0(3) \ \dots \ r_0(M)]^T \quad (4.101)$$

e

$$\mathbf{P} = \begin{bmatrix} p_0(0) & p_0(1) & \dots & p_0(M-1) \\ p_0(1) & p_0(0) & \dots & p_0(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ p_0(M-1) & p_0(M-2) & \dots & p_0(0) \end{bmatrix} \quad (4.102)$$

$$\mathbf{Q} = \begin{bmatrix} q_0(2) & q_0(3) & \dots & q_0(M+1) \\ q_0(3) & \tilde{q}_0(4) & \dots & q_0(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ q_0(M+1) & q_0(M+2) & \dots & q_0(2M) \end{bmatrix} \quad (4.103)$$

onde  $r_0(k)$ ,  $p_0(k)$  e  $q_0(k)$  são expressos por

$$r_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 0, 1, \dots, 2M \quad (4.104)$$

$$p_0(k) = r_0(k), \quad k = 0, 1, \dots, M - 1 \quad (4.105)$$

e

$$q_0(k) = r_0(k), \quad k = 2, 3, \dots, 2M \quad (4.106)$$

onde

$$D(e^{j\omega}) = \exp\left(\sum_{n=1}^M c_0^{(i)}(n) e^{-j\omega n}\right) \quad (4.107)$$

que corresponde ao caso  $\gamma = 0$  de  $D(e^{j\omega})$  expresso por (4.6). A computação de  $D(e^{j\omega})$  e da integral para  $r_0(k)$  é abordada no final deste tópic.

Uma vez calculados o gradiente e a Hessiana de  $\varepsilon$ , resolve-se o sistema

$$\mathbf{H} \cdot \Delta \mathbf{c}_0'^{(i)} = -\nabla_{\mathbf{b}'} \varepsilon \quad (4.108)$$

e atualiza-se  $\mathbf{c}_0'^{(i)}$  através de

$$\mathbf{c}_0'^{(i+1)} = \mathbf{c}_0'^{(i)} + \Delta \mathbf{c}_0'^{(i)} \quad (4.109)$$

### Computação dos Coeficientes Cepstrais

Na iteração  $i$ , para o cálculo de  $D(e^{j\omega})$ , computa-se a FFT da expressão no interior dos parênteses em (4.107). Com base nela computa-se  $r_0(k)$  via FFT e  $p_0(k)$  e  $q_0(k)$  são determinados. Obtidos o vetor  $\mathbf{r}$  e as matrizes  $\mathbf{P}$  e  $\mathbf{Q}$ , resolve-se, para  $\Delta \mathbf{c}_0'^{(i)}$ , o sistema (4.108) através de

$$[\mathbf{P} + \mathbf{Q}] \Delta \mathbf{c}_0'^{(i)} = \mathbf{r} \quad (4.110)$$

Como  $\mathbf{P}$  e  $\mathbf{Q}$  são respectivamente uma matriz de Toeplitz e uma matriz de Hankel, algoritmos eficientes, como os sugeridos em [26] e [27], podem ser utilizados para a computação de sua soma. Calcula-se  $\mathbf{c}_0'^{(i+1)}$  através de (4.109) e inicia-se uma nova iteração.

Como critério de parada calcula-se a taxa de variação de  $\varepsilon$  a cada iteração. Para isso, antes que se complete a iteração  $i$ , computa-se  $r_0(k)$  e  $\varepsilon^{(i)}$  é dado por

$$\varepsilon^{(i)} = r_0(0) \quad (4.111)$$

De posse de  $\varepsilon^{(i-1)}$ , calculado na iteração anterior, determina-se a taxa de variação de  $\varepsilon^{(i)}$  para  $\varepsilon^{(i-1)}$ . A convergência pode ser considerada suficiente se a taxa for inferior a 0.1%. Caso contrário, completa-se a iteração para o cálculo de  $\mathbf{c}'_0^{(i+1)}$ .

Um conjunto de coeficientes iniciais

$$\mathbf{c}'_0^{(0)} = [c_0^{(0)}(1) \ c_0^{(0)}(2) \ \dots \ c_0^{(0)}(M)]^T \quad (4.112)$$

próximo à solução, dado pelo cepstrum complexo (para  $k > 0$ ), é calculado por (4.96), através de uma FFT de  $N_F$  pontos<sup>§</sup> ( $N_F$  é comprimento do periodograma  $I_N(\omega)$  amostrado).

## 4.6 Computação dos Coeficientes Mel-Cepstrais Generalizados ( $\alpha \neq 0$ ) por meio de FFTs e Fórmulas Recursivas

Na seção anterior, foi visto que para os casos em que  $\alpha = 0$  (GC e Cepstral), as integrais para o cálculo dos elementos do gradiente e da Hessiana de  $\varepsilon$  correspondem a transformadas inversas de Fourier, permitindo que seu cálculo seja feito via FFT. Para os casos em que  $\alpha \neq 0$  (MGC e Mel-Cepstral), tais integrais, por não se apresentarem como transformadas de Fourier, exigem outro método para a sua computação. Esta seção aborda os métodos apresentados em [2] e [7], respectivamente, para o cálculo dos coeficientes MGC e dos coeficientes Mel Cepstrais. Ambos se baseiam no uso de fórmulas recursivas para os cálculos dos elementos do gradiente e da Hessiana de  $\varepsilon$  e transformações de coeficientes exigidas. Uma abordagem alternativa, em que tais cálculos são efetuados via FFT, será apresentada na seção 4.8.

---

<sup>§</sup> Idealmente, tal cálculo exigiria infinitos pontos para que fossem evitadas distorções por *aliasing* no cepstrum. A distorção pelo cálculo com  $N_F$  pontos é, porém, aceitável para os coeficientes iniciais.



#### 4.6.1 Caso Mel-Cepstral Generalizado ( $\alpha \neq 0, \gamma \neq 0$ )

Os cálculos da  $i$ -ésima iteração do método de Newton-Raphson iniciam-se, dado o conjunto de coeficientes MGC transformados

$$\mathbf{b}'_{\gamma}{}^{(i)} = [ b'_{\gamma}{}^{(i)}(1) b'_{\gamma}{}^{(i)}(2) \dots b'_{\gamma}{}^{(i)}(M) ]^T \quad (4.113)$$

O primeiro passo para o cálculo dos elementos  $\tilde{r}_{\gamma}(k)$ ,  $\tilde{p}_{\gamma}(k)$  e  $\tilde{q}_{\gamma}(k)$ , constituintes do gradiente e da Hessiana de  $\varepsilon$ , consiste na aquisição do cepstrum generalizado

$$\mathbf{c}'_{\gamma, N'}{}^{(i)} = [ c'_{\gamma}{}^{(i)}(1) c'_{\gamma}{}^{(i)}(2) \dots c'_{\gamma}{}^{(i)}(N') ]^T \quad (4.114)$$

associado a  $\mathbf{b}'_{\gamma}{}^{(i)}$ . Para tal, utiliza-se a seguinte transformação

$$\left[ 0 \ \mathbf{c}'_{\gamma, N'}{}^{(i)T} \right]^T = T_1 \left( \left[ 0 \ \mathbf{b}'_{\gamma, M}{}^{(i)T} \right]^T, M, N', -\alpha \right) \quad (4.115)$$

que é computada através da fórmula recursiva (C.45) e abordada na seção C.3 do Apêndice C. Uma especificação completa de  $\mathbf{b}'_{\gamma, M}{}^{(i)}$ , exigiria, idealmente,  $N' = \infty$  componentes do cepstrum generalizado  $\mathbf{c}'_{\gamma, N'}{}^{(i)}$ . É suficiente, porém, que se adote um valor grande  $N' \gg 2M$ , sob o teto de  $N' = N_F - 1$  ( $N_F$  é o comprimento do periodograma amostrado). Aconselha-se a adoção de, no mínimo,  $N' \approx \frac{1+|\alpha|}{1-|\alpha|} (2M)$ <sup>§</sup>.

A partir de  $\mathbf{c}'_{\gamma, N'}{}^{(i)}$ , são calculados

$$\begin{aligned} \mathbf{r}_{\gamma} &= [ r_{\gamma}(0) r_{\gamma}(1) \dots r_{\gamma}(N') ]^T \\ \mathbf{p}_{\gamma} &= [ p_{\gamma}(0) p_{\gamma}(1) \dots p_{\gamma}(N') ]^T \\ \mathbf{q}_{\gamma} &= [ q_{\gamma}(0) q_{\gamma}(1) \dots q_{\gamma}(N') ]^T \end{aligned} \quad (4.116)$$

onde

$$r_{\gamma}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}^{1+\gamma}(e^{j\omega})|^2} \bar{D}^{\gamma}(e^{j\omega}) e^{j\omega k} d\omega, \quad k = 0, 1, \dots, N' \quad (4.117)$$

---

<sup>§</sup> Como é assumido um único valor de  $N'$  para todas as transformações em frequência aqui efetuadas, adota-se o mínimo de  $N' + 1 \approx \frac{1+|\alpha|}{1-|\alpha|} (2M + 1)$  amostras, com base na aplicação do critério de suavização da seção 3.4.3 à transformação (4.123), que gera o maior número de amostras dentre todas as transformações aqui envolvidas ( $2M + 1$  amostras) e conta, portanto, com o maior número  $N' + 1$  de amostras de entrada.

$$p_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}^{1+\gamma}(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 0, 1, \dots, N' \quad (4.118)$$

$$q_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}^{1+2\gamma}(e^{j\omega})|^2} \bar{D}^{2\gamma}(e^{j\omega}) e^{j\omega k} d\omega, \quad k = 0, 1, \dots, N' \quad (4.119)$$

com  $\bar{D}(e^{j\omega})$  dado por

$$\bar{D}(e^{j\omega}) = \left( 1 + \gamma \sum_{n=1}^{N'} c_\gamma^{(i)}(n) e^{-j\omega n} \right)^{1/\gamma} \quad (4.120)$$

A expressão no interior dos parênteses em (4.120) e as expressões (4.117) - (4.119) podem ser computadas via FFT.

O cálculo de  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  (expressões (4.66),(4.67) e (4.68)), é feito aplicando-se as transformações (E.16), (E.23) e (E.33) às quais estão associadas as fórmulas recursivas (E.15), (E.22) e (E.32), apontadas e tratadas no Apêndice E. Em especial para o cálculo de  $\tilde{p}_\gamma(k)$ , utiliza-se o vetor  $\mathbf{p}'_\gamma = [\frac{1}{2}p_\gamma(0) p_\gamma(1) \dots p_\gamma(N')]^T$  no lugar de  $\mathbf{p}_\gamma = [p_\gamma(0) p_\gamma(1) \dots p_\gamma(N')]^T$ . Assim, as transformações são dadas por

$$\tilde{\mathbf{r}}_{\gamma, M} = T_1(\mathbf{r}_{\gamma, N'}, N', M, \alpha) \quad (4.121)$$

$$\tilde{\mathbf{p}}'_{\gamma, M-1} = (1 - \alpha^2) \cdot T_\Lambda(\mathbf{p}'_{\gamma, N'}, N', M - 1, \alpha) \quad (4.122)$$

$$\tilde{\mathbf{q}}_{\gamma, 2M} = T_{1, \Phi_1}(\mathbf{q}_{\gamma, N'}, N', 2M, \alpha) \quad (4.123)$$

Em (4.122)  $\tilde{\mathbf{p}}'_{\gamma, M-1}$  é dado por  $\tilde{\mathbf{p}}'_{\gamma, M-1} = [\frac{1}{2}\tilde{p}_\gamma(0) \tilde{p}_\gamma(1) \dots \tilde{p}_\gamma(M-1)]^T$ . Considerado isso, as transformações (4.121) - (4.123) fornecem os vetores

$$\begin{aligned} \tilde{\mathbf{r}}_{\gamma, M} &= [\tilde{r}_\gamma(0) \tilde{r}_\gamma(1) \dots \tilde{r}_\gamma(M)]^T \\ \tilde{\mathbf{p}}_{\gamma, M-1} &= [\tilde{p}_\gamma(0) \tilde{p}_\gamma(1) \dots \tilde{p}_\gamma(M-1)]^T \\ \tilde{\mathbf{q}}_{\gamma, 2M} &= [\tilde{q}_\gamma(0) \tilde{q}_\gamma(1) \dots \tilde{q}_\gamma(2M)]^T \end{aligned} \quad (4.124)$$

onde  $\tilde{\mathbf{p}}_{\gamma, M-1}$  é obtido dobrando-se a zero-ésima componente de  $\tilde{\mathbf{p}}'_{\gamma, M-1}$ . De posse de tais vetores, obtém-se o vetor  $\mathbf{r}$  dado por (4.63) e as matrizes  $\mathbf{P}$  e  $\mathbf{Q}$ , dadas por (4.64) e (4.65) e resolve-se o sistema

$$[\mathbf{P} + (1 + \gamma)\mathbf{Q}]\Delta\mathbf{b}'_{\gamma}{}^{(i)} = \mathbf{r} \quad (4.125)$$

para

$$\Delta\mathbf{b}'_{\gamma}{}^{(i)} = [\Delta b'_{\gamma}{}^{(i)}(1) \Delta b'_{\gamma}{}^{(i)}(2) \dots \Delta b'_{\gamma}{}^{(i)}(M)]^T \quad (4.126)$$

Como  $\mathbf{P}$  e  $\mathbf{Q}$  são respectivamente uma matriz de Toeplitz e uma matriz de Hankel, algoritmos eficientes podem ser utilizados para computação de sua soma [26][27].

Atualizando  $\mathbf{b}'_{\gamma}{}^{(i)}$  através de

$$\mathbf{b}'_{\gamma}{}^{(i+1)} = \mathbf{b}'_{\gamma}{}^{(i)} + \Delta\mathbf{b}'_{\gamma}{}^{(i)} \quad (4.127)$$

obtém-se o novo conjunto de coeficientes MGC transformados

$$\mathbf{b}'_{\gamma}{}^{(i+1)} = [b'_{\gamma}{}^{(i+1)}(1) b'_{\gamma}{}^{(i+1)}(2) \dots b'_{\gamma}{}^{(i+1)}(M)]^T \quad (4.128)$$

Como critério de parada calcula-se a taxa de variação de  $\varepsilon$  por iteração. Para isso, antes que se complete a iteração  $i$ , computa-se  $\tilde{r}_{\gamma}(k)$  e

$$\varepsilon^{(i)} = \tilde{r}_{\gamma}(0) + \gamma \mathbf{r}^T \cdot \mathbf{b}'_{\gamma}{}^{(i)} \quad (4.129)$$

com  $\mathbf{r}$  dado por (4.63). Tal expressão é deduzida no Apêndice E. De posse de  $\varepsilon^{(i-1)}$ , calculado na iteração anterior, determina-se a taxa de variação de  $\varepsilon^{(i)}$  para  $\varepsilon^{(i-1)}$ . A convergência pode ser considerada suficiente se a taxa for inferior a 0.1%. Caso contrário, completa-se a iteração para o cálculo de  $\mathbf{b}'_{\gamma}{}^{(i+1)}$ .

### **Cálculo dos Coeficientes Iniciais $\mathbf{b}'_{\gamma}{}^{(0)}$**

Para o cálculo dos coeficientes iniciais, parte-se do cepstrum complexo com  $N' + 1$  amostras dado pelo vetor

$$\mathbf{c}_{0,N'} = [c_0(0) c_0(1) c_0(2) \dots c_0(N')]^T \quad (4.130)$$

onde, uma vez que o coeficiente  $c_0(0)$  não tem influência nos cálculos subsequentes pode-se assumir  $c_0(0) = 0$ . As amostras de índice  $k > 0$  do cepstrum complexo são dadas por

$$c_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(I_N(\omega)) e^{j\omega k} d\omega \quad k = 1, 2, \dots, N' \quad (4.131)$$

que pode ser computada através de uma FFT de  $N_F$  pontos<sup>§</sup> (comprimento do periodograma  $I_N(\omega)$  amostrado).

O mel-cepstrum de ordem  $M$ , dado por

$$\tilde{\mathbf{c}}_{0,M} = [ \tilde{c}_0(0) \ \tilde{c}_0(1) \ \tilde{c}_0(2) \ \dots \ \tilde{c}_0(M) ]^T \quad (4.132)$$

é calculado a partir do cepstrum pela transformação em frequência

$$\tilde{\mathbf{c}}_{0,M} = T_\Lambda(\mathbf{c}_{0,N'}, N', M, \alpha) \quad (4.133)$$

à qual está associada a fórmula recursiva (3.29) e cujas propriedades são abordadas na seção 3.4.3. O coeficiente  $\tilde{c}_0(0)$  não é usado nos cálculos subseqüentes. Assim no lugar do vetor  $\tilde{\mathbf{c}}_{0,M}$ , assume-se o vetor

$$\tilde{\mathbf{c}}'_{0,M} = [ \tilde{c}'_0(1) \ \tilde{c}'_0(2) \ \tilde{c}'_0(3) \ \dots \ \tilde{c}'_0(M) ]^T \quad (4.134)$$

a partir do qual, é calculado o mel-cepstrum generalizado

$$\tilde{\mathbf{c}}'_{\gamma,M} = [ \tilde{c}'_\gamma(1) \ \tilde{c}'_\gamma(2) \ \tilde{c}'_\gamma(3) \ \dots \ \tilde{c}'_\gamma(M) ]^T \quad (4.135)$$

obtido pela transformação

$$\tilde{\mathbf{c}}'_{\gamma,M} = T_{cg}(\tilde{\mathbf{c}}'_{0,M}, M, M, 0, \gamma) \quad (4.136)$$

à qual está associada a fórmula recursiva (2.78) e cujas propriedades são abordadas na seção 2.7.4. A aplicação da transformação (4.29) ao vetor  $[ 0 \ \tilde{\mathbf{c}}'^T_{\gamma,M} ]^T$  expressa por

$$\mathbf{b}_{\gamma,M} = \mathbf{A}_{M+1} \cdot [ 0 \ \tilde{\mathbf{c}}'^T_{\gamma,M} ]^T \quad (4.137)$$

fornece os coeficientes MGC transformados

$$\mathbf{b}_{\gamma,M} = [ b_\gamma(0) \ b_\gamma(1) \ b_\gamma(2) \ \dots \ b_\gamma(M) ]^T \quad (4.138)$$

que, normalizados, fornecem o vetor de coeficientes iniciais

$$\mathbf{b}'^{(0)}_\gamma = [ b'^{(0)}_\gamma(1) \ b'^{(0)}_\gamma(2) \ \dots \ b'^{(0)}_\gamma(M) ]^T \quad (4.139)$$

onde

$$b'^{(0)}_\gamma(n) = \frac{b_\gamma(n)}{1 + \gamma b_\gamma(0)} \quad (4.140)$$

---

<sup>§</sup> Idealmente, tal cálculo exigiria infinitos pontos para que fossem evitadas distorções por *aliasing* no cepstrum. A distorção pelo cálculo com  $N_F$  pontos é, porém, aceitável para os coeficientes iniciais.

Observa-se que para  $\gamma = -1$  (caso Mel-LPC) não é necessário o cálculo de coeficientes iniciais pois, uma vez que  $\nabla_{\mathbf{b}'} \varepsilon$  se torna linear, uma solução exata é obtida em uma única *iteração*.

#### 4.6.2 Caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ )

Os cálculos da  $i$ -ésima iteração do método de Newton-Raphson iniciam-se, dado o conjunto de coeficientes mel-cepstrais transformados

$$\mathbf{b}'_0^{(i)} = [ b_0^{(i)}(1) b_0^{(i)}(2) \dots b_0^{(i)}(M) ]^T \quad (4.141)$$

O primeiro passo para o cálculo dos elementos  $\tilde{r}_0(k)$ ,  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$ , constituintes do gradiente e da Hessiana de  $\varepsilon$ , consiste na aquisição do cepstrum

$$\mathbf{c}'_{0,N'}^{(i)} = [ c_0^{(i)}(1) c_0^{(i)}(2) \dots c_0^{(i)}(N') ]^T \quad (4.142)$$

associado a  $\mathbf{b}'_0^{(i)}$ . Para tal, utiliza-se a seguinte transformação

$$\left[ 0 \ \mathbf{c}'_{0,N'}^{(i)T} \right]^T = T_1 \left( \left[ 0 \ \mathbf{b}'_0^{(i)T} \right]^T, M, N', -\alpha \right) \quad (4.143)$$

que é computada através da fórmula recursiva (C.45) e abordada na seção C.3 do Apêndice C. Uma especificação completa de  $\mathbf{b}'_0^{(i)}$  exigiria, idealmente,  $N' = \infty$  componentes do cepstrum  $\mathbf{c}'_{0,N'}^{(i)}$ . É suficiente, porém, que se adote um valor grande  $N' \gg 2M$ , sob o teto de  $N' = N_F - 1$  ( $N_F$  é o comprimento do periodograma amostrado). Aconselha-se a adoção de, no mínimo,  $N' \approx \frac{1+|\alpha|}{1-|\alpha|} (2M)^\S$ .

A partir de  $\mathbf{c}'_{0,N'}^{(i)}$  calcula-se

$$\mathbf{r}_{0,N'} = [ r_0(0) r_0(1) \dots r_0(N') ]^T \quad (4.144)$$

onde

---

<sup>§</sup> Como é assumido um único valor de  $N'$  para as transformações em frequência aqui efetuadas, adota-se o mínimo de  $N' + 1 \approx \frac{1+|\alpha|}{1-|\alpha|} (2M + 1)$  amostras, com base na aplicação do critério de suavização da seção 3.4.3 à transformação (4.147), que gera o maior número de amostras dentre as transformações aqui envolvidas ( $2M + 1$  amostras) e conta, portanto, com o maior número  $N' + 1$  de amostras de entrada.

$$r_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 0, 1, \dots, N' \quad (4.145)$$

com  $\bar{D}(e^{j\omega})$  dado por

$$\bar{D}(e^{j\omega}) = \exp\left(\sum_{n=1}^{N'} c_0(n) e^{-j\omega n}\right) \quad (4.146)$$

O argumento de (4.146) e a expressão (4.145) podem ser computados via FFT.

O cálculo de  $\tilde{r}_0(k)$  (expressão (4.76)) é feito aplicando-se a transformação

$$\tilde{\mathbf{r}}_{0,2M} = T_1(\mathbf{r}_{0,N'}, N', 2M, \alpha) \quad (4.147)$$

à qual está associada a fórmula recursiva (E.45), apontada e deduzida no Apêndice E. O vetor  $\tilde{\mathbf{r}}_{0,2M}$  é dado por

$$\tilde{\mathbf{r}}_{0,2M} = [\tilde{r}_0(0) \tilde{r}_0(1) \tilde{r}_0(2) \dots \tilde{r}_0(2M)]^T \quad (4.148)$$

De posse da seqüência das  $2M + 1$  amostras de  $\tilde{r}_0(k)$ , calcula-se  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$  por

$$\tilde{p}_0(k) = \begin{cases} (1 - \alpha^2)\tilde{r}_0(0) + 2\alpha\tilde{r}_0(1), & k = 0 \\ \tilde{r}_0(k) + \alpha\tilde{r}_0(k+1), & k = 1, 2, \dots, M-1 \end{cases} \quad (\text{expressão (4.77)})$$

e

$$\tilde{q}_0(k) = \tilde{r}_0(k) + \alpha\tilde{r}_0(k-1), \quad k = 2, 3, \dots, 2M \quad (\text{expressão (4.78)})$$

expressões deduzidas na seção D.2.2 do Apêndice D. As seqüências  $\tilde{r}_0(k)$ ,  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$  calculadas permitem a obtenção do vetor  $\mathbf{r}$  dado por (4.73) e das matrizes  $\mathbf{P}$  e  $\mathbf{Q}$ , dadas por (4.74) e (4.75). Assim, resolve-se o sistema

$$[\mathbf{P} + \mathbf{Q}]\Delta\mathbf{b}'_0^{(i)} = \mathbf{r} \quad (4.149)$$

para

$$\Delta\mathbf{b}'_0^{(i)} = [\Delta b_0^{(i)}(1) \Delta b_0^{(i)}(2) \dots \Delta b_0^{(i)}(M)]^T \quad (4.150)$$

Como  $\mathbf{P}$  e  $\mathbf{Q}$  são respectivamente uma matriz de Toeplitz e uma matriz de Hankel, algoritmos eficientes podem ser utilizados para computação de sua soma [26][27].

Atualizando  $\mathbf{b}'_0^{(i)}$  através de

$$\mathbf{b}'_0^{(i+1)} = \mathbf{b}'_0^{(i)} + \Delta \mathbf{b}'_0^{(i)} \quad (4.151)$$

obtém-se o conjunto de coeficientes

$$\mathbf{b}'_0^{(i+1)} = [ b'_0{}^{(i+1)}(1) \ b'_0{}^{(i+1)}(2) \ \dots \ b'_0{}^{(i+1)}(M) ]^T \quad (4.152)$$

Como critério de parada calcula-se a taxa de variação de  $\varepsilon$  por iteração. Para isso, antes que se complete a iteração  $i$ , computa-se  $\tilde{r}_0(k)$  e  $\varepsilon^{(i)}$  é dado por

$$\varepsilon^{(i)} = \tilde{r}_0(0) \quad (4.153)$$

(o que pode ser verificado assumindo-se  $k = 0$  em (4.76)). De posse de  $\varepsilon^{(i-1)}$ , calculado na iteração anterior, determina-se a taxa de variação de  $\varepsilon^{(i)}$  para  $\varepsilon^{(i-1)}$ . A convergência pode ser considerada suficiente se a taxa for inferior a 0.1%. Caso contrário, completa-se a iteração para o cálculo de  $\mathbf{b}'_0^{(i+1)}$ .

### **Cálculo dos Coeficientes Iniciais $\mathbf{b}'_0^{(0)}$**

Para o cálculo dos coeficientes iniciais, parte-se do cepstrum complexo  $c_0^{(0)}(n)$  com  $N'$  amostras dado pelo vetor

$$\mathbf{c}'_{0,N'}^{(0)} = [ c_0^{(0)}(1) \ c_0^{(0)}(2) \ \dots \ c_0^{(0)}(N') ]^T \quad (4.154)$$

que não inclui a amostra  $c_0(0)$  uma vez que ela não tem influência nos cálculos subsequentes. As amostras de índice  $k > 0$  do cepstrum complexo são dadas por (4.131), e podem ser computadas por uma FFT de  $N_F$  pontos<sup>§</sup> (comprimento do periodograma  $I_N(\omega)$  amostrado). A partir de  $\mathbf{c}'_{0,N'}^{(0)}$  a transformação

$$\left[ 0 \ \mathbf{b}'_0{}^{(0)T} \right]^T = \frac{1}{1 - \alpha^2} T_1 \left( \left[ 0 \ \mathbf{c}'_{0,N'}{}^{(0)T} \right]^T, N', M, \alpha \right) \quad (4.155)$$

fornece os coeficientes iniciais

$$\mathbf{b}'_0{}^{(0)} = [ b_0^{(0)}(1) \ b_0^{(0)}(2) \ \dots \ b_0^{(0)}(M) ]^T \quad (4.156)$$

A transformação (4.155) é computada através da fórmula recursiva (C.43) e abordada na seção C.3 do Apêndice C

---

<sup>§</sup> Idealmente, tal cálculo exigiria infinitos pontos para que fossem evitadas distorções por *aliasing* no cepstrum. A distorção pelo cálculo com  $N_F$  pontos é, porém, aceitável para os coeficientes iniciais.

## 4.7 Características do Modelo e Emprego dos Parâmetros MGC em Síntese de Fala Via HMM

Esta seção aborda as características do modelo MGC, dando exemplos de como o modelo se ajusta a diferentes características do espectro de fala, sob diferentes atribuições de valores aos parâmetros  $\alpha$  e  $\gamma$ . São tratados a ressíntese do sinal de fala e o uso dos parâmetros MGC em síntese de fala via HMM.

### 4.7.1 Perfis do Modelo e o Espectro de Fala

Nesta seção são avaliadas as diferenças entre modelos MGC associados a parâmetros  $(\alpha, \gamma)$  distintos, e como diferentes modelos descrevem diferentes características do espectro de fala. O espectro de fala é primordialmente caracterizado pela presença de formantes e anti-formantes, que representam, respectivamente, as ressonâncias (picos) e anti-ressonâncias (vales) características da configuração instantânea do trato vocal e, portanto, do fonema em emissão. Vogais orais são caracterizadas pela presença exclusiva de formantes em seu espectro. Já vogais e soantes nasais<sup>§</sup> (e laterais), caracterizam-se pela presença de anti-formantes, além de formantes, em seu espectro. A seguir faz-se uma análise de características intrínsecas ao modelo MGC, posteriormente tratadas no contexto do espectro de fala.

Como ponto de partida para uma avaliação dos perfis do modelo MGC, considere dois modelos distintos: um associado aos parâmetros  $(\alpha_1, \gamma_1)$  e outro, aos parâmetros  $(\alpha_2, \gamma_2)$ , ambos aplicados a um mesmo espectro. Quanto mais elevada a ordem de ambos os modelos, melhor o seu ajuste ao espectro e mais proximidade entre si eles adquirem. Por outro lado, quanto menor as suas ordens, mais se evidenciam os perfis característicos de cada modelo. Considerando isso, a avaliação será feita levando a representação trigonométrica de modelos MGCs distintos ao limite de somente uma parcial (senóide). Uma vez evidenciadas suas características, explora-se o ajuste do modelo ao espectro de fala.

Na seção 3.5.2 (ou 2.3.2 para o caso  $\alpha = 0$ ), foi apontado que o mel-cepstrum generalizado integra os coeficientes de uma série trigonométrica que representa a

---

<sup>§</sup> Alguns exemplos de fonemas com base na vogal "a" são dados a seguir. Exemplo de vogal nasal: "ã". Exemplos de soantes nasais: "m" e "n" representando, respectivamente, os *murmúrios* contidos nas pronúncias das sílabas *ma* e *na*. Ex. de soante lateral: "l" representando o *murmúrio* na pronúncia de *la*.



magnitude mel-logarítmica generalizada do espectro. Os coeficientes MGC integram um polinômio trigonométrico de ordem  $M$  que representa a magnitude mel-logarítmica generalizada do modelo, dada por

$$\tilde{f}_\gamma(\tilde{\omega}) = \ln_\gamma \left( |\tilde{H}(e^{j\tilde{\omega}})|^2 \right) = \tilde{h}_\gamma(0) + 2 \sum_{n=0}^M \tilde{h}_\gamma(n) \cos(\tilde{\omega}n) \quad (4.157)$$

onde  $\tilde{H}(e^{j\tilde{\omega}})$  corresponde ao modelo MGC na escala distorcida de frequências, ou seja  $\tilde{H}(e^{j\tilde{\omega}}) = H(e^{j\omega(\tilde{\omega})})$ , e os coeficientes  $\tilde{h}_\gamma(n)$  estão relacionados aos coeficientes MGC  $\tilde{c}_\gamma(n)$  por

$$\tilde{h}_\gamma(n) = \begin{cases} 2 \cdot \tilde{c}_\gamma(0) + \gamma \cdot \sum_{k=0}^M \tilde{c}_\gamma^2(k), & n = 0 \\ \tilde{c}_\gamma(n) + \gamma \cdot \sum_{k=0}^{M-n} \tilde{c}_\gamma(k+n) \tilde{c}_\gamma(k), & 1 \leq n \leq M \end{cases} \quad (4.158)$$

Para que se obtenha a magnitude em dB do modelo, três operações são aplicadas à expressão (4.157): a substituição  $\tilde{\omega} = \tilde{\omega}(\omega)$ , que conduz a expressão à escala linear de frequências; a aplicação da função  $\exp_\gamma(\cdot)$ , que conduz a expressão à escala linear de amplitudes; e a aplicação da função  $10 \log_{10}(\cdot)$ , que leva a expressão à escala logarítmica em dB. Assim, a magnitude espectral do modelo em dB é dada por

$$10 \log_{10} \left( |H(e^{j\omega})|^2 \right) = 10 \log_{10} \left[ \exp_\gamma \left( \tilde{f}_\gamma(\tilde{\omega}(\omega)) \right) \right] \quad (4.159)$$

Destaca-se que, quando  $\gamma = 0$ , a exponencial generalizada em (4.159) é invertida pela função logarítmica, de tal maneira que

$$10 \log_{10} \left( |H(e^{j\omega})|^2 \right) = 10 \log_{10}(e) \cdot \tilde{f}_0(\tilde{\omega}(\omega)) \quad (4.160)$$

onde

$$\tilde{f}_0(\tilde{\omega}) = 2 \sum_{k=0}^M \tilde{c}_0(n) \cos(\tilde{\omega}k) \quad (4.161)$$

já que em (4.158) os coeficientes assumem os valores  $\tilde{h}_0(0) = 0$  e  $\tilde{h}_\gamma(n) = \tilde{c}_0(n)$  ( $1 \leq n \leq M$ ). Considere as curvas da função  $10 \log_{10} \left( \exp_\gamma(x) \right)$  com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ , dadas na Figura 4.2. Para  $\gamma = 0$ , a curva é linear e, conforme  $\gamma$  é reduzido ( $\gamma = -1/2$  e  $-1$ ), as curvas ganham caráter não-linear, de maneira que as

variações em  $x$  resultam em variações progressivamente mais lentas pela função no intervalo  $-1 < x < 0$  e mais rápidas no intervalo  $0 < x < 1$ .

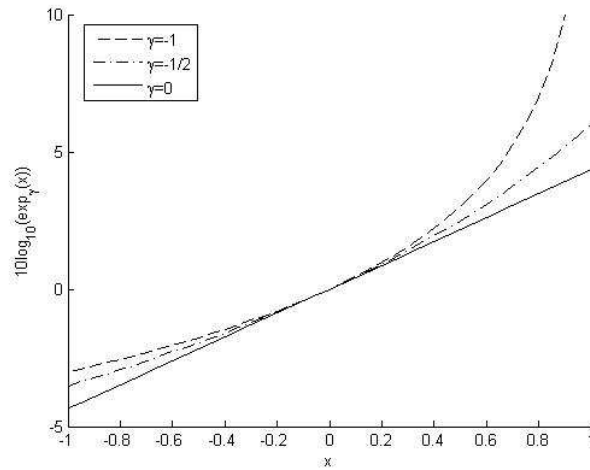


Figura 4.2: Curvas para  $\gamma = -1$ ,  $\gamma = -1/2$  e  $\gamma = 0$  da função  $10 \log_{10}(\exp_{\gamma}(x))$ . Para  $\gamma = 0$  a função se torna linear na forma  $10 \log_{10}(e) \cdot x$ .

Tomando os polinômios  $\tilde{f}_{\gamma}(\tilde{\omega}(\omega))$  associados aos três valores de  $\gamma$  assumidos acima e considerando em cada um deles somente os coeficientes  $\tilde{h}_{\gamma}(0)$  e  $\tilde{h}_{\gamma}(5)$  não-nulos (o que seleciona a parcial  $n = 5$  em (4.157)), ao substituí-los em  $x$  na função da Figura 4.2 são obtidas<sup>§</sup> as curvas da Figura 4.3, dadas para  $\alpha = 0$  e  $\alpha = 0.42$ .

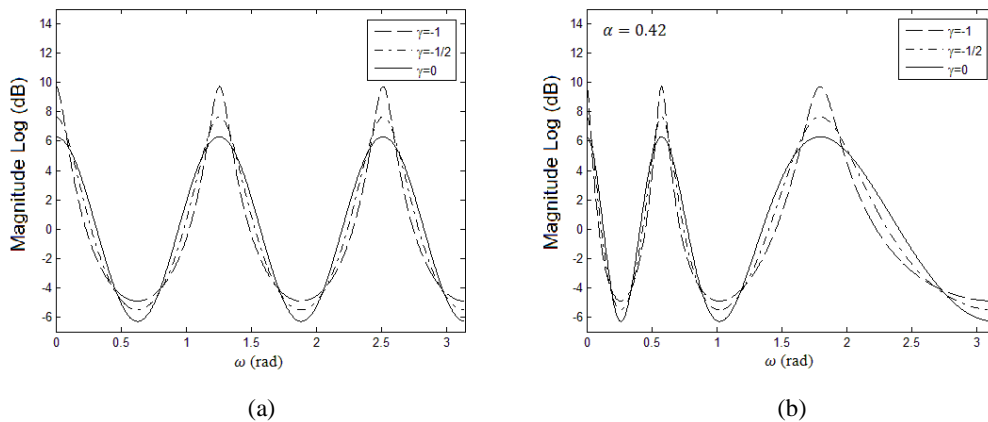


Figura 4.3: Curvas para  $\gamma = -1$ ,  $\gamma = -1/2$  e  $\gamma = 0$  da função  $10 \log_{10}[\exp_{\gamma}(\tilde{f}_{\gamma}(\tilde{\omega}(\omega)))]$  utilizando o harmônico  $k = 5$  de  $\tilde{f}_{\gamma}(\tilde{\omega})$ . (a) com  $\alpha = 0$  ( $\tilde{\omega}(\omega) = \omega$ ). (b) com  $\alpha = 0.42$ .

<sup>§</sup> com devidas atribuições de valores aos pares  $(\tilde{h}_{\gamma}(0), \tilde{h}_{\gamma}(5))$

A Figura 4.3 (a) representa curvas obtidas com parâmetro  $\alpha = 0$ . Nesse caso, as funções de base do modelo se manifestam em variações equivalentes ao longo do eixo de frequências o que caracteriza um ajuste descritivo uniforme ao longo do eixo de frequências, o que é esperado da análise GC. A Figura 4.3 (b) representa curvas obtidas com  $\alpha = 0.42$ , se manifestando em variações rápidas na região de baixas frequências, e gradualmente mais lentas conforme o aumento da frequência. Isso caracteriza um ajuste descritivo desigual ao longo do eixo das frequências, o que se espera da análise MGC.

Na Figura 4.3, em ambas as curvas de (a) e (b), a curva para  $\gamma = -1$  (caso LPC/Mel-LPC), tem picos estreitos e vales largos. Com o aumento progressivo de  $\gamma$ , segue-se a tendência ao alargamento dos picos e ao estreitamento dos vales, o que pode ser observado na curva para  $\gamma = -1/2$  e que culmina na curva para  $\gamma = 0$  (caso Cepstral/Mel-Cepstral). Nesse caso, picos e vales têm larguras equivalentes, o que é esperado, uma vez que a curva é uma senóide proveniente da expressão (4.160) somente com a parcial  $k = 5$ . A característica de estreitamento dos picos e suavização dos vales quando  $\gamma$  se aproxima de  $-1$  faz com que a análise com  $\gamma = -1$  (LPC/Mel-LPC) tenha as melhores condições para a modelagem de formantes, mas seja a mais problemática para a modelagem de vales acentuados no espectro (que freqüentemente associam-se a anti-formantes no espectro de fala). Por outro lado, quando  $\gamma$  se aproxima de  $0$  (caso cepstral), a tendência de abertura dos picos e estreitamento dos vales beneficia a descrição de vales acentuados, ao custo de alargar a banda dos formantes, podendo ocorrer ambigüidade na representação daqueles próximos entre si.

Em síntese: Fixada a ordem do modelo, a escolha do parâmetro  $\gamma$  pesa uma relação compromissada entre as descrições de formantes e anti-formantes. Assim, considerando  $\gamma = -1$  (valor que melhor descreve formantes e portanto emissões orais), ao se variar  $\gamma$  no sentido positivo, fonemas nasalizados, caracterizados pela presença de anti-formantes, são gradualmente melhor descritos. O custo do aumento de  $\gamma$  para a descrição de emissões orais é uma abertura na banda dos formantes.

Em [28] e [29] são feitas avaliações da descrição do espectro de vogais pelo modelo GC. A Figura 4.4 (a) - (d) mostra a desambiguação de formantes próximos entre si pelo modelo GC ( $\alpha = 0$ ) aplicado ao espectro da vogal oral “ó”, conforme a redução do valor de  $\gamma$ . Tal característica também é observada na Figura 4.4 (e), pela adoção de  $\alpha = 0.42$ . A Figura 4.5 (a) - (c) mostra o ajuste aos anti-formantes da soante nasal “m”

pelo modelo GC ( $\alpha = 0$ ) conforme o aumento do valor de  $\gamma$ . Tal ajuste também é observado na Figura 4.5 (d), pela adoção de  $\alpha = 0.42$ .

Destaca-se que a discussão aqui efetuada considera a adoção de uma ordem  $M$  baixa o suficiente para que envoltória do modelo não registre as interferências devido ao período fundamental no espectro de fala. Tal questão será abordada no contexto do emprego dos parâmetros MGC em síntese paramétrica estatística (seção 4.7.3).

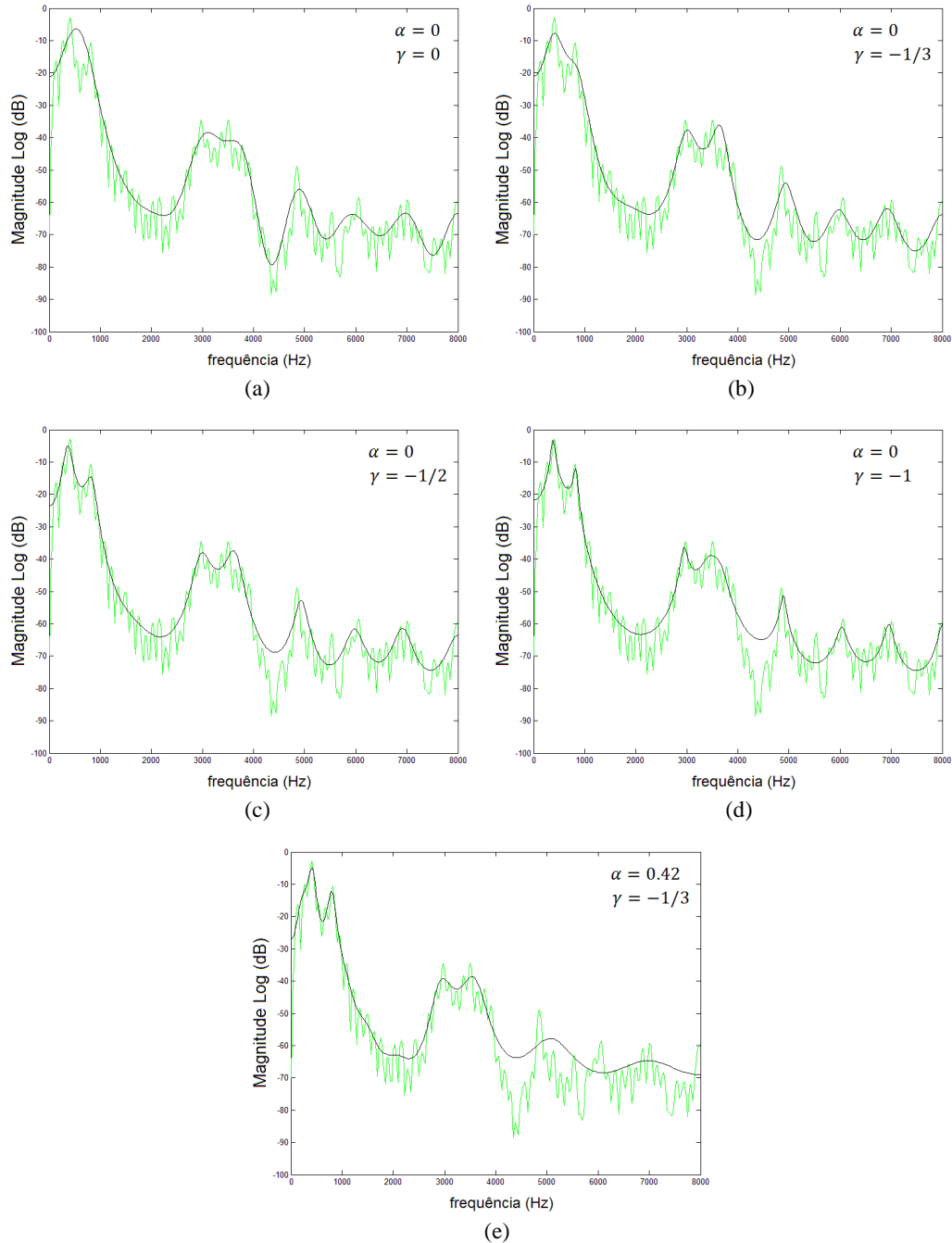


Figura 4.4: Envoltórias espectrais obtidas por análises MGC de ordem 20 aplicadas à emissão da vogal “ó” amostrada a 16kHz (espectro em verde). (a) Com  $(\alpha, \gamma) = (0, 0)$ , o modelo não faz distinção entre o par de formantes  $f_1 \sim 500$ Hz e  $f_2 \sim 900$ Hz, o mesmo vale para o par  $f_3 \sim 3000$ Hz e  $f_4 \sim 3500$ Hz. (b) Com  $(\alpha, \gamma) = (0, -1/3)$ , o modelo passa a distinguir  $f_3$  e  $f_4$ . (c) Com  $(\alpha, \gamma) = (0, -1/2)$ , adicionalmente, o modelo passa a distinguir entre  $f_1$  e  $f_2$ . (d) Com  $(\alpha, \gamma) = (0, -1)$ , bandas dos formantes seguem se estreitando e o vale em  $\sim 4300$ Hz tem sua descrição mais pobre. (e) Com  $(\alpha, \gamma) = (0.42, -1/3)$ , o modelo descreve frequências baixas com maior resolução e passa a distinguir os formantes  $f_1$  e  $f_2$ , antes ambíguos no caso  $(\alpha, \gamma) = (0, -1/3)$ .

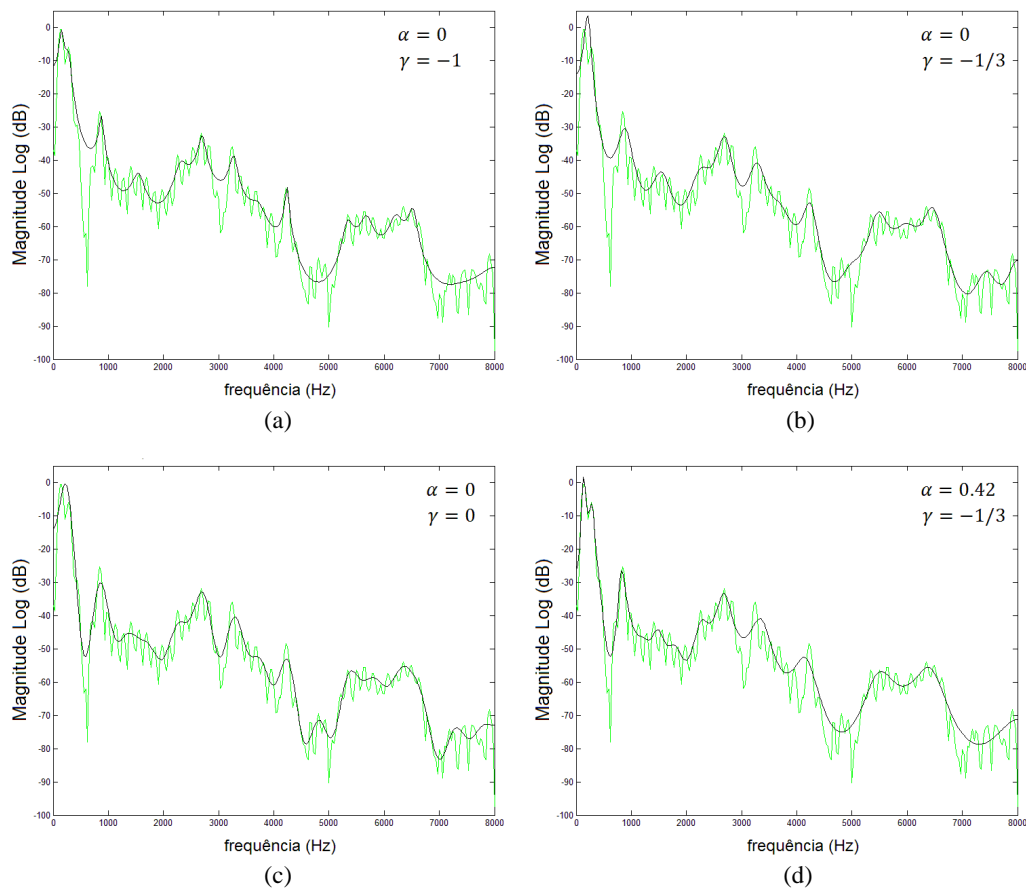


Figura 4.5: Envoltórias espectrais obtidas por análises MGC de ordem 32 aplicadas à emissão da soante nasal “m” amostrada a 16kHz (espectro em verde). (a) Com  $(\alpha, \gamma) = (0, -1)$  o modelo não descreve os vales associados aos anti-formantes  $af_1 \sim 600\text{Hz}$  e  $af_2 \sim 3000\text{Hz}$  (b)  $(\alpha, \gamma) = (0, -1/3)$ : o aumento em  $\gamma$  ainda não contempla tais vales. (c) Com  $(\alpha, \gamma) = (0, 0)$ , os vales em  $af_1$  e  $af_2$  são contemplados pelo modelo. (d) Com  $(\alpha, \gamma) = (0.42, -1/3)$ , o vale em  $af_1$ , não contemplado no caso  $(\alpha, \gamma) = (0, -1/3)$ , passa a ser bem descrito em função da resolução superior na região de baixas frequências.

## 4.7.2 Ressíntese do Sinal de Fala

A ressíntese do sinal de fala, efetuada com base no modelo fonte-filtro, requer a representação do modelo MGC por um filtro racional, o que é obtido pelos filtros MLSA [30] (*Mel Log Spectrum Approximation Filter*) e MGLSA [31] (*Mel Generalized Log Spectrum Approximation Filter*). Destaca-se que a adoção dos valores para o parâmetro  $\gamma$ , na forma  $\gamma = -1/n$ , com  $n$  natural não-nulo, leva o modelo MGC à representação

$$D(z) = \left( \frac{1}{1 + \gamma \sum_{n=1}^M b'_n(n) \Phi_n(z)} \right)^n \quad (4.162)$$

que corresponde a  $n$  filtros *all-pole* (distorcidos em frequência) em cascata e é, portanto, diretamente implementável. Uma avaliação subjetiva básica da síntese é feita em [2].

Destaca-se que, na ressíntese são utilizados, a priori, coeficientes diretamente extraídos pela análise MGC, já que, sob quantização, interpolação, ou quando gerados por estatísticas, os coeficientes MGC podem levar o modelo à instabilidade (nos casos  $-1 \leq \gamma < 0$ ), da qual se previne somente no caso  $\gamma = 0$  (coeficientes Mel-Cepstrais ou Cepstrais). Isso pode ser solucionado pela transformação dos coeficientes em parâmetros do tipo LSP [10] (MGC-LSP), para os quais um critério simples garante estabilidade ao modelo. Após manipulados, os parâmetros são novamente convertidos nos coeficientes MGC para a ressíntese.

## 4.7.3 Emprego dos Parâmetros MGC em Síntese de Fala via HMM

A técnica de seleção de unidades evoluiu de forma a se tornar a abordagem dominante em conversão texto-fala. A qualidade da síntese por tal técnica é derivada diretamente da qualidade das gravações de sua base, já que ela retém unidades de fala basicamente inalteradas. Já a síntese paramétrica estatística [32], cuja instância principal é a síntese de fala via HMM, que tem crescido em popularidade nos últimos anos, pode ser descrita, de maneira simples, como uma geração de médias de conjuntos de segmentos de fala com sonoridades similares. Apesar do contraste direto com a abordagem dominante, a síntese via HMM tem trazido alguns benefícios em qualidade, o que tem sido constatado em testes subjetivos. A maioria de tais benefícios recai sobre

a flexibilidade devido ao processo de modelagem estatística. Um deles corresponde à possibilidade de modificação das características da voz e do estilo de fala do locutor, por técnicas como adaptação e interpolação de locutor. Outro corresponde à cobertura do espaço acústico. Na síntese por seleção de unidades, a seleção é feita a partir de um conjunto finito de segmentos, o que leva a problemas na junção dos mesmos. A síntese paramétrica estatística, por sua vez, usa a estatística para *gerar* a fala, de maneira que, uma vez que o contexto de um segmento afeta a geração dos parâmetros de fala através do vínculo estabelecido por parâmetros dinâmicos, um intervalo bem mais amplo de *unidades* fica efetivamente disponível. Outra vantagem da síntese estatística é o espaço requerido pelo sistema, que é reduzido, uma vez que são armazenadas estatísticas de modelos acústicos, ao invés de segmentos de fala completos.

Na síntese paramétrica estatística são modelados, basicamente, o espectro, a frequência fundamental, e a duração da fala. A síntese dos parâmetros gerados é efetuada pelo modelo fonte filtro. Algumas das desvantagens da síntese via HMM são: a baixa qualidade da síntese pela técnica de *vocoding*, cujo artefato típico é o *zumbido*; a baixa precisão do modelo estatístico, que faz uso de simplificações que levam a afetar a qualidade da fala sintética; a sobre-suavização das trajetórias dos parâmetros gerados, que levam a síntese a soar *abafada*. Esta última decorre do processo de médias estatísticas que, apesar de tornar os modelos robustos e levar os parâmetros dinâmicos a gerarem trajetórias suaves (sem problemas de *junção* como é o caso na síntese por seleção de unidades), remove características detalhadas dos parâmetros de fala. Feitas tais considerações a respeito da síntese paramétrica estatística, aborda-se a seguir o uso dos parâmetros MGC no sistema de síntese e alternativas para amenizar alguns dos problemas mencionados.

A flexibilidade oferecida pelo modelo MGC mediante o ajuste dos parâmetros  $\alpha$  e  $\gamma$ , e a garantia de estabilidade e bom desempenho sob manipulações [10] (pelo uso dos parâmetros MGC-LSP), têm feito dos MGC, parâmetros de uso recorrente na modelagem estatística de espectro na síntese via HMM [14].

Um dos problemas do uso dos parâmetros do modelo MGC decorre da seguinte questão: a envoltória do modelo MGC pode vir a registrar as interferências devido ao período fundamental no espectro de fala. Tais interferências se manifestam em oscilações periódicas na curva do espectro. Conforme a frequência fundamental aumenta, essas oscilações ficam mais lentas, tornando-se passíveis de serem registradas pela curva do modelo MGC. Assim, o uso dos parâmetros MGC obtidos diretamente do



espectro de fala, nos HMM, leva os modelos estatísticos de espectro a incorporarem as variabilidades decorrentes de interferências devido à frequência fundamental. Isso diverge do propósito de se modelar separadamente espectro e frequência fundamental no sistema de síntese, e conduz a modelos espectrais mais pobres. Uma solução para o problema é o uso de parâmetros MGC obtidos da aplicação da análise MGC a envoltórias pré-suavizadas. O espectro STRAIGHT, que se mostra uma representação de qualidade para a envoltória espectral de fala, tem sido uma alternativa recorrente para tal propósito, e possibilita ordens mais altas para os parâmetros MGC a serem modelados nos HMM. O STRAIGHT [33] é um sistema de *vocoding* de alta qualidade que extrai os valores de frequência fundamental  $F_0$  e os utiliza em uma análise espectral adaptativa a  $F_0$  combinada a um método de reconstrução de superfície na região tempo-frequência para remover as interferências da periodicidade do sinal de fala. A Figura 4.6 compara as curvas obtidas pela análise Mel-Cepstral convencional ( $\gamma = 0$ ) de ordens 24, 39, e a análise Mel-Cepstral de ordem 39 aplicada ao espectro STRAIGHT. Nota-se que ambas as análises Mel-Cepstrais convencionais de ordem 24 e 39 registram as interferências de  $F_0$  no espectro; o que não ocorre na análise Mel-Cepstral aplicada ao espectro STRAIGHT [34] (Figura 4.6).

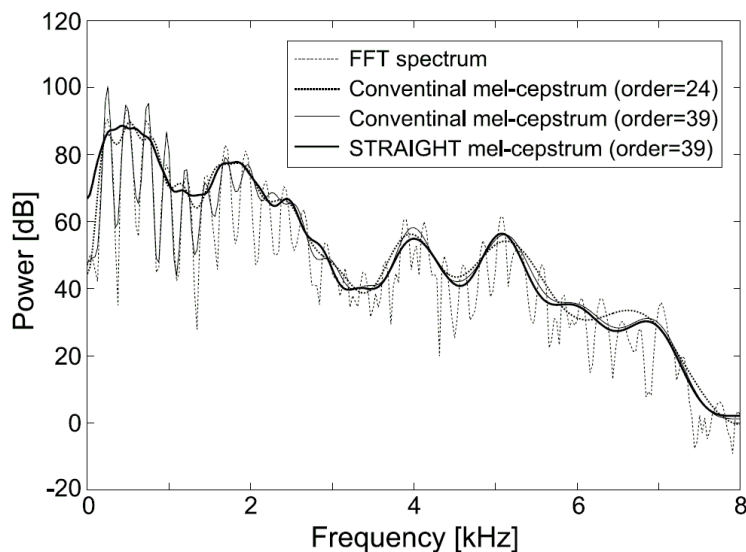


Figura 4.6: envoltórias obtidas por análise mel-cepstral ( $\alpha = 0$ ) aplicada diretamente a um espectro de fala (FFT) (ordem 24 e 39) e ao espectro pré-suavizado STRAIGHT (ordem 39) [34].

A seguir são comentados alguns dos métodos que levam a amenizar os problemas da síntese paramétrica estatística, já citados.

O problema da presença do *zumbido* na fala sintética, artefato típico de *Vocoders* que fazem uso de fontes simples do tipo *trem de impulsos/ruído*, tem sido amenizado pelo adoção de modelos de excitação/resíduo mista, cujos parâmetros são integrados no treinamento dos HMM. Alguns exemplos são os modelos de excitação mista do STRAIGHT [35] e aqueles de [36] e [37] [38].

Para o problema da sobre-suavização das trajetórias dos parâmetros gerados, que leva a síntese a soar *abafada*, melhoras significativas tem sido obtidas pelo uso de um algoritmo de geração de parâmetros considerando Variância Global (GV – *Global Variance*) [39]. Tal algoritmo se propõe a recuperar o intervalo dinâmico das trajetórias dos parâmetros gerados para aproximá-las das trajetórias naturais. Destaca-se que o desempenho da Variância Global é altamente dependente dos parâmetros de espectro. Bons desempenhos têm sido obtidos com o uso dos coeficientes Mel-Cepstrais ( $\gamma = 0$ ) extraídos, como parâmetros de espectro, a partir da envoltória do STRAIGHT. Isso se relaciona ao fato de sua representação não registrar interferências de F0, o que ocorre na representação pelos Mel-Cepstrais convencionais [34] (ver Figura 4.6). Apesar de os parâmetros MGC-LSP ( $\gamma \neq 0$ ) terem bom desempenho sob quantização e interpolação [10], o que os tornaria, a priori, uma boa escolha para parâmetros de espectro nos HMM<sup>§</sup>, testes subjetivos mostram que, sob a adoção de GV e transformações como a MLLT (*Maximum Likelihood Linear Transform*) seu desempenho é inferior àquele dos parâmetros Mel-Cepstrais [14] (todos extraídos da análise aplicada à envoltória pré-suavizada do STRAIGHT). Alguns dos problemas da adoção de parâmetros do tipo LSP em síntese paramétrica estatística são abordados em [32].

---

<sup>§</sup> A modelagem estatística tem relação próxima à quantização e a síntese tem relação próxima à interpolação [14].

## 4.8 Método para a Computação dos Coeficientes Mel-Cepstrais Generalizados ( $\alpha \neq 0$ ) por meio de FFTs

Na seção 4.6 foram abordados os métodos de [2] e [7] para a determinação dos coeficientes MGC e Mel-Cepstrais que, além de FFTs, fazem uso de fórmulas recursivas para os cálculos do gradiente e da Hessiana de  $\varepsilon$ . Na presente seção são apresentados métodos para tais cálculos, que contam somente com o uso de FFTs. Para sua aplicação é necessário que se disponha do periodograma original (dado por (4.51)) adotado na escala distorcida de frequências  $\tilde{\omega}$ , expresso por

$$\tilde{I}_N(\tilde{\omega}) = I_N(\omega(\tilde{\omega})) \quad (4.163)$$

A computação do periodograma  $I_N(\omega)$  via FFT, fornece no mínimo  $N$  amostras  $I_N(\omega_m)$  igualmente espaçadas em  $\omega$ , ou seja, calculadas em  $\omega = \omega_m = \frac{2\pi}{N}m$ , com  $m = 0, 1, \dots, N$ . Os cálculos com base na formulação aqui apresentada devem dispor de  $N_\alpha$  amostras  $\tilde{I}_N(\tilde{\omega}_k)$  igualmente espaçadas em  $\tilde{\omega}$ , ou seja, calculadas em  $\tilde{\omega} = \tilde{\omega}_k = \frac{2\pi}{N_\alpha}k$ , com  $k = 0, 1, \dots, N_\alpha$ . Os métodos disponíveis a priori para tal cálculo (apontados na seção 4.8.3) são custosos, o que é agravado pela exigência de valores altos para  $N_\alpha$  para que sejam evitadas degradações na representação de  $\tilde{I}_N(\tilde{\omega})$ . Como alternativa, uma interpolação linear do periodograma amostrado  $I_N(\omega_m)$  permite um cálculo mais econômico de  $\tilde{I}_N(\tilde{\omega}_k)$ . Através de tal técnica, a economia nos cálculos é reforçada pelo fato de que a boa qualidade da descrição de  $\tilde{I}_N(\tilde{\omega})$ , através de  $\tilde{I}_N(\tilde{\omega}_k)$ , se mostra tolerante a valores mais baixos de  $N_\alpha$ . O método para o cálculo de  $\tilde{I}_N(\tilde{\omega})$  será abordado no tópico 4.8.3. A formulação e a computação para a solução dos modelos MGC e Mel-Cepstral são abordadas seguir, e contam com as demonstrações do Apêndice F.

O modelo MGC considerado no plano  $\tilde{z} = |\tilde{z}|e^{j\tilde{\omega}}$  é expresso por

$$\tilde{D}(\tilde{z}) = \exp_\gamma \left( \sum_{n=1}^M b_\gamma'(n) \tilde{\Phi}_n(\tilde{z}) \right) \quad (4.164)$$

onde

$$\tilde{\Phi}_n(\tilde{z}) = \begin{cases} 1, & n = 0 \\ (\tilde{z}^{-1} + \alpha)\tilde{z}^{-(n-1)}, & n > 0 \end{cases} \quad (4.165)$$

A solução para os coeficientes do modelo é obtida pela minimização de

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \left( \frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} \right) d\tilde{\omega} \quad (4.166)$$

obtido pela mudança de variável  $\omega = \omega(\tilde{\omega})$  aplicada a (4.52) assumindo  $\tilde{I}_N(\tilde{\omega}) = I_N(\omega(\tilde{\omega}))$  e  $\tilde{D}(e^{j\tilde{\omega}}) = D(e^{j\omega(\tilde{\omega})})$ . As seções 4.8.1 e 4.8.2 trazem as expressões para os elementos  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  (caso MGC) e  $\tilde{r}_0(k)$  (caso Mel-Cepstral) de  $\nabla_{\mathbf{b}'} \varepsilon$  e  $\mathbf{H}$  obtidas com base em (4.166), e abordam sua computação.

#### 4.8.1 Caso Mel-Cepstral Generalizado ( $\alpha \neq 0, \gamma \neq 0$ )

Para a análise MGC os elementos  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  do gradiente e da Hessiana de  $\varepsilon$  (na forma (4.166)) são dados por

$$\tilde{r}_\gamma(k) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \frac{(1-\alpha^2)}{(e^{j\tilde{\omega}} + \alpha)(e^{-j\tilde{\omega}} + \alpha)} d\tilde{\omega}, & k = 0 \\ (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \frac{1}{(1 + \alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}k} d\tilde{\omega}, & k = 1, 2, \dots, M \end{cases} \quad (4.167)$$

$$\tilde{p}_\gamma(k) = (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} e^{j\tilde{\omega}k} d\tilde{\omega} \quad k = 0, 1, \dots, M-1 \quad (4.168)$$

e

$$\tilde{q}_\gamma(k) = (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+2\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^{2\gamma}(e^{j\tilde{\omega}}) \left( \frac{e^{-j\tilde{\omega}} + \alpha}{1 + \alpha e^{-j\tilde{\omega}}} \right) e^{-j\tilde{\omega}(k-1)} d\tilde{\omega} \quad (4.169)$$

$$k = 2, 3, \dots, 2M$$

onde em (4.167), inclui-se  $\tilde{r}_\gamma(0)$  necessário para o cálculo de  $\varepsilon$  a cada iteração do método de Newton-Raphson. As integrais (4.168) e (4.169) podem ser calculadas via FFT. A integral para  $k \geq 1$  em (4.167), se calculada via FFT, não contemplaria  $\tilde{r}_\gamma(0)$ . Uma maneira de se determinar  $\tilde{r}_\gamma(k)$  para  $k \geq 0$  é calculando-se a seguinte integral auxiliar

$$a(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) e^{j\tilde{\omega}k} d\tilde{\omega} \quad k = -\frac{N}{2}, \dots, -1, 0, 1, \frac{N}{2} - 1 \quad (4.170)$$

de onde determina-se recursivamente

$$a_1(n) = \begin{cases} a\left(\frac{N}{2} - 1\right), & n = \frac{N}{2} - 1 \\ a(n) - \alpha a_1(n+1), & 0 \leq n < \frac{N}{2} - 1 \end{cases} \quad (4.171)$$

e

$$a_2(n) = \begin{cases} a\left(-\frac{N}{2}\right), & n = -\frac{N}{2} \\ a(n) - \alpha a_2(n-1), & -\frac{N}{2} < n \leq -1 \end{cases} \quad (4.172)$$

Com base em  $a_1(n)$  e  $a_2(n)$ ,  $\tilde{r}_\gamma(k)$  é dado por

$$\tilde{r}_\gamma(k) = \begin{cases} a_1(0) - \alpha a_2(-1), & k = 0 \\ (1 - \alpha^2) a_1(k), & k = 1, 2, \dots, M \end{cases} \quad (4.173)$$

Na iteração  $i$  do método de Newton-Raphson, de posse do conjunto de coeficientes MGC  $\mathbf{b}'_\gamma(i) = [b'_\gamma(i)(1) b'_\gamma(i)(2) \dots b'_\gamma(i)(M)]^T$ , as integrais para  $a(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  ((4.170), (4.168) e (4.169)), são calculadas com base no modelo  $\tilde{D}(e^{j\tilde{\omega}})$  que, na forma

$$\tilde{D}(e^{j\tilde{\omega}}) = \left( 1 + \gamma(1 + \alpha e^{j\tilde{\omega}}) \sum_{n=1}^M b'_\gamma(i)(n) e^{-j\tilde{\omega}n} \right)^{1/\gamma} \quad (4.174)$$

pode ter o fator do somatório calculado via FFT.

Uma vez computados  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$ , resolve-se o sistema

$$[\mathbf{P} + (1 + \gamma)\mathbf{Q}]\Delta\mathbf{b}'_\gamma(i) = \mathbf{r} \quad \text{expressão (4.125)}$$

para  $\Delta\mathbf{b}'_\gamma(i)$ , com  $\mathbf{r}$ ,  $\mathbf{P}$  e  $\mathbf{Q}$  dados respectivamente por (4.63), (4.64) e (4.65), e o novo conjunto de coeficientes MGC  $\mathbf{b}'_\gamma(i+1) = [b'_\gamma(i+1)(1) b'_\gamma(i+1)(2) \dots b'_\gamma(i+1)(M)]^T$  é obtido por  $\mathbf{b}'_\gamma(i+1) = \mathbf{b}'_\gamma(i) + \Delta\mathbf{b}'_\gamma(i)$ .

e os termos associados aos elementos da Hessiana  $\mathbf{H}$ , por (F.13) e (F.14).

Uma vez calculado  $\tilde{r}_\gamma(k)$  ( $k = 0, 1, \dots, M$ ) o cálculo de  $\varepsilon$  pode ser feito através de

$$\varepsilon^{(i)} = \tilde{r}_\gamma(0) + \gamma \mathbf{r}^T \cdot \mathbf{b}_\gamma'^{(i)} \quad (4.175)$$

onde  $\mathbf{r} = [\tilde{r}_\gamma(1) \ \tilde{r}_\gamma(2) \ \dots \ \tilde{r}_\gamma(M)]^T$ . Tal cálculo, efetuado anteriormente aos cálculos de  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$ , permite determinar a taxa de variação de  $\varepsilon$  por iteração. A convergência pode ser considerada suficiente quando a taxa for inferior a 0.1%.

### **Cálculo dos Coeficientes Iniciais $\mathbf{b}_\gamma'^{(0)}$**

O método para o cálculo dos coeficientes iniciais se assemelha àquele descrito na seção 4.6.1, com a diferença de que com  $\tilde{I}_N(\tilde{\omega})$  em mãos, o mel-cepstrum com  $M$  amostras dado pelo vetor

$$\tilde{\mathbf{c}}'_{0,M} = [ \tilde{c}'_0(1) \ \tilde{c}'_0(2) \ \tilde{c}'_0(3) \ \dots \ \tilde{c}'_0(M) ]^T \quad (4.176)$$

(onde não se inclui o coeficiente  $\tilde{c}'_0(0)$ ), pode ser calculado, para  $k > 0$ , por

$$\tilde{c}'_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(\tilde{I}_N(\tilde{\omega})) e^{j\tilde{\omega}k} d\tilde{\omega} \quad k = 1, 2, \dots, M \quad (4.177)$$

através de uma FFT com número de pontos igual ao do comprimento do periodograma  $\tilde{I}_N(\tilde{\omega})$  amostrado. De posse do mel-cepstrum  $\tilde{\mathbf{c}}'_{0,M}$ , os coeficientes iniciais  $\mathbf{b}_\gamma'^{(0)} = [ b_\gamma'^{(0)}(1) \ b_\gamma'^{(0)}(2) \ \dots \ b_\gamma'^{(0)}(M) ]^T$  são obtidos por passos equivalentes àqueles descritos através das expressões (4.134)-(4.140) da seção 4.6.1.

Observa-se que para  $\gamma = -1$  (caso Mel-LPC), não é necessário o cálculo de coeficientes iniciais, uma vez que  $\nabla_{\mathbf{b}'} \varepsilon$  se torna linear e uma solução exata é obtida com somente uma iteração.

### **4.8.2 Caso Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ )**

Para a análise Mel-Cepstral a expressão para  $\tilde{r}_0(k)$ , obtida com base em  $\varepsilon$  (na forma (4.166)) é dada por

$$\tilde{r}_0(l) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \frac{(1-\alpha^2)}{(e^{j\tilde{\omega}} + \alpha)(e^{-j\tilde{\omega}} + \alpha)} d\tilde{\omega} & l = 0 \\ (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \frac{1}{(1 + \alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}l} d\tilde{\omega}, & l = 1, 2, \dots, 2M \end{cases} \quad (4.178)$$

De maneira que para seu cálculo, computa-se, via FFT

$$a_0(l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \frac{1}{(1 + \alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}l} d\tilde{\omega} \quad l = 0, 1, \dots, 2M \quad (4.179)$$

e  $\tilde{r}_0(l)$  é obtido por

$$\tilde{r}_0(k) = \begin{cases} a_0(0) - \alpha a_0(1), & k = 0 \\ (1 - \alpha^2) a_0(k), & k = 1, 2, \dots, 2M \end{cases} \quad (4.180)$$

Na iteração  $i$  do método de Newton-Raphson, de posse do conjunto de coeficientes Mel-Cepstrais  $\mathbf{b}_0^{(i)} = [b_0^{(i)}(1) b_0^{(i)}(2) \dots b_0^{(i)}(M)]^T$ , a integral (4.179) para  $a_0(l)$  é calculada com base no modelo  $\tilde{D}(e^{j\tilde{\omega}})$  que, na forma

$$\tilde{D}(e^{j\tilde{\omega}}) = \exp \left( (1 + \alpha e^{j\tilde{\omega}}) \sum_{n=1}^M b_0^{(i)}(n) e^{-j\tilde{\omega}n} \right) \quad (4.181)$$

pode ter o fator do somatório calculado via FFT. Os termos associados aos elementos da Hessiana  $\mathbf{H}$  são calculados, partir de  $\tilde{r}_0(k)$  por

$$\tilde{p}_0(k) = \begin{cases} (1 - \alpha^2)\tilde{r}_0(0) + 2\alpha\tilde{r}_0(1), & k = 0 \\ \tilde{r}_0(k) + \alpha\tilde{r}_0(k+1), & k = 1, 2, \dots, M-1 \end{cases} \quad \text{expressão (4.77)}$$

e

$$\tilde{q}_0(k) = \tilde{r}_0(k) + \alpha\tilde{r}_0(k-1), \quad k = 2, 3, \dots, 2M \quad \text{expressão (4.78)}$$

Uma vez computados  $\tilde{r}_0(k)$ ,  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$ , resolve-se o sistema

$$[\mathbf{P} + \mathbf{Q}]\Delta\mathbf{b}_0^{(i)} = \mathbf{r} \quad \text{expressão (4.149)}$$

para  $\Delta\mathbf{b}_0^{(i)}$ , com  $\mathbf{r}$ ,  $\mathbf{P}$  e  $\mathbf{Q}$  dados respectivamente por (4.73), (4.74) e (4.75), e um novo conjunto de coeficientes  $\mathbf{b}_0^{(i+1)} = [b_0^{(i+1)}(1) b_0^{(i+1)}(2) \dots b_0^{(i+1)}(M)]^T$  é obtido por  $\mathbf{b}_0^{(i+1)} = \mathbf{b}_0^{(i)} + \Delta\mathbf{b}_0^{(i)}$ .

Uma vez calculado  $\tilde{r}_0(k)$  ( $k = 0, 1, \dots, M$ ),  $\varepsilon$  é dado por

$$\varepsilon^{(i)} = \tilde{r}_0(0) \quad (4.182)$$

Seu valor permite determinar a taxa de variação de  $\varepsilon$  por iteração. A convergência pode ser considerada suficiente quando a taxa for inferior a 0.1%.

### Cálculo dos Coeficientes Iniciais $\mathbf{b}'^{(0)}$

Para a obtenção dos coeficientes iniciais, o mel-cepstrum com  $M$  amostras  $\tilde{\mathbf{c}}'_{0,M} = [ \tilde{c}_0(1) \tilde{c}_0(2) \dots \tilde{c}_0(M) ]^T$  (onde não se inclui o coeficiente  $\tilde{c}_0(0)$ ) é dado pela expressão (4.177) e calculado via FFT com número de pontos igual ao do comprimento do periodograma  $\tilde{I}_N(\tilde{\omega})$  amostrado. Os coeficientes iniciais  $\mathbf{b}'^{(0)} = [ b_0^{(0)}(1) b_0^{(0)}(2) \dots b_0^{(0)}(M) ]^T$  são obtidos pela transformação

$$\mathbf{b}'^{(0)} = \mathbf{A}_M \cdot \tilde{\mathbf{c}}'_{0,M} \quad (4.183)$$

(a matriz  $\mathbf{A}_{M+1}$  é dada por (4.31)).

### 4.8.3 Método para a Computação do Periodograma $\tilde{I}_N(\tilde{\omega})$

Para o cálculo discreto do periodograma distorcido em frequência  $\tilde{I}_N(\tilde{\omega})$ , dois métodos são imediatos. Considere

$$I_N(\omega) = |S(e^{j\omega})|^2 \quad (4.184)$$

onde

$$S(e^{j\omega}) = \sum_{n=0}^{N-1} s(n)e^{-j\omega n} \quad (4.185)$$

com

$$s(n) = \frac{w(n)x(n)}{\sqrt{\sum_{n=0}^{N-1} w^2(n)}} \quad (4.186)$$

O periodograma  $\tilde{I}_N(\tilde{\omega})$ , em termos de  $S(e^{j\omega})$ , é dado por

$$\tilde{I}_N(\tilde{\omega}) = |\tilde{S}(\tilde{\omega})|^2 = |S(e^{j\omega(\tilde{\omega})})|^2 \quad (4.187)$$



Um método para o cálculo de  $\tilde{I}_N(\tilde{\omega})$  amostrado seria através da transformação de  $s(n)$  em uma seqüência distorcida em freqüência  $\tilde{s}(n)$  através de

$$\tilde{\mathbf{s}}_{N_\alpha} = T_\Lambda(\mathbf{s}_N, N, N_\alpha, \alpha)^\S \quad (4.188)$$

onde  $\mathbf{s}_N = [s(0) \ s(1) \ s(2) \ \dots \ s(N)]^T$  e  $\tilde{\mathbf{s}}_{N_\alpha} = [\tilde{s}(0) \ \tilde{s}(1) \ \tilde{s}(2) \ \dots \ \tilde{s}(N_\alpha)]^T$ , de maneira que  $\tilde{S}(\tilde{\omega})$  seria obtido pela DFT

$$\tilde{S}(\tilde{\omega}_l) = \sum_{n=0}^{N_\alpha-1} \tilde{s}(n) e^{-j\tilde{\omega}_l n} \quad \tilde{\omega}_l = \frac{2\pi}{N_\alpha} l \quad (4.189)$$

$$l = 0, 1, \dots, N_\alpha - 1$$

O segundo método seria o cálculo da transformada de Fourier de  $s(n)$  nos pontos  $\omega = \omega(\tilde{\omega}_l)$  ( $\omega(\tilde{\omega})$  é definida pela expressão (3.8)), dada por

$$\tilde{S}(\tilde{\omega}_l) = \sum_{n=0}^{N-1} s(n) e^{-j\omega(\tilde{\omega}_l)n} \quad \tilde{\omega}_l = \frac{2\pi}{N_\alpha} l \quad (4.190)$$

$$l = 0, 1, \dots, N_\alpha - 1$$

Ambos os métodos, com base no critério da seção 3.4.3 para representação espectral completa, exigiriam o cálculo de no mínimo  $N_\alpha \approx \frac{1+|\alpha|}{1-|\alpha|} N$  amostras  $\tilde{S}(\tilde{\omega}_l)$  para que distorções fossem minimizadas em  $\tilde{I}_N(\tilde{\omega})$ . No cálculo por (4.189), a redução do número de pontos da DFT em relação ao exigido para  $N_\alpha$  corresponderia a uma truncagem da seqüência  $\tilde{s}(n)$ , o que levaria a distorções devido à resolução reduzida a partir de um ponto no eixo de freqüências (ver seção 3.4.4). Já no cálculo por (4.190), a redução do número de amostras  $N_\alpha$  levaria a distorções por *aliasing* temporal (ver seção 3.4.5). Em experimentos, tal conduta não se mostrou uma boa estratégia para uma diminuição no custo nos cálculos, já que as distorções são consideráveis. Além disso, tais métodos são intrinsecamente custosos; um por envolver uma filtragem em cascata de  $N_\alpha$  etapas (transformação (4.188)) e o outro, por representar o cálculo de  $N_\alpha$  amostras desigualmente espaçadas em  $\omega$  (expressão (4.190)), o que exige que ele seja efetuado ponto a ponto, e não via FFT.

Uma solução que se mostra menos custosa e minimiza distorções em  $\tilde{I}_N(\tilde{\omega})$  é aqui proposta. A partir do sinal janelado  $w(n)x(n)$  de comprimento  $N$ , calcula-se o

---

<sup>§</sup> A transformação em freqüência é abordada na seção 3.4.3.

periodograma regular por uma DFT de  $N_1$  pontos ( $N_1 \geq N$ ). Deseja-se obter  $N_2$  amostras de  $\tilde{I}_N(\tilde{\omega}) = I_N(\omega(\tilde{\omega}))$ , ou seja, obter as amostras  $\tilde{I}_N(\tilde{\omega}_k)$ , nas frequências  $\omega(\tilde{\omega}_k)$  onde  $\tilde{\omega}_k = \frac{2\pi}{N_2}k$ , com  $k = 0, 1, \dots, N_2$ . Tais amostras, das quais não se dispõe a priori, são obtidas por uma interpolação linear aplicada às  $N_1$  amostras do periodograma regular, dadas por  $I_N(\omega_m)$  onde  $\omega_m = \frac{2\pi}{N_1}m$  com  $m = 0, 1, \dots, N_1$ . Considera-se aqui, que  $I_N(\omega_m)$  é calculado por uma FFT de  $N_1 \geq N$  pontos (a partir das  $N$  amostras de  $w(n)x(n)$ ). Pelo critério da seção 3.4.5, seriam necessários no mínimo  $N_2 = N_\alpha \approx \frac{1+|\alpha|}{1-|\alpha|}N$  amostras para que fossem evitadas distorções por *aliasing* temporal no periodograma. Um critério para o valor de  $N_1$ , que minimizaria distorções decorrentes do cálculo de amostras por interpolação, seria obtido adotando-se o mínimo de  $N_1 \approx \frac{1+|\alpha|}{1-|\alpha|}N_2 = \left(\frac{1+|\alpha|}{1-|\alpha|}\right)^2 N$ . Dessa maneira, se fosse adotado  $\alpha = 0.5$ , por exemplo, teríamos as razões  $\frac{N_2}{N} = 3$  e  $\frac{N_1}{N_2} = 3$ . Com  $N = 512$ , seria necessário o cálculo de uma DFT de  $N_1 \approx 4608$  pontos (amostras  $I_N(\omega_m)$ ) para o periodograma  $I_N(\omega)$ , e o cálculo de  $N_2 \approx 1536$  amostras  $\tilde{I}_N(\tilde{\omega}_k)$  de  $\tilde{I}_N(\tilde{\omega})$  por interpolação linear das amostras  $I_N(\omega_m)$  de  $I_N(\omega)$ .

Como se pode constatar, as razões  $\frac{N_2}{N} \approx \frac{1+|\alpha|}{1-|\alpha|}$  e  $\frac{N_1}{N_2} \approx \frac{1+|\alpha|}{1-|\alpha|}$  (exigidas para um cálculo *ideal* de  $\tilde{I}_N(\tilde{\omega})$ ) levam a valores altos de  $N_1$  e  $N_2$  que, se adotados, tornariam o método custoso. Avaliações preliminares têm mostrado que a adoção de valores mais baixos para as razões  $\frac{N_2}{N}$  e  $\frac{N_1}{N_2}$  leva a degradações pouco significativas na curva do modelo. Como exemplo, o periodograma  $I_N(\omega_m)$  de um bloco de 24ms de um sinal de fala amostrado a 16kHz, tem comprimento  $N = 384$  amostras. Para  $\alpha = 0.42$ , o critério acima exigiria que as razões  $\frac{N_2}{N} \approx 2.45$  e  $\frac{N_1}{N_2} \approx 2.45$  fossem respeitadas, de maneira que deveriam ser utilizados, no mínimo,  $N_1 \approx 2305$  e  $N_2 \approx 941$ . No entanto, a adoção de  $N_1 = 512$  e  $N_2 = 512$ , ou seja, das razões  $\frac{N_2}{N} \approx 1.33$  e  $\frac{N_1}{N_2} = 1$  leva às curvas dos gráficos (a) (c) e (e) da Figura 4.7, obtidas pelas análises MGC de ordem 32 efetuadas através de uma janela de Blackman (de comprimento  $N = 384$ ) à emissão da vogal “i” por uma voz masculina, respectivamente, com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ . Cada gráfico contém a sobreposição da curva do modelo MGC obtida pelo método *FFT/recursão* [2][7] (em azul) e da mesma curva, obtida pelo método *somente FFT* (em

vermelho). Em um segundo exemplo, o periodograma  $I_N(\omega_m)$  de um bloco de 23.2ms de um sinal de fala amostrado a 22.050kHz, tem comprimento  $N = 512$  amostras. Para  $\alpha = 0.45$ , o critério acima exigiria que as razões  $\frac{N_2}{N} \approx 2.64$  e  $\frac{N_1}{N_2} \approx 2.64$  fossem respeitadas, de maneira que deveriam ser utilizados, no mínimo,  $N_1 \approx 3570$  e  $N_2 \approx 1352$ . No entanto, a adoção de  $N_1 = 512$  e  $N_2 = 512$ , ou seja, das razões  $\frac{N_2}{N} = 1$  e  $\frac{N_1}{N_2} = 1$ , leva às curvas dos gráficos (b) (d) e (f) da Figura 4.7, obtidas pelas análises MGC de ordem 32, efetuadas através de uma janela de Blackman (de comprimento  $N = 512$ ) à mesma emissão do exemplo anterior, respectivamente, com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ . Cada gráfico contém a sobreposição da curva do modelo MGC obtida pelo método *FFT/recursão* [2][7] (em azul), e da mesma curva obtida pelo método *somente FFT* (em vermelho).

Na Figura 4.7, a visualização de degradações nas curvas do método *somente FFT* é possível pela adoção de razões  $\frac{N_2}{N}$  e  $\frac{N_1}{N_2}$  baixas. Quando se aumenta os valores de  $N_2$  e  $N_1$  (mantendo-os como potências de 2 para o ganho do cálculo via FFT), tais degradações deixam de estar visualmente evidentes nas curvas. Degradações devidas a baixos valores de  $\frac{N_2}{N}$  tendem a ocorrer em regiões de frequências altas, já que, para  $\alpha > 0$ , baixos valores de  $N_2$  levam  $\tilde{I}_N(\tilde{\omega})$  a ter sua maior *sub-amostragem* em  $\tilde{\omega} = \pi$ . Já degradações devidas a baixos valores de  $\frac{N_1}{N_2}$  tendem a ocorrer em regiões de baixa frequência, uma vez que o valor  $N_2 > N$ , adotado no sentido de se minimizar uma *sub-amostragem* em torno de  $\tilde{\omega} = \pi$ , leva a uma *super-amostragem* nas proximidades de  $\tilde{\omega} = 0$ , sendo o cálculo de tais amostras por interpolação linear tão mais impreciso quanto menor o valor  $N_1$ . Tais distorções ficam evidentes na Figura 4.7 (b), onde  $\frac{N_2}{N} = \frac{N_1}{N_2} = 1$  e  $\gamma = 0$ . Destaca-se tal valor de  $\gamma$ , uma vez que a análise Mel-Cepstral ( $\gamma = 0$ ) tem mostrado ser a mais sensível a distorções em  $\tilde{I}_N(\tilde{\omega})$ .

Avalia-se que a baixa degradação observada na análise MGC pela adoção de valores inferiores a  $\frac{1+|\alpha|}{1-|\alpha|}$  para a relação  $\frac{N_2}{N}$  decorre do fato de que uma sub-amostragem do periodograma corresponde a um *aliasing* temporal na seqüência de auto-correlação do sinal janelado. Tal seqüência tem um comportamento de queda que leva amostras de longo termo a valores baixos, de modo que sua superposição tem efeitos pequenos. Isso não seria esperado de uma seqüência de fase genérica. Outro fator que reforçaria a

tolerância ao *aliasing*, viria da consideração de que o modelo se baseia em correlações de alcance curto o suficiente, de modo que correlações de longo termo teriam pouco efeito sobre a análise.

As análises e experimentos relacionados ao método proposto encontram-se em estágio inicial e devem ser completados no sentido de se identificar a relação *ganho computacional*  $\times$  *qualidade da representação*. Uma avaliação da ressíntese do sinal de fala, com base nos coeficientes MGC extraídos pelo método, trará uma referência qualitativa inicial para o método. A constatação de economia computacional associada a uma boa qualidade na representação de ambos, o espectro convencional e o espectro STRAIGHT, poderia fazer do método um meio mais veloz para a extração de parâmetros das extensas bases utilizadas no treinamento de sistemas de síntese via HMM. A constatação de bons desempenhos do método, sob a adoção de eixos de frequência distorcidos com base em transformações bilineares (*passa-tudo*) de ordens mais altas, poderia trazer benefícios em algumas aplicações. Um exemplo de aplicação em identificação de locutor, dos coeficientes mel-cepstrais com distorção frequencial baseada em transformação bilineares de ordem 2, está em [40].

Outras aplicações dependeriam de mais investigações e de integração com outras técnicas. A adoção de transformações bilineares (*passa-tudo*) de ordens mais altas, para o mapeamento frequencial, possibilitaria escolhas mais específicas, ao longo do eixo de frequências, da resolução da análise MGC. Tal flexibilidade, associada a uma extração de informações apropriadas do espectro ao longo do eixo de frequências, poderia auxiliar na obtenção de uma envoltória robusta a interferências de F0 e que, ao mesmo tempo, incorpora detalhes relevantes do espectro.

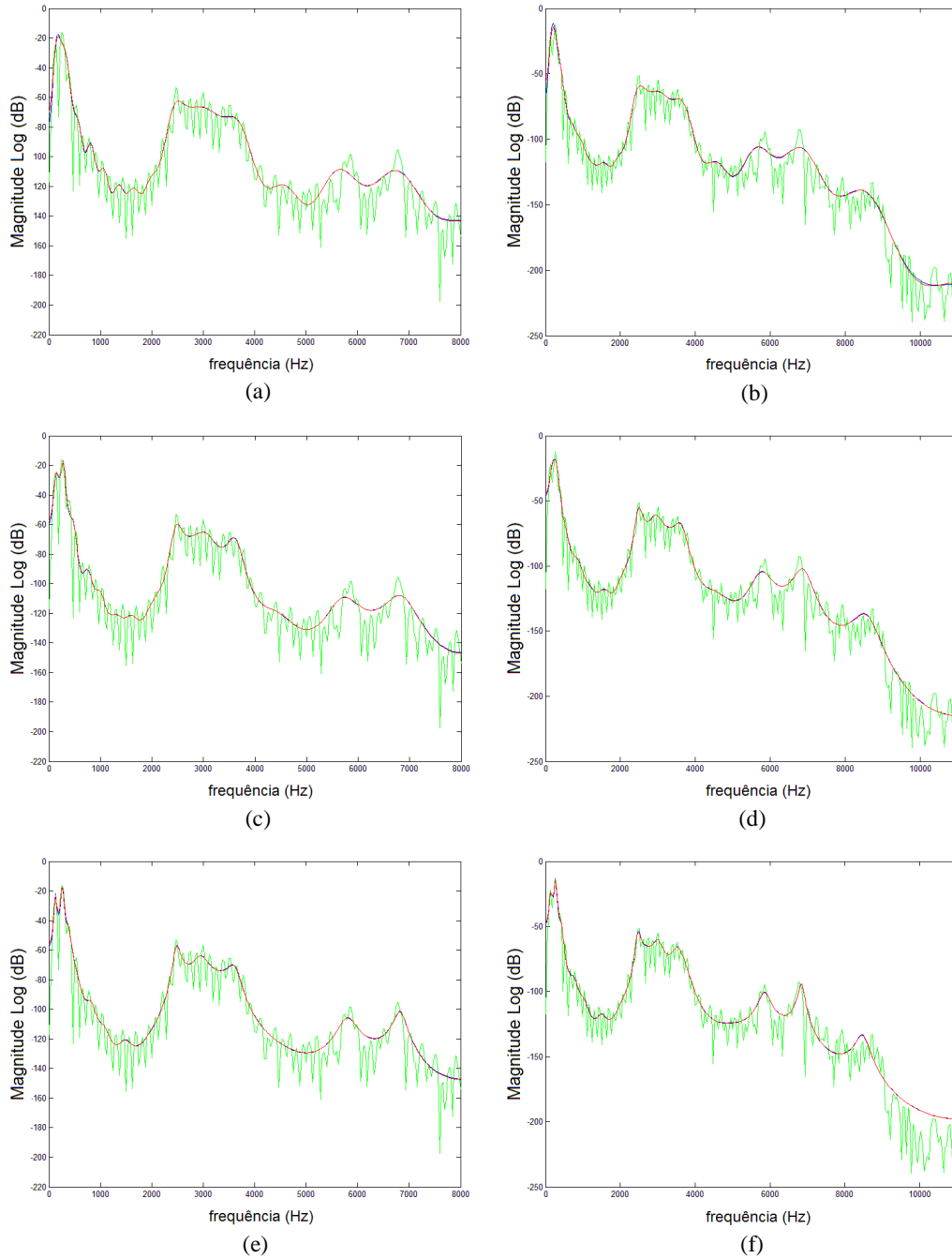


Figura 4.7: Envoltórias espectrais obtidas por análises MGC de ordem 32 aplicadas à emissão da vogal “i” (curva espectral em verde). Curvas em azul: calculadas pelo método *FFT/recursão* [2][7]. Curvas em vermelho: calculadas pelo método *somente FFT*. (a) (c) (e): Análises efetuadas respectivamente com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ , a 16kHz, com  $\alpha = 0.42$ ,  $N = 384$  e  $N_1 = N_2 = 512$ . (b) (d) (f) Análises efetuadas respectivamente com  $\gamma = 0$ ,  $\gamma = -1/2$  e  $\gamma = -1$ , a 22.05kHz com  $\alpha = 0.45$  e  $N = N_1 = N_2 = 512$ .

## 4.9 Conclusão

Este trabalho trouxe uma abordagem do modelo Mel-Cepstral Generalizado e de suas bases, sob prisma resultante de esforços de dedução e interpretação investidos no sentido de compensar a escassez bibliográfica sobre o tema. O trabalho resultou em uma extensa abordagem analítica e computacional do modelo Mel-Cepstral Generalizado, seus casos especiais, e suas bases. Adicionalmente, deu origem à proposta de um método alternativo para a computação dos coeficientes MGC.

A abordagem dada ao cepstrum generalizado tratou da unicidade do espectro logarítmico generalizado, da condição de fase mínima, e demonstrou a relação recursiva para o cálculo do cepstrum-gen. A condição de fase mínima, associada ao cepstrum, foi tratada no contexto amplo do cepstrum generalizado. A solução do modelo GC (e Cepstral-UELS), nesse contexto, foi tratada como um ajuste do cepstrum generalizado de maneira a remover problemas característicos da suavização linear do espectro logarítmico generalizado.

O tratamento dado às seqüências distorcidas em frequência levou à dedução das fórmulas recursivas para o seu cálculo. Uma abordagem dada à análise espectral com resolução não-uniforme permitiu a dedução de um critério para o comprimento de seqüências envolvidas em transformações frequenciais. O mel-cepstrum generalizado foi tratado considerando as transformações frequencial e logarítmica generalizada. A solução do modelo MGC (e Mel-Cepstral), nesse contexto, foi tratada como um ajuste do mel-cepstrum generalizado para remover problemas característicos da suavização linear do espectro logarítmico generalizado.

A abordagem para o modelo MGC foi conduzida de forma unificada até o ponto em que especificidades de metodologia computacional exigiram tratamentos separados a diferentes casos do modelo. A formulação para os cálculos foi toda deduzida nos apêndices. Os métodos para a computação dos casos MGC ( $\alpha \neq 0, \gamma \neq 0$ ) e Mel Cepstral ( $\alpha \neq 0, \gamma = 0$ ) foram tratados em uma seção dedicada, por exigirem a computação de fórmulas recursivas, além das FFTs (os casos GC e Cepstral exigem somente a computação de FFTs), tornando a abordagem mais complexa do que aquelas dos casos GC e Cepstral. Foram tratados os perfis do modelo MGC e o seu ajuste a diferentes características do espectro de fala, e o emprego dos parâmetros MGC em síntese paramétrica estatística.

Propôs-se um método alternativo para a solução do modelo MGC (e Mel-Cepstral). Esse método diferencia-se daquele de [2][7], por contar somente com o cálculo de FFTs a cada iteração do método de Newton-Raphson, sem necessidade do uso de fórmulas recursivas. O método exige que se disponha do periodograma discreto tomado no eixo não-linear de frequências. Constatou-se que a adoção de um periodograma distorcido em frequência, com amostras obtidas por interpolação linear das amostras do periodograma regular, leva a degradações na curva do modelo MGC inferiores às esperadas, com um número de amostras próximo ao comprimento do sinal em análise. As análises e experimentos relacionados ao método proposto encontram-se em estágio inicial e devem ser completados no sentido de se identificar a relação *ganho computacional*  $\times$  *qualidade da representação*.

A constatação de economia computacional associada a uma boa qualidade na representação de ambos o espectro convencional e o espectro STRAIGHT, poderia fazer do método proposto um meio mais veloz para a extração de parâmetros das extensas bases utilizadas no treinamento de sistemas de síntese via HMM. Investigações serão conduzidas no sentido de se verificar o desempenho do método sob a adoção de eixos de frequência distorcidos com base em transformações bilineares (*passa-tudo*) de ordens superiores a 1. A constatação de um bom desempenho *custo-qualitativo* viabilizaria a aplicação da análise com escolhas mais específicas de resolução ao longo do eixo de frequências. Tal flexibilidade, associada a uma extração de informações apropriadas do espectro, poderia auxiliar na obtenção de uma envoltória robusta a interferências de F0 e que, ao mesmo tempo, incorpora detalhes relevantes do espectro.

# Apêndice A

## Função log-gen e sua Inversa

Este apêndice lista algumas propriedades das funções logarítmica generalizada e exponencial generalizada.

### A.1 Função log-gen ( $\ln_\gamma(\cdot)$ )

A função logarítmica generalizada (log-gen) é dada por

$$\ln_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & 0 < |\gamma| \leq 1 \\ \ln(x), & \gamma = 0 \end{cases} \quad (\text{A.1})$$

onde  $\ln_0(x) = \ln(x)$ , já que

$$\lim_{\gamma \rightarrow 0} [\ln_\gamma(x)] = \lim_{\gamma \rightarrow 0} \left( \frac{x^\gamma - 1}{\gamma} \right) = \ln(x) \quad (\text{A.2})$$

#### Demonstração de (A.2)

Expandindo  $\ln_\gamma(x)$  em torno de  $x = 1$ , temos



$$\ln_{\gamma}(x) = x - 1 + \sum_{k=2}^{\infty} \frac{(\gamma - 1)(\gamma - 2) \cdots (\gamma - k + 1)}{k!} (x - 1)^k, \quad (\text{A.3})$$

$$|x - 1| < 1$$

Assim

$$\lim_{\gamma \rightarrow 0} [\ln_{\gamma}(x)] = x - 1 - \frac{(x - 1)^2}{2} + \frac{(x - 1)^3}{3} - \cdots = \ln(x) \quad (\text{A.4})$$

### A.1.1 Relações para a Função log-gen

Considere  $a$  e  $b$ , dois números complexos não-nulos. A relação para o logaritmo generalizado de um produto é dada por

$$\ln_{\gamma}(a \cdot b) = \ln_{\gamma}(a) + \ln_{\gamma}(b) + \gamma \ln_{\gamma}(a) \ln_{\gamma}(b) \quad (\text{A.5})$$

que, para  $\gamma = 0$ , se reduz à relação para o logaritmo de um produto, dada por

$$\ln(a \cdot b) = \ln(a) + \ln(b) \quad (\text{A.6})$$

A relação para o logaritmo generalizado de um quociente é dada por

$$\ln_{\gamma}\left(\frac{a}{b}\right) = \frac{\ln_{\gamma}(a) - \ln_{\gamma}(b)}{1 + \gamma \ln_{\gamma}(b)} \quad (\text{A.7})$$

que, para  $\gamma = 0$ , se reduz à relação para o logaritmo de um quociente, dada por

$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b) \quad (\text{A.8})$$

Considerando  $f(z)$  uma função complexa, a derivada de  $\ln_{\gamma}(f(z))$  é dada por

$$\frac{d}{dz} \ln_{\gamma}(f(z)) = [1 + \gamma \ln_{\gamma}(f(z))] \frac{1}{f(z)} \frac{df(z)}{dz} \quad (\text{A.9})$$

que, para  $\gamma = 0$ , se reduz a

$$\frac{d}{dz} \ln(f(z)) = \frac{1}{f(z)} \frac{df(z)}{dz} \quad (\text{A.10})$$

## A.2 Função exp-gen ( $\exp_\gamma(\cdot)$ )

A inversa da função log-gen, a função exponencial generalizada (exp-gen) é dada por

$$\exp_\gamma(x) = \begin{cases} (1 + \gamma \cdot x)^{1/\gamma}, & 0 < |\gamma| \leq 1 \\ \exp(x), & \gamma = 0 \end{cases} \quad (\text{A.11})$$

onde  $\exp_0(x) = \exp(x)$ , já que

$$\lim_{\gamma \rightarrow 0} [\exp_\gamma(x)] = \lim_{\gamma \rightarrow 0} (1 + \gamma \cdot x)^{1/\gamma} = \exp(x) \quad (\text{A.12})$$

### Demonstração de (A.12)

Expandindo  $\exp_\gamma(x)$  em torno de  $x = 0$ , temos

$$\exp_\gamma(x) = 1 + x + \frac{1}{2!}(1 - \gamma)x^2 + \frac{1}{3!}(1 - \gamma)(1 - 2\gamma)x^3 + \dots \quad (\text{A.13})$$

assim

$$\lim_{\gamma \rightarrow 0} [\exp_\gamma(x)] = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots = \exp(x) \quad (\text{A.14})$$

### A.2.1 Relações para a Função exp-gen

Considere  $a$  e  $b$ , dois números complexos não-nulos. A relação para o produto entre exponenciais generalizadas é dada por

$$\exp_\gamma(a)\exp_\gamma(b) = \exp_\gamma(a + b + \gamma ab) \quad (\text{A.15})$$

que, para  $\gamma = 0$ , se reduz à relação para o produto entre exponenciais, dada por

$$\exp(a)\exp(b) = \exp(a + b) \quad (\text{A.16})$$

A relação para um quociente de exponenciais generalizadas é dada por

$$\frac{\exp_\gamma(a)}{\exp_\gamma(b)} = \exp_\gamma\left(\frac{a - b}{1 + \gamma b}\right) \quad (\text{A.17})$$

que, para  $\gamma = 0$ , se reduz à relação para um quociente de exponenciais, dada por

$$\frac{\exp(a)}{\exp(b)} = \exp(a - b) \quad (\text{A.18})$$

A relação para a exponencial generalizada da soma de dois números é dada por

$$\exp_\gamma(a + b) = \exp_\gamma(a) \exp_\gamma\left(\frac{b}{1 + \gamma a}\right) \quad (\text{A.19})$$

que, para  $\gamma = 0$ , se reduz a (A.16). A exponencial generalizada de uma subtração é dada por

$$\exp_\gamma(a - b) = \frac{\exp_\gamma(a(1 + \gamma b))}{\exp_\gamma(b(1 + \gamma a))} \quad (\text{A.20})$$

que, para  $\gamma = 0$ , se reduz a (A.18). Considerando  $f(z)$  uma função complexa, a derivada de  $\exp_\gamma(f(z))$  é dada por

$$\frac{d}{dz} \exp_\gamma(f(z)) = \frac{1}{1 + \gamma f(z)} \frac{df(z)}{dz} \exp_\gamma(f(z)) \quad (\text{A.21})$$

que, para  $\gamma = 0$ , se reduz a

$$\frac{d}{dz} \exp(f(z)) = \frac{df(z)}{dz} \exp(f(z)) \quad (\text{A.22})$$

## Apêndice B

### Relações úteis para $\Phi_n(e^{j\omega})$

$\Phi_n(e^{j\omega})$  é dado por

$$\Phi_n(e^{j\omega}) = \begin{cases} 1, & n = 0 \\ \frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} [\Psi(e^{j\omega})]^{n-1}, & n > 0 \end{cases} \quad (\text{B.1})$$

com

$$\Psi(e^{j\omega}) = e^{-j\tilde{\omega}(\omega)} = \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \quad (\text{B.2})$$

Considere também

$$\Lambda(e^{j\omega}) = \frac{1}{1 - \alpha e^{-j\omega}} \quad (\text{B.3})$$

As deduções contidas nos próximos apêndices recorrerão às relações envolvendo  $\Phi_n(e^{j\omega})$ ,  $\Psi(e^{j\omega})$  e  $\Lambda(e^{j\omega})$  apontadas a seguir.

#### B.1 Relações entre $\Phi_k(e^{j\omega})$ e $\Psi(e^{j\omega})$

Com base em (B.1), conclui-se que

$$\Phi_{k+n}(e^{j\omega}) = \Phi_k(e^{j\omega})[\Psi(e^{j\omega})]^n, \quad k > 0 \quad (\text{B.4})$$

que leva às relações

$$\Phi_n(e^{j\omega}) = \Phi_1(e^{j\omega})[\Psi(e^{j\omega})]^{n-1}, \quad n > 0 \quad (\text{B.5})$$

e

$$\Phi_n(e^{j\omega}) = \Phi_{n-1}(e^{j\omega}) \cdot \Psi(e^{j\omega}), \quad n > 1 \quad (\text{B.6})$$

$\Phi_n(e^{j\omega})$  pode também ser expresso por

$$\Phi_n(e^{j\omega}) = [\Psi(e^{j\omega})]^n + \alpha[\Psi(e^{j\omega})]^{n-1}, \quad n > 0 \quad (\text{B.7})$$

A partir de tais expressões, relações úteis para os produtos  $\Phi_m(e^{j\omega})\Phi_l(e^{j\omega})$  e  $\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega})$  serão obtidas no próximo item. A relação (B.7) encontra-se deduzida a seguir.

### Dedução da relação (B.7)

Comparando (B.5) com (B.1), conclui-se que

$$\Phi_1(e^{j\omega}) = \frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} \quad (\text{B.8})$$

Tal expressão pode ser manipulada da seguinte maneira

$$\frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} = \frac{e^{-j\omega} - \alpha + \alpha - \alpha^2 e^{-j\omega}}{1 - \alpha e^{-j\omega}} = \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} + \alpha \quad (\text{B.9})$$

Assim

$$\Phi_1(e^{j\omega}) = \Psi(e^{j\omega}) + \alpha \quad (\text{B.10})$$

que, substituída na relação (B.5), fornece

$$\Phi_n(e^{j\omega}) = [\Psi(e^{j\omega}) + \alpha] \cdot [\Psi(e^{j\omega})]^{n-1} \quad (\text{B.11})$$

de onde a relação (B.7) é obtida.

## B.2 Relações para os produtos $\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega})$ e $\Phi_m(e^{j\omega})\Phi_l(e^{j\omega})$

Será útil que se disponha da seguinte relação para  $|\Phi_1(e^{j\omega})|^2$ :

$$|\Phi_1(e^{j\omega})|^2 = (1 - \alpha^2) + \alpha[\Phi_1(e^{j\omega}) + \Phi_1^*(e^{j\omega})] \quad (\text{B.12})$$

cuja dedução encontra-se no final do presente item.

Usando as relações (B.5) e (B.6) nos produtos  $\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega})$  e  $\Phi_m(e^{j\omega})\Phi_l(e^{j\omega})$ , onde  $1 \leq m \leq M$ ,  $1 \leq l \leq M$ , pode-se obter, respectivamente

$$\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega}) = \begin{cases} \Phi_1(e^{j\omega})\Phi_{l-m+1}^*(e^{j\omega}), & 0 \leq l-m \leq M-1 \\ \Phi_1^*(e^{j\omega})\Phi_{m-l+1}(e^{j\omega}), & -M+1 \leq l-m \leq 1 \end{cases} \quad (\text{B.13})$$

e

$$\Phi_m(e^{j\omega})\Phi_l(e^{j\omega}) = \Phi_1(e^{j\omega})\Phi_{l+m-1}(e^{j\omega}), \quad 2 \leq l+m \leq 2M \quad (\text{B.14})$$

Tais produtos podem ser expressos nas formas

$$\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega}) = \begin{cases} \Phi_{l-m}^*(e^{j\omega}) + \alpha\Phi_{l-m+1}^*(e^{j\omega}), & 1 \leq l-m \leq M-1 \\ (1 - \alpha^2) + \alpha[\Phi_1(e^{j\omega}) + \Phi_1^*(e^{j\omega})] & l-m = 0 \\ \Phi_{m-l}(e^{j\omega}) + \alpha\Phi_{m-l+1}(e^{j\omega}), & -M+1 \leq l-m \leq -1 \end{cases} \quad (\text{B.15})$$

e

$$\Phi_m(e^{j\omega})\Phi_l(e^{j\omega}) = \Phi_{l+m}(e^{j\omega}) + \alpha\Phi_{l+m-1}(e^{j\omega}), \quad 2 \leq l+m \leq 2M \quad (\text{B.16})$$

A expressão (B.15), para  $l-m=0$ , pode ser obtida utilizando (B.12) em (B.13). As relações (B.15), para  $1 \leq |l-m| \leq M-1$ , e (B.16) são obtidas manipulando as relações (B.13) e (B.14) através de (B.7) (com  $n=1$ ) e (B.6).

### Dedução da relação (B.12)

Adotando  $n=1$  em (B.7), obtém-se

$$\Phi_1(e^{j\omega}) = \Psi(e^{j\omega}) + \alpha \quad (\text{expressão (B.10)})$$

Assim  $|\Phi_1(e^{j\omega})|^2$  é dado por

$$|\Phi_1(e^{j\omega})|^2 = 1 + \alpha^2 + \alpha[\Psi(e^{j\omega}) + \Psi^*(e^{j\omega})] \quad (\text{B.17})$$

que pode ser rearranjada da seguinte maneira

$$|\Phi_1(e^{j\omega})|^2 = 1 - \alpha^2 + \alpha\{[\Psi(e^{j\omega}) + \alpha] + [\Psi^*(e^{j\omega}) + \alpha]\} \quad (\text{B.18})$$

onde a substituição de (B.10) fornece (B.12).

### B.3 Relações envolvendo $\Lambda(e^{j\omega})$

$\Phi_1(e^{j\omega})$  (eq. (B.8)) está relacionada a  $\Lambda(e^{j\omega})$  (eq. (B.3)) por

$$\Phi_1(e^{j\omega}) = \frac{1 - \alpha^2}{\alpha} [\Lambda(e^{j\omega}) - 1] \quad (\text{B.19})$$

expressão que, invertida, fornece

$$\Lambda(e^{j\omega}) = 1 + \frac{\alpha}{1 - \alpha^2} \Phi_1(e^{j\omega}) \quad (\text{B.20})$$

O produto  $\Phi_1^*(e^{j\omega})\Phi_{k+1}(e^{j\omega})$  pode ser expresso por

$$\Phi_1^*(e^{j\omega})\Phi_{k+1}(e^{j\omega}) = \begin{cases} \alpha\Phi_1^*(e^{j\omega}) + (1 - \alpha^2)\Lambda(e^{j\omega}), & k = 0 \\ (1 - \alpha^2)\Lambda(e^{j\omega})\Phi_k(e^{j\omega}), & k > 0 \end{cases} \quad (\text{B.21})$$

Em termos de  $\Psi(e^{j\omega})$ ,  $\Lambda(e^{j\omega})$  é dado por

$$\Lambda(e^{j\omega}) = \frac{1}{1 - \alpha^2} [1 + \alpha\Psi(e^{j\omega})] \quad (\text{B.22})$$

Tais relações encontram-se deduzidas a seguir.

#### Dedução da relação (B.19)

A partir da expressão (B.8) faz-se o seguinte desenvolvimento

$$\begin{aligned} \Phi_1(e^{j\omega}) &= \frac{(1 - \alpha^2)e^{-j\omega}}{1 - \alpha e^{-j\omega}} = \frac{1 - \alpha^2}{\alpha} \left( \frac{\alpha e^{-j\omega}}{1 - \alpha e^{-j\omega}} \right) = \\ &= \frac{1 - \alpha^2}{\alpha} \left( \frac{-1 + \alpha e^{-j\omega} + 1}{1 - \alpha e^{-j\omega}} \right) = \frac{1 - \alpha^2}{\alpha} \left[ \frac{1}{1 - \alpha e^{-j\omega}} - 1 \right] \end{aligned} \quad (\text{B.23})$$

e (B.19) é obtido.

### Dedução da relação (B.21)

Para  $k = 0$ , o lado esquerdo da igualdade na expressão (B.21) resulta em  $|\Phi_1(e^{j\omega})|^2$ . Substituindo a expressão (B.19) na expressão para  $|\Phi_1(e^{j\omega})|^2$  dada por (B.12), obtém-se a expressão (B.21) ( $k = 0$ ).

Para  $k > 0$ , o lado esquerdo da igualdade na expressão (B.21) pode ser, com base em (B.6), expresso por

$$\Phi_1^*(e^{j\omega})\Phi_{k+1}(e^{j\omega}) = \Phi_1^*(e^{j\omega})\Psi(e^{j\omega})\Phi_k(e^{j\omega}) \quad (\text{B.24})$$

onde, explicitando o produto  $\Phi_1^*(e^{j\omega})\Psi(e^{j\omega})$ , e desenvolvendo, temos

$$\begin{aligned} \Phi_1^*(e^{j\omega})\Psi(e^{j\omega}) &= \frac{(1 - \alpha^2)e^{j\omega}}{1 - \alpha e^{j\omega}} \left( \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \right) = \\ &= \frac{(1 - \alpha^2)}{e^{-j\omega} - \alpha} \left( \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \right) = \frac{1 - \alpha^2}{1 - \alpha e^{-j\omega}} \end{aligned} \quad (\text{B.25})$$

Assim

$$\Phi_1^*(e^{j\omega})\Psi(e^{j\omega}) = (1 - \alpha^2)\Lambda(e^{j\omega}) \quad (\text{B.26})$$

que, substituída em (B.24), fornece a expressão (B.21) para  $k > 0$ .

### Dedução da relação (B.22)

A partir de (B.3) faz-se o seguinte desenvolvimento

$$\begin{aligned} \Lambda(e^{j\omega}) &= \frac{1}{1 - \alpha e^{-j\omega}} = \frac{1}{1 - \alpha^2} \left( \frac{1 - \alpha^2}{1 - \alpha e^{-j\omega}} \right) = \\ &= \frac{1}{1 - \alpha^2} \left( \frac{1 - \alpha e^{-j\omega} - \alpha^2 + \alpha e^{-j\omega}}{1 - \alpha e^{-j\omega}} \right) = \frac{1}{1 - \alpha^2} \left[ 1 + \alpha \frac{e^{-j\omega} - \alpha}{1 - \alpha e^{-j\omega}} \right] \end{aligned} \quad (\text{B.27})$$

e (B.22) é obtido.



## Apêndice C

### Modelo MGC Baseado nos Coeficientes $b_\gamma(n)$ e Transformações $\tilde{c}_\gamma(n) \Leftrightarrow b_\gamma(n)$ e $c_\gamma(n) \Leftrightarrow b_\gamma(n)$

Este Apêndice demonstra como se pode chegar a uma representação para o modelo MGC com base no conjunto de coeficientes transformados  $b_\gamma(n)$ . Aborda-se a transformação que leva  $\tilde{c}_\gamma(n)$  em  $b_\gamma(n)$  e uma transformação que fornece  $b_\gamma(n)$  diretamente a partir dos coeficientes  $c_\gamma(n)$  (GC).

#### C.1 Representação do Modelo MGC com base nos Coeficientes $b_\gamma(n)$

O modelo para a análise MGC, em função dos coeficientes  $\tilde{c}_\gamma(n)$ , é dado por

$$H(z) = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) [\tilde{z}(z)]^{-n} \right) \quad (\text{C.1})$$

onde

$$[\tilde{z}(z)]^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (\text{C.2})$$

Como  $H(z)$  é causal, os termos com atraso podem ser eliminados do interior de  $\exp_\gamma(\cdot)$  aplicando-se  $\lim_{|z| \rightarrow \infty} H(z)$ . Tal limite leva à *zero-ésima* amostra da resposta ao impulso de  $H(z)$ , que é a constante de ganho  $K$  do modelo. Assim

$$K = \lim_{|z| \rightarrow \infty} \left[ \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) [\tilde{z}(z)]^{-n} \right) \right] = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n \right) \quad (\text{C.3})$$

já que

$$\lim_{|z| \rightarrow \infty} [\tilde{z}(z)]^{-1} = \lim_{|z| \rightarrow \infty} \left( \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \right) = -\alpha \quad (\text{C.4})$$

Em (C.3) pode-se ver que o termo livre de atraso  $\sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n$  envolve todos os coeficientes MGC. Para a aplicação do método de cálculo dos coeficientes MGC, o modelo deve ter o termo livre de atraso representado somente por um coeficiente. Os passos a seguir mostrarão como se pode chegar uma representação do modelo que contemple tal característica. A demonstração ficará facilitada assumindo-se, por hora, a função  $\tilde{z}(z)$  como variável independente ( $\tilde{z}(z) = \tilde{z}$ ). Assim expressaremos (C.1) na forma

$$\tilde{H}(\tilde{z}) = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) \tilde{z}^{-n} \right) \quad (\text{C.5})$$

de maneira  $H(z) = \tilde{H}(\tilde{z}(z))$ . Como ponto de partida para uma representação alternativa para  $H(z)$ , considere (C.5) na forma

$$\tilde{H}(\tilde{z}) = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n + B_a(\tilde{z}) \right) \quad (\text{C.6})$$

onde  $\sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n$  é o termo livre de atraso e

$$B_a(\tilde{z}) = \sum_{n=0}^M \tilde{c}_\gamma(n) \tilde{z}^{-n} - \sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n \quad (\text{C.7})$$

corresponde, ainda que não explicitamente, à porção da estrutura de  $H(z)$  com atrasos, já que  $\lim_{|z| \rightarrow \infty} B_a(\tilde{z}(z)) = 0$ . Expressando  $B_a(\tilde{z})$  por

$$B_a(\tilde{z}) = \sum_{n=1}^M \tilde{c}_\gamma(n) [\tilde{z}^{-n} - (-\alpha)^n] \quad (\text{C.8})$$

o fator  $[[\tilde{z}]^{-n} - (-\alpha)^n]$ , por ser um número binomial, pode ser representado na forma

$$\tilde{z}^{-n} - (-\alpha)^n = \frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}} \sum_{m=1}^n (-\alpha)^{n-m} \tilde{z}^{-m} \quad (\text{C.9})$$

como enunciado no final deste tópic. Assim

$$B_a(\tilde{z}) = \frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}} \left( \sum_{n=1}^M \tilde{c}_\gamma(n) \sum_{m=1}^n (-\alpha)^{n-m} \tilde{z}^{-m} \right) \quad (\text{C.10})$$

Considerando que  $\tilde{\mathbf{c}}_1 = [\tilde{c}_\gamma(1) \tilde{c}_\gamma(2) \dots \tilde{c}_\gamma(M)]^T$ ,  $\tilde{\mathbf{z}}_- = [\tilde{z}^{-1} \tilde{z}^{-2} \dots \tilde{z}^{-M}]^T$  e

$$\mathbf{A}_M = \begin{bmatrix} 1 & (-\alpha)^1 & (-\alpha)^2 & (-\alpha)^3 & \dots & (-\alpha)^{M-1} \\ 0 & 1 & (-\alpha)^1 & (-\alpha)^2 & \dots & (-\alpha)^{M-2} \\ 0 & 0 & 1 & (-\alpha)^1 & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \ddots & (-\alpha)^2 \\ \vdots & \vdots & \ddots & \ddots & \ddots & (-\alpha)^1 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \quad (\text{C.11})$$

então, com base em  $\tilde{\mathbf{c}}_1$ ,  $\tilde{\mathbf{z}}_-$  e  $\mathbf{A}_M$ , o fator entre parênteses na expressão (C.10) é representado por

$$(\tilde{\mathbf{z}}_-^T \mathbf{A}_M) \tilde{\mathbf{c}}_1 = \sum_{n=1}^M \tilde{c}_\gamma(n) \sum_{m=1}^n (-\alpha)^{n-m} \tilde{z}^{-m} \quad (\text{C.12})$$

onde o vetor  $\tilde{\mathbf{z}}_-^T \mathbf{A}_M$  tem a componente  $n$  dada por

$$(\tilde{\mathbf{z}}_-^T \mathbf{A}_M)_{1,n} = \sum_{m=1}^n (-\alpha)^{n-m} \tilde{z}^{-m} \quad (\text{C.13})$$

Pela propriedade associativa

$$(\tilde{\mathbf{z}}_-^T \mathbf{A}_M) \tilde{\mathbf{c}}_1 = \tilde{\mathbf{z}}_-^T (\mathbf{A}_M \tilde{\mathbf{c}}_1) \quad (\text{C.14})$$

onde é imediato que o lado direito da igualdade pode ser expresso por

$$\tilde{\mathbf{z}}_-^T (\mathbf{A}_M \tilde{\mathbf{c}}_1) = \sum_{n=1}^M \tilde{z}^{-n} \sum_{m=n}^M (-\alpha)^{m-n} \tilde{c}_\gamma(m) \quad (\text{C.15})$$

já que

$$(\mathbf{A}_M \tilde{\mathbf{c}}_1)_{n,1} = \sum_{m=n}^M (-\alpha)^{m-n} \tilde{c}_\gamma(m) \quad (\text{C.16})$$

Assim, verifica-se que

$$\sum_{n=1}^M \tilde{c}_\gamma(n) \sum_{m=1}^n (-\alpha)^{n-m} \tilde{z}^{-m} = \sum_{n=1}^M \tilde{z}^{-n} \sum_{m=n}^M (-\alpha)^{m-n} \tilde{c}_\gamma(m) \quad (\text{C.17})$$

Substituindo (C.17) em (C.10), obtém-se

$$B_a(\tilde{z}) = \frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}} \left( \sum_{n=1}^M \tilde{z}^{-n} \sum_{m=n}^M (-\alpha)^{m-n} \tilde{c}_\gamma(m) \right) \quad (\text{C.18})$$

que, por sua vez, substituída em (C.6) com os fatores rearranjados, fornece

$$\tilde{H}(\tilde{z}) = \exp_\gamma \left( \sum_{n=0}^M \tilde{c}_\gamma(n) (-\alpha)^n + \frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}} \left[ \sum_{n=1}^M \left( \sum_{m=n}^M (-\alpha)^{m-n} \tilde{c}_\gamma(m) \right) \tilde{z}^{-n} \right] \right) \quad (\text{C.19})$$

ou

$$\tilde{H}(\tilde{z}) = \exp_\gamma \left( b_\gamma(0) + \frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}} \left( \sum_{n=1}^M b_\gamma(n) \tilde{z}^{-n} \right) \right) \quad (\text{C.20})$$

onde

$$b_\gamma(n) = \sum_{m=n}^M (-\alpha)^{m-n} \tilde{c}_\gamma(m), \quad n = 0, 1, \dots, M \quad (\text{C.21})$$

Substituindo  $\tilde{z}^{-1} = [\tilde{z}(z)]^{-1}$  no fator  $\frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}}$  obtém-se

$$\frac{\tilde{z}^{-1} + \alpha}{\tilde{z}^{-1}} = \frac{(1 - \alpha^2)z^{-1}}{1 - \alpha z^{-1}} \tilde{z}(z) \quad (\text{C.22})$$

que substituída em (C.20) e posicionada no interior do somatório, fornece

$$H(z) = \exp_{\gamma} \left( b_{\gamma}(0) + \sum_{n=1}^M b_{\gamma}(n) \frac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}} [\tilde{z}(z)]^{-(n-1)} \right) \quad (\text{C.23})$$

Assim, o modelo para a análise MGC pode ser expresso por

$$H(z) = \exp_{\gamma} \left( \sum_{n=0}^M b_{\gamma}(n) \Phi_n(z) \right) \quad (\text{C.24})$$

onde

$$\Phi_n(z) = \begin{cases} 1, & n = 0 \\ \frac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}} [\tilde{z}(z)]^{-(n-1)}, & n > 0 \end{cases} \quad (\text{C.25})$$

O modelo na forma (C.24) tem o termo livre de atraso dado pelo coeficiente  $b_{\gamma}(0)$ .

### Número binomial

Um número binomial tem a forma  $x^n \pm y^n$  sendo que  $x^n - y^n$  pode ser fatorado da seguinte forma

$$x^n - y^n = (x - y) \left( \sum_{k=0}^{n-1} x^k y^{n-1-k} \right) \quad (\text{C.26})$$

onde efetuando a substituição  $m = k + 1$ , obtém-se

$$x^n - y^n = \frac{x - y}{x} \sum_{m=1}^n x^m y^{n-m} \quad (\text{C.27})$$

Substituindo  $x = \tilde{z}^{-1}$  e  $y = -\alpha$  em (C.27), obtém-se a expressão (C.9).

## C.2 Transformação $\tilde{c}_{\gamma}(n) \Leftrightarrow b_{\gamma}(n)$

Os coeficientes  $b_{\gamma}(n)$  são obtidos a partir dos coeficientes  $\tilde{c}_{\gamma}(n)$  pela transformação

$$\mathbf{b}_\gamma = \mathbf{A}_{M+1} \tilde{\mathbf{c}}_\gamma \quad (\text{C.28})$$

que corresponde à forma matricial de (C.21). Os coeficientes  $\tilde{c}_\gamma(n)$ , por sua vez são obtidos a partir dos coeficientes  $b_\gamma(n)$  pela transformação inversa

$$\tilde{\mathbf{c}}_\gamma = \mathbf{A}_{M+1}^{-1} \mathbf{b}_\gamma \quad (\text{C.29})$$

com  $\mathbf{b}_\gamma = [b_\gamma(0) b_\gamma(1) b_\gamma(2) \dots b_\gamma(M)]^T$  e  $\tilde{\mathbf{c}}_\gamma = [\tilde{c}_\gamma(0) \tilde{\mathbf{c}}_1^T]^T$ , ambos de dimensão  $M + 1$ . As matrizes  $\mathbf{A}_{M+1}$  e  $\mathbf{A}_{M+1}^{-1}$ , de dimensão  $(M + 1) \times (M + 1)$ , são dadas por

$$\mathbf{A}_{M+1} = \begin{bmatrix} 1 & (-\alpha)^1 & (-\alpha)^2 & (-\alpha)^3 & \dots & (-\alpha)^M \\ 0 & 1 & (-\alpha)^1 & (-\alpha)^2 & \dots & (-\alpha)^{M-1} \\ 0 & 0 & 1 & (-\alpha)^1 & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \ddots & (-\alpha)^2 \\ \vdots & \vdots & \ddots & \ddots & \ddots & (-\alpha)^1 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \quad (\text{C.30})$$

e

$$\mathbf{A}_{M+1}^{-1} = \begin{bmatrix} 1 & \alpha & 0 & 0 & \dots & 0 \\ 0 & 1 & \alpha & 0 & \dots & 0 \\ 0 & 0 & 1 & \alpha & \ddots & \vdots \\ 0 & 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \alpha \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \quad (\text{C.31})$$

Com base em (C.29), os coeficientes  $\tilde{c}_\gamma(n)$  podem ser obtidos a partir dos coeficientes  $b_\gamma(n)$  pela fórmula

$$\tilde{c}_\gamma(n) = \begin{cases} b_\gamma(n), & n = M \\ b_\gamma(n) + \alpha b_\gamma(n+1), & 0 \leq n < M \end{cases} \quad (\text{C.32})$$

que, invertida, fornece a fórmula para a obtenção de  $b_\gamma(n)$  a partir de  $\tilde{c}_\gamma(n)$

$$b_\gamma(n) = \begin{cases} \tilde{c}_\gamma(n), & n = M \\ \tilde{c}_\gamma(n) - \alpha b_\gamma(n+1), & 0 \leq n < M \end{cases} \quad (\text{C.33})$$

### C.3 Transformação $c_\gamma(n) \rightleftharpoons b_\gamma(n)$

Os coeficientes MGC transformados  $b_\gamma(n)$  podem ser obtidos diretamente a partir dos coeficientes cepstrais generalizados  $c_\gamma(k)$ , o que será demonstrado a seguir. Para o desenvolvimento, será útil que  $\Phi_n(z)$  seja expresso na forma

$$\Phi_n(z) = \begin{cases} 1 & n = 0 \\ \Phi_1(z)[\Psi(z)]^{n-1} & n > 0 \end{cases} \quad (\text{C.34})$$

com

$$\Phi_1(z) = \frac{(1 - \alpha^2)z^{-1}}{1 - \alpha z^{-1}} \quad (\text{C.35})$$

e

$$\Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (\text{C.36})$$

Com base na expressão (3.20) para a obtenção de uma seqüência distorcida em frequência, o mel-cepstrum generalizado é obtido a partir do cepstrum generalizado por

$$\tilde{c}_\gamma(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) \Lambda(e^{j\omega}) \Phi_n(e^{j\omega}) d\omega \quad (\text{C.37})$$

onde

$$C_\gamma(e^{j\omega}) = \sum_{k=0}^{\infty} c_\gamma(k) e^{-j\omega k} \quad (\text{C.38})$$

Em (C.37), substituindo (B.20) para  $n = 0$ , e (B.22) para  $n > 0$ , obtém-se

$$\tilde{c}_\gamma(n) = \begin{cases} \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) d\omega + \alpha \left[ \frac{1}{1 - \alpha^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) \Phi_1(e^{j\omega}) d\omega \right] \right] & n = 0 \\ \left[ \frac{1}{1 - \alpha^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) \Phi_n(e^{j\omega}) d\omega + \alpha \left[ \frac{1}{1 - \alpha^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) \Phi_{n+1}(e^{j\omega}) d\omega \right] \right] & n > 0 \end{cases} \quad (\text{C.39})$$

onde para se determinar a expressão  $\tilde{c}_\gamma(n)$  ( $n > 0$ ) faz-se uso de (B.6). Comparando (C.39) com  $\tilde{c}_\gamma(n) = b_\gamma(n) + \alpha b_\gamma(n+1)$  (expressão (C.32)), verifica-se que

$$b_\gamma(n) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) d\omega & n = 0 \\ \frac{1}{1-\alpha^2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) \Phi_n(e^{j\omega}) d\omega \right) & n > 0 \end{cases} \quad (\text{C.40})$$

Tais integrais são obtidas assumindo  $i = 0$  nas transformadas inversas de Fourier

$$b_\gamma(n, i) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) e^{j\omega i} d\omega & n = 0 \\ \frac{1}{1-\alpha^2} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} C_\gamma^*(e^{j\omega}) \Phi_n(e^{j\omega}) e^{j\omega i} d\omega \right) & n > 0 \end{cases} \quad (\text{C.41})$$

Assim

$$b_\gamma(n, 0) = b_\gamma(n) \quad (\text{C.42})$$

Com base na forma (C.34) de  $\Phi_n(e^{j\omega})$ , conclui-se que a expressão (C.41) é contemplada pela filtragem em cascata representada na Figura C.1.

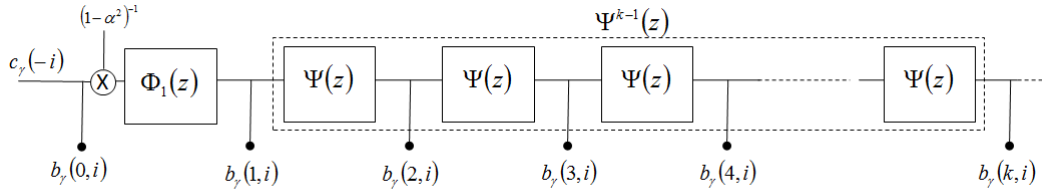


Figura C.1: Banco de filtros em cascata com entrada  $c_\gamma(-i)$  e saída  $b_\gamma(k, i)$  em sua  $k$ -ésima seção. Os valores  $b_\gamma(k, 0)$  das seqüências de saída constituem a seqüência  $b_\gamma(k)$ .

A filtragem da Figura C.1 pode ser computada pela fórmula recursiva

$$b_\gamma(k, i) = \begin{cases} c_\gamma(-i) & k = 0 \\ b_\gamma(0, i-1) + \alpha b_\gamma(1, i-1), & k = 1 \\ b_\gamma(k-1, i-1) + \alpha (b_\gamma(k, i-1) - b_\gamma(k-1, i)), & k = 2, 3, \dots, M \end{cases} \quad (\text{C.43})$$

$$i = -N', \dots, -2, -1, 0$$

$$b_\gamma(k) = b_\gamma(k, 0)$$

que pode ser deduzida com base em (C.35) e (C.36)<sup>§</sup>. Assumindo  $\mathbf{c}_{\gamma, N'} = [c_\gamma(0) \ c_\gamma(1) \ \dots \ c_\gamma(N')]^T$  e  $\mathbf{b}_\gamma = [b_\gamma(0) \ b_\gamma(1) \ \dots \ b_\gamma(M)]^T$ , a transformação que leva  $\mathbf{c}_{\gamma, N'}$  em  $\mathbf{b}_\gamma$  através de (C.43) é representada por<sup>§§</sup>

<sup>§</sup> A dedução é feita nos mesmos moldes daquela para a fórmula (3.29) na seção 3.4.1.



$$\mathbf{b}_\gamma = \frac{1}{(1 - \alpha^2)} T_1(\mathbf{c}_{\gamma, N'}, N', M, \alpha) \quad (\text{C.44})$$

Para o cálculo de  $\mathbf{c}_{\gamma, N'}$  a partir de  $\mathbf{b}_\gamma$ , utiliza-se a fórmula

$$c_\gamma(k, i) = \begin{cases} b_\gamma(-i) & k = 0 \\ (1 - \alpha^2)c_\gamma(0, i - 1) - \alpha c_\gamma(1, i - 1), & k = 1 \\ c_\gamma(k - 1, i - 1) - \alpha (c_\gamma(k, i - 1) - c_\gamma(k - 1, i)), & k = 2, 3, \dots, N' \end{cases} \quad (\text{C.45})$$

$$i = -M, \dots, -2, -1, 0$$

$$c_\gamma(k) = c_\gamma(k, 0)$$

A transformação que leva  $\mathbf{b}_\gamma$  em  $\mathbf{c}_{\gamma, N'}$  através de (C.45) é representada por

$$\mathbf{c}_{\gamma, N'} = T_1(\mathbf{b}_\gamma, M, N', -\alpha) \quad (\text{C.46})$$

---

<sup>§§</sup> A transformação  $T_1(\cdot)$  também é empregada no método de Newton-Raphson para a determinação do gradiente do critério  $\varepsilon$  (casos MGC e Mel-Cepstral) e aparece no Apêndice E em sua forma padrão dada pela fórmula recursiva (E.15).

# Apêndice D

## Gradiente e Hessiana de $\varepsilon$

Neste apêndice é feita a dedução do vetor gradiente  $\nabla_{\mathbf{b}'}\varepsilon$  e da matriz Hessiana  $\mathbf{H}$  do critério  $\varepsilon$ , dado por

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (\text{D.1})$$

onde

$$D(e^{j\omega}) = \exp_{\gamma} \left( \sum_{n=1}^M b_{\gamma}'(n) \Phi_n(e^{j\omega}) \right) \quad (\text{D.2})$$

As deduções apresentadas têm base nas relações para  $\Phi_n(e^{j\omega})$  dadas no Apêndice B.

### D.1 Análise Mel-Cepstral Generalizada ( $\alpha \neq 0, \gamma \neq 0$ )

Aqui são deduzidas as componentes  $\frac{\partial \varepsilon}{\partial b_{\gamma}'(l)}$  do vetor gradiente  $\nabla_{\mathbf{b}'}\varepsilon$  e os elementos  $\frac{\partial^2 \varepsilon}{\partial b_{\gamma}'(m) \partial b_{\gamma}'(l)}$  da matriz Hessiana  $\mathbf{H}$  do critério  $\varepsilon$ . Para tal, será útil o uso da expressão

$$D^{\gamma}(e^{j\omega}) = 1 + \gamma \sum_{n=1}^M b_{\gamma}'(n) \Phi_n(e^{j\omega}) \quad (\text{D.3})$$

e de sua derivada parcial

$$\frac{\partial D^\gamma(e^{j\omega})}{\partial b_{\gamma'}(l)} = \gamma \cdot \Phi_l(e^{j\omega}) \quad (\text{D.4})$$

### D.1.1 Cálculo do Vetor Gradiente $\nabla_{b'} \varepsilon$

A componente  $l$  do vetor gradiente é dada por

$$\frac{\partial \varepsilon}{\partial b_{\gamma'}(l)} = \frac{\partial}{\partial b_{\gamma'}(l)} \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} I_N(\omega) \frac{\partial}{\partial b_{\gamma'}(l)} \left( \frac{1}{|D(e^{j\omega})|^2} \right) d\omega \quad (\text{D.5})$$

Assim, o problema de se encontrar  $\frac{\partial \varepsilon}{\partial b_{\gamma'}(l)}$  reduz-se à determinação de  $\frac{\partial}{\partial b_{\gamma'}(l)} \left( \frac{1}{|D(e^{j\omega})|^2} \right)$  que, em termos de  $D^\gamma(e^{j\omega})$ , tem a forma

$$\frac{\partial}{\partial b_{\gamma'}(l)} \left( \frac{1}{|D(e^{j\omega})|^2} \right) = \frac{\partial}{\partial b_{\gamma'}(l)} \left[ \left( |D^\gamma(e^{j\omega})|^2 \right)^{-\frac{1}{\gamma}} \right] \quad (\text{D.6})$$

Com  $u = |D^\gamma(e^{j\omega})|^2$  a expressão (D.6) assume a forma

$$\frac{\partial}{\partial b_{\gamma'}(l)} \left( u^{-\frac{1}{\gamma}} \right) = \frac{d}{du} \left( u^{-\frac{1}{\gamma}} \right) \frac{\partial u}{\partial b_{\gamma'}(l)} \quad (\text{D.7})$$

onde

$$\frac{d}{du} \left( u^{-\frac{1}{\gamma}} \right) = -\frac{1}{\gamma} u^{-\left(\frac{1+\gamma}{\gamma}\right)} = -\frac{1}{\gamma} \left( \frac{1}{|D^{1+\gamma}(e^{j\omega})|^2} \right) \quad (\text{D.8})$$

e

$$\frac{\partial u}{\partial b_{\gamma'}(l)} = \frac{\partial \left( |D^\gamma(e^{j\omega})|^2 \right)}{\partial b_{\gamma'}(l)} = D^\gamma(e^{j\omega}) \frac{\partial D^{*\gamma}(e^{j\omega})}{\partial b_{\gamma'}(l)} + D^{*\gamma}(e^{j\omega}) \frac{\partial D^\gamma(e^{j\omega})}{\partial b_{\gamma'}(l)} \quad (\text{D.9})$$

Substituindo (D.3) e (D.4) na última expressão, obtém-se

$$\frac{\partial u}{\partial b_{\gamma'}(l)} = \gamma \left( D^\gamma(e^{j\omega}) \Phi_l^*(e^{j\omega}) + D^{*\gamma}(e^{j\omega}) \Phi_l(e^{j\omega}) \right) \quad (\text{D.10})$$

Substituindo (D.10) e (D.8) em (D.7), obtém-se

$$\frac{\partial}{\partial b_{\gamma'}(l)} \left( \frac{1}{|D(e^{j\omega})|^2} \right) = -\frac{1}{|D^{1+\gamma}(e^{j\omega})|^2} \left( D^\gamma(e^{j\omega}) \Phi_l^*(e^{j\omega}) + D^{*\gamma}(e^{j\omega}) \Phi_l(e^{j\omega}) \right) \quad (\text{D.11})$$

que, por sua vez, substituída em (D.5) fornece

$$\begin{aligned} \frac{\partial \varepsilon}{\partial b_{\gamma'}(l)} = & -\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{\gamma}(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \\ & - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{*\gamma}(e^{j\omega}) \Phi_l(e^{j\omega}) d\omega \end{aligned} \quad (\text{D.12})$$

Considerando que o fator  $\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2}$  dos integrandos dos termos de (D.12) é real, é possível concluir que um termo corresponde ao complexo conjugado do outro. Dado que as integrais são reais, já que seus integrandos obedecem às condições de simetria para tal, os termos são equivalentes. Assim  $\frac{\partial \varepsilon}{\partial b_{\gamma'}(l)}$  é dado por

$$\frac{\partial \varepsilon}{\partial b_{\gamma'}(l)} = -2 \cdot \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{\gamma}(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \right] \quad (\text{D.13})$$

Assim, o vetor gradiente é dado por

$$\mathbf{\nabla}_{b'} \varepsilon = -2 \cdot \mathbf{r} \quad (\text{D.14})$$

onde

$$\mathbf{r} = [\tilde{r}_{\gamma}(1) \quad \tilde{r}_{\gamma}(2) \quad \tilde{r}_{\gamma}(3) \quad \dots \quad \tilde{r}_{\gamma}(M)]^T \quad (\text{D.15})$$

com

$$\tilde{r}_{\gamma}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{\gamma}(e^{j\omega}) \Phi_k^*(e^{j\omega}) d\omega, \quad k = 1, 2, \dots, M \quad (\text{D.16})$$

Assumindo  $k = 0$  em (D.16), obtém-se

$$\tilde{r}_{\gamma}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{\gamma}(e^{j\omega}) d\omega \quad (\text{D.17})$$

que, em conjunto com  $\mathbf{r}$ , é utilizado no cálculo de  $\varepsilon$ , a cada iteração do método de Newton Raphson (ver Apêndice E, seção E.1.4).

### D.1.2 Cálculo da Matriz Hessiana $\mathbf{H}$

O elemento  $\frac{\partial^2 \varepsilon}{\partial b_{\gamma'}(m) \partial b_{\gamma'}(l)}$  da linha  $m$  e da coluna  $l$  da matriz Hessiana  $\mathbf{H}$  é dado por

$$\frac{\partial^2 \varepsilon}{\partial b_{\gamma'}(m) \partial b_{\gamma'}(l)} = \frac{\partial}{\partial b_{\gamma'}(m)} (-2 \cdot \tilde{r}_{\gamma}(l)) = \quad (\text{D.18})$$

$$= -2. \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} I_N(\omega) \frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{D^{\gamma}(e^{j\omega})}{|D^{1+\gamma}(e^{j\omega})|^2} \right) \Phi_l^*(e^{j\omega}) d\omega \right]$$

Assim, o problema de se encontrar  $\frac{\partial^2 \varepsilon}{\partial b_{\gamma'}(m) \partial b_{\gamma'}(l)}$  é reduzido à determinação de

$\frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{D^{\gamma}(e^{j\omega})}{|D^{1+\gamma}(e^{j\omega})|^2} \right)$  que pode ser expressa por

$$\begin{aligned} \frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{D^{\gamma}(e^{j\omega})}{|D^{1+\gamma}(e^{j\omega})|^2} \right) &= \frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{1}{|D(e^{j\omega})|^2} \cdot \frac{1}{D^{*\gamma}(e^{j\omega})} \right) = \\ &= \frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{1}{|D(e^{j\omega})|^2} \right) \cdot \frac{1}{D^{*\gamma}(e^{j\omega})} + \frac{1}{|D(e^{j\omega})|^2} \cdot \frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{1}{D^{*\gamma}(e^{j\omega})} \right) \end{aligned} \quad (D.19)$$

onde

$$\frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{1}{D^{*\gamma}(e^{j\omega})} \right) = - \frac{\gamma \cdot \Phi_m^*(e^{j\omega})}{D^{*2\gamma}(e^{j\omega})} \quad (D.20)$$

Substituindo (D.11) (com  $l = m$ ) e (D.20) em (D.19), obtém-se

$$\begin{aligned} \frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{D^{\gamma}(e^{j\omega})}{|D^{1+\gamma}(e^{j\omega})|^2} \right) &= \\ &= - \frac{1}{|D^{1+\gamma}(e^{j\omega})|^2} \left( D^{\gamma}(e^{j\omega}) \Phi_m^*(e^{j\omega}) + D^{*\gamma}(e^{j\omega}) \Phi_m(e^{j\omega}) \right) \frac{1}{D^{*\gamma}(e^{j\omega})} \\ &\quad - \frac{1}{|D(e^{j\omega})|^2} \cdot \frac{\gamma \cdot \Phi_m^*(e^{j\omega})}{D^{*2\gamma}(e^{j\omega})} \end{aligned} \quad (D.21)$$

Desenvolvendo (D.21) e rearranjando seus termos, obtém-se

$$\frac{\partial}{\partial b_{\gamma'}(m)} \left( \frac{D^{\gamma}(e^{j\omega})}{|D^{1+\gamma}(e^{j\omega})|^2} \right) = - \frac{\Phi_m(e^{j\omega})}{|D^{1+\gamma}(e^{j\omega})|^2} - (1 + \gamma) \frac{D^{2\gamma}(e^{j\omega}) \Phi_m^*(e^{j\omega})}{|D^{1+2\gamma}(e^{j\omega})|^2} \quad (D.22)$$

que substituída em (D.18) fornece

$$\begin{aligned} \frac{\partial^2 \varepsilon}{\partial b_{\gamma'}(m) \partial b_{\gamma'}(l)} &= \\ &= 2. \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_m(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \right. \\ &\quad \left. + (1 + \gamma) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_m^*(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \right] \end{aligned} \quad (D.23)$$

Portanto

$$\frac{\partial^2 \varepsilon}{\partial b_{\gamma'}(m) \partial b_{\gamma'}(l)} = 2. [\tilde{p}_{\gamma}(m, l) + (1 + \gamma) \tilde{q}_{\gamma}(m, l)] \quad (D.24)$$

onde

$$\tilde{p}_\gamma(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_m(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \quad (\text{D.25})$$

e

$$\tilde{q}_\gamma(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_m^*(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \quad (\text{D.26})$$

Substituindo (B.13) e (B.14) em (D.25) e (D.26), respectivamente, obtém-se<sup>§</sup>

$$\tilde{p}_\gamma(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_1(e^{j\omega}) \cdot \Phi_{|l-m|+1}^*(e^{j\omega}) d\omega, \quad (\text{D.27})$$

$$0 \leq |l - m| \leq M - 1$$

e

$$\tilde{q}_\gamma(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_1^*(e^{j\omega}) \cdot \Phi_{l+m-1}^*(e^{j\omega}) d\omega, \quad (\text{D.28})$$

$$2 \leq l + m \leq 2M$$

onde se observa que  $\tilde{p}_\gamma(l, m)$  é um elemento de matriz do tipo  $p_{|l-m|}$  e, portanto, compõe uma matriz de Toeplitz simétrica. Já,  $\tilde{q}_\gamma(m, l)$  é um elemento do tipo  $q_{l+m}$ , constituindo então, uma matriz de Hankel. Representando  $\tilde{p}_\gamma(m, l)$  e  $\tilde{q}_\gamma(m, l)$ , respectivamente, por  $\tilde{p}_\gamma(l - m)$  e  $\tilde{q}_\gamma(l + m)$ , com base em (D.24), a matriz Hessiana é dada por

$$\mathbf{H} = 2 \cdot [\mathbf{P} + (1 + \gamma)\mathbf{Q}] \quad (\text{D.29})$$

onde

$$\mathbf{P} = \begin{bmatrix} \tilde{p}_\gamma(0) & \tilde{p}_\gamma(1) & \cdots & \tilde{p}_\gamma(M-1) \\ \tilde{p}_\gamma(1) & \tilde{p}_\gamma(0) & \cdots & \tilde{p}_\gamma(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_\gamma(M-1) & \tilde{p}_\gamma(M-2) & \cdots & \tilde{p}_\gamma(0) \end{bmatrix} \quad (\text{D.30})$$

e

---

<sup>§</sup> O produto  $\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega})$  dado pela relação (B.13) leva a expressões simétricas em relação a  $l - m = 0$ , distintas nos intervalos  $1 \leq l - m \leq M - 1$  e  $-M + 1 \leq l - m \leq -1$  somente por serem conjugadas entre si. A expressão (D.25), porém, é equivalente mediante a conjugação de  $\Phi_m(e^{j\omega})\Phi_l^*(e^{j\omega})$ , já que a integral e o fator  $\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2}$  do integrando são reais. (D.25) tem, portanto, simetria par, o que se manifesta em (D.27) em expressão única associada ao intervalo  $0 \leq |l - m| \leq M - 1$ .

$$\mathbf{Q} = \begin{bmatrix} \tilde{q}_\gamma(2) & \tilde{q}_\gamma(3) & \cdots & \tilde{q}_\gamma(M+1) \\ \tilde{q}_\gamma(3) & \tilde{q}_\gamma(4) & \cdots & \tilde{q}_\gamma(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}_\gamma(M+1) & \tilde{q}_\gamma(M+2) & \cdots & \tilde{q}_\gamma(2M) \end{bmatrix} \quad (\text{D.31})$$

com

$$\tilde{p}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_1(e^{j\omega}) \cdot \Phi_{k+1}^*(e^{j\omega}) d\omega, \quad (\text{D.32})$$

$$k = 0, 1, \dots, M-1$$

e

$$\tilde{q}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_1^*(e^{j\omega}) \cdot \Phi_{k-1}^*(e^{j\omega}) d\omega, \quad (\text{D.33})$$

$$k = 2, 3, \dots, 2M$$

## D.2 Análise Mel-Cepstral ( $\alpha \neq 0, \gamma = 0$ )

### D.2.1 Vetor Gradiente $\nabla_{\mathbf{b}'} \varepsilon$

Para o caso mel-cepstral, no vetor gradiente, dado por

$$\nabla_{\mathbf{b}'} \varepsilon = -2 \cdot \mathbf{r} \quad (\text{expressão (D.14)})$$

assume-se  $\gamma = 0$ , assim

$$\mathbf{r} = [\tilde{r}_0(1) \quad \tilde{r}_0(2) \quad \tilde{r}_0(3) \quad \dots \quad \tilde{r}_0(M)]^T \quad (\text{D.34})$$

com

$$\tilde{r}_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_k^*(e^{j\omega}) d\omega, \quad k = 1, 2, \dots, M \quad (\text{D.35})$$

expressão obtida substituindo  $\gamma = 0$  em (D.16).

Com  $k = 0$ , a expressão (D.35), reduz-se à expressão (D.1), fornecendo  $\tilde{r}_0(0) = \varepsilon$ , necessário a cada iteração do método de Newton Raphson.  $\tilde{r}_0(0)$  também é utilizado no cálculo da matriz Hessiana, abordado a seguir.

## D.2.2 Matriz Hessiana $H$

Assumindo  $\gamma = 0$  em (D.24), o elemento  $\frac{\partial^2 \varepsilon}{\partial b_0(m) \partial b_0(l)}$  da linha  $m$  e da coluna  $l$  da matriz Hessiana  $H$  é dado por

$$\frac{\partial^2 \varepsilon}{\partial b_0(m) \partial b_0(l)} = 2 \cdot [\tilde{p}_0(m, l) + \tilde{q}_0(m, l)] \quad (\text{D.36})$$

com

$$\tilde{p}_0(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_m(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \quad (\text{D.37})$$

e

$$\tilde{q}_0(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_m^*(e^{j\omega}) \Phi_l(e^{j\omega}) d\omega \quad (\text{D.38})$$

expressões obtidas substituindo  $\gamma = 0$  em (D.25) e (D.26). Substituindo (B.15) e (B.16) em (D.37) e (D.38), respectivamente, obtém-se<sup>§</sup>

$$\tilde{p}_0(m, l) = \begin{cases} (1 - \alpha^2) \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \right) + 2\alpha \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_1^*(e^{j\omega}) d\omega \right), & l - m = 0 \\ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_{|l-m|}^*(e^{j\omega}) d\omega + \alpha \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_{|l-m|+1}^*(e^{j\omega}) d\omega \right), & 1 \leq |l - m| \leq M - 1 \end{cases} \quad (\text{D.39})$$

e

$$\tilde{q}_0(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_{l+m}^*(e^{j\omega}) d\omega + \alpha \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_{l+m-1}^*(e^{j\omega}) d\omega \right) \quad (\text{D.40})$$

$$2 \leq l + m \leq 2M$$

Comparando tais expressões com (D.35), conclui-se que

$$\tilde{p}_0(m, l) = \begin{cases} (1 - \alpha^2) \tilde{r}_0(0) + 2\alpha \tilde{r}_0(1), & l - m = 0 \\ \tilde{r}_0(|l - m|) + \alpha \tilde{r}_0(|l - m| + 1), & 1 \leq |l - m| \leq M - 1 \end{cases} \quad (\text{D.41})$$

e

---

<sup>§</sup> O produto  $\Phi_m(e^{j\omega}) \Phi_l^*(e^{j\omega})$  dado pela relação (B.15) leva a expressões simétricas em relação a  $l - m = 0$ , distintas nos intervalos  $1 \leq l - m \leq M - 1$  e  $-M + 1 \leq l - m \leq -1$  somente por serem conjugadas entre si. A expressão (D.37), porém, é equivalente mediante a conjugação de  $\Phi_m(e^{j\omega}) \Phi_l^*(e^{j\omega})$ , já que a integral e o fator  $\frac{I_N(\omega)}{|D(e^{j\omega})|^2}$  do integrando são reais. (D.37) tem, portanto, simetria par, o que se manifesta em (D.39) em expressão única associada ao intervalo  $1 \leq |l - m| \leq M - 1$ .



$$\tilde{q}_0(m, l) = \tilde{r}_0(l + m) + \alpha \tilde{r}_0(l + m - 1), \quad 2 \leq l + m \leq 2M \quad (\text{D.42})$$

$\tilde{p}_0(m, l)$  é um elemento do tipo  $p_{|l-m|}$  e, portanto, compõe uma matriz de Toeplitz simétrica. Já,  $\tilde{q}_0(m, l)$  é do tipo  $q_{l+m}$ , constituindo então, uma matriz de Hankel. Representando  $\tilde{p}_0(m, l)$  e  $\tilde{q}_0(m, l)$ , respectivamente, por  $\tilde{p}_0(l - m)$  e  $\tilde{q}_0(l + m)$ , com base em (D.36), a matriz Hessiana é dada por

$$\mathbf{H} = 2. [\mathbf{P} + \mathbf{Q}] \quad (\text{D.43})$$

onde

$$\mathbf{P} = \begin{bmatrix} \tilde{p}_0(0) & \tilde{p}_0(1) & \cdots & \tilde{p}_0(M-1) \\ \tilde{p}_0(1) & \tilde{p}_0(0) & \cdots & \tilde{p}_0(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}_0(M-1) & \tilde{p}_0(M-2) & \cdots & \tilde{p}_0(0) \end{bmatrix} \quad (\text{D.44})$$

e

$$\mathbf{Q} = \begin{bmatrix} \tilde{q}_0(2) & \tilde{q}_0(3) & \cdots & \tilde{q}_0(M+1) \\ \tilde{q}_0(3) & \tilde{q}_0(4) & \cdots & \tilde{q}_0(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}_0(M+1) & \tilde{q}_0(M+2) & \cdots & \tilde{q}_0(2M) \end{bmatrix} \quad (\text{D.45})$$

com

$$\tilde{p}_0(k) = \begin{cases} (1 - \alpha^2)\tilde{r}_0(0) + 2\alpha\tilde{r}_0(1), & k = 0 \\ \tilde{r}_0(k) + \alpha\tilde{r}_0(k+1), & k = 1, 2, \dots, M-1 \end{cases} \quad (\text{D.46})$$

e

$$\tilde{q}_0(k) = \tilde{r}_0(k) + \alpha\tilde{r}_0(k-1), \quad k = 2, 3, \dots, 2M \quad (\text{D.47})$$

onde

$$\tilde{r}_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_k^*(e^{j\omega}) d\omega \quad (\text{expressão (D.35)})$$

### D.3 Análise Cepstral Generalizada ( $\alpha = 0, \gamma \neq 0$ )

Para a análise cepstral generalizada, assume-se  $\alpha = 0$ . Nesse caso os coeficientes  $b_\gamma'(k)$  se reduzem aos coeficientes  $c_\gamma'(k)$  e

$$\Phi_k(e^{j\omega}) = e^{-j\omega k} \quad \text{para } \alpha = 0 \quad (\text{D.48})$$

### D.3.1 Vetor Gradiente $\nabla_{\mathbf{b}'} \varepsilon$

Para o caso cepstral generalizado, no vetor gradiente, dado por

$$\nabla_{\mathbf{b}'} \varepsilon = -2 \cdot \mathbf{r} \quad (\text{expressão (D.14)})$$

assume-se  $\alpha = 0$ , assim

$$\mathbf{r} = [r_\gamma(1) \quad r_\gamma(2) \quad r_\gamma(3) \quad \dots \quad r_\gamma(M)]^T \quad (\text{D.49})$$

com

$$r_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) e^{j\omega k} d\omega \quad k = 1, 2, \dots, M \quad (\text{D.50})$$

expressão obtida substituindo  $\alpha = 0$  em (D.16) (já que (D.48) é válido).

Assumindo  $k = 0$  em (D.50), obtém-se

$$r_\gamma(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) d\omega \quad (\text{D.51})$$

que, em conjunto com  $\mathbf{r}$ , é utilizada no cálculo de  $\varepsilon$ , a cada iteração do método de Newton Raphson. Tal cálculo é efetuado de forma semelhante àquele abordado no Apêndice E (seção E.1.4) no contexto da análise Mel-Cepstral Generalizada ( $\alpha \neq 0, \gamma \neq 0$ ).

### D.3.2 Matriz Hessiana $H$

Adotando-se  $\alpha = 0$  em (D.24), (D.25) e (D.26), obtém-se

$$\frac{\partial^2 \varepsilon}{\partial c_\gamma'(m) \partial c_\gamma'(l)} = 2 \cdot [p_\gamma(m, l) + (1 + \gamma)q_\gamma(m, l)] \quad (\text{D.52})$$

onde

$$p_\gamma(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} e^{j\omega(l-m)} d\omega \quad (\text{D.53})$$

e

$$q_\gamma(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) e^{j\omega(l+m)} d\omega \quad (\text{D.54})$$

$p_\gamma(m, l)$  é um elemento do tipo  $p_{|l-m|}$  e, portanto, compõe uma matriz de Toeplitz simétrica. Já,  $q_\gamma(m, l)$  é do tipo  $q_{l+m}$ , constituindo então, uma matriz de Hankel. Representando  $p_\gamma(m, l)$  e  $q_\gamma(m, l)$ , respectivamente, por  $p_\gamma(l - m)$  e  $q_\gamma(l + m)$ , com base em (D.52), a matriz Hessiana é dada por

$$\mathbf{H} = 2. [\mathbf{P} + (1 + \gamma)\mathbf{Q}] \quad (\text{D.55})$$

onde

$$\mathbf{P} = \begin{bmatrix} p_\gamma(0) & p_\gamma(1) & \cdots & p_\gamma(M-1) \\ p_\gamma(1) & p_\gamma(0) & \cdots & p_\gamma(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ p_\gamma(M-1) & p_\gamma(M-2) & \cdots & p_\gamma(0) \end{bmatrix} \quad (\text{D.56})$$

e

$$\mathbf{Q} = \begin{bmatrix} q_\gamma(2) & q_\gamma(3) & \cdots & q_\gamma(M+1) \\ q_\gamma(3) & q_\gamma(4) & \cdots & q_\gamma(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ q_\gamma(M+1) & q_\gamma(M+2) & \cdots & q_\gamma(2M) \end{bmatrix} \quad (\text{D.57})$$

com

$$p_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 0, 1, \dots, M-1 \quad (\text{D.58})$$

e

$$q_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) e^{j\omega k} d\omega \quad k = 2, 3, \dots, 2M \quad (\text{D.59})$$

## D.4 Análise Cepstral ( $\alpha = 0, \gamma = 0$ )

### D.4.1 Vetor Gradiente $\nabla_{\mathbf{b}'} \varepsilon$

Para o caso cepstral, no vetor gradiente, dado por

$$\nabla_{\mathbf{b}'} \varepsilon = -2. \mathbf{r} \quad (\text{expressão (D.14)})$$

assume-se  $\gamma = 0$  e  $\alpha = 0$  e assim

$$\mathbf{r} = [r_0(1) \quad r_0(2) \quad r_0(3) \quad \dots \quad r_0(M)]^T \quad (\text{D.60})$$

com

$$r_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 1, 2, \dots, M \quad (\text{D.61})$$

expressão obtida substituindo  $\gamma = 0$  e  $\alpha = 0$  em (D.16) (considerando a igualdade (D.48)).

Com  $k = 0$ , a expressão (D.61) reduz-se à expressão (D.1), fornecendo  $r_0(0) = \varepsilon$ , necessário a cada iteração do método de Newton Raphson. A componente  $r_0(0)$  também é utilizado no cálculo da matriz Hessiana, abordado a seguir

#### D.4.2 Matriz Hessiana $H$

Assumindo  $\gamma = 0$  e  $\alpha = 0$  em (D.24), o elemento  $\frac{\partial^2 \varepsilon}{\partial c_0(m) \partial c_0(l)}$  da linha  $m$  e da coluna  $l$  da matriz Hessiana  $H$  é dado por

$$\frac{\partial^2 \varepsilon}{\partial c_0(m) \partial c_0(l)} = 2 \cdot [p_0(m, l) + q_0(m, l)] \quad (\text{D.62})$$

Substituindo  $\gamma = 0$  e  $\alpha = 0$  nas expressões (D.25) e (D.26) obtém-se

$$p_0(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} e^{j\omega|l-m|} d\omega \quad 0 \leq |l-m| \leq M-1 \quad (\text{D.63})$$

e

$$q_0(m, l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} e^{j\omega(l+m)} d\omega, \quad 2 \leq l+m \leq 2M \quad (\text{D.64})$$

Comparando tais expressões com (D.61), conclui-se que

$$p_0(m, l) = r_0(|l-m|), \quad 0 \leq |l-m| \leq M-1 \quad (\text{D.65})$$

e

$$q_0(m, l) = r_0(l+m), \quad 2 \leq l+m \leq 2M \quad (\text{D.66})$$

$p_0(m, l)$  é um elemento do tipo  $p_{|l-m|}$  e, portanto, compõe uma matriz de Toeplitz simétrica. Já,  $q_0(m, l)$  é do tipo  $q_{l+m}$ , constituindo então, uma matriz de Hankel.

Representando  $p_0(m, l)$  e  $q_0(m, l)$ , respectivamente, por  $p_0(l - m)$  e  $q_0(l + m)$ , com base em (D.62), a matriz Hessiana é dada por

$$\mathbf{H} = 2 \cdot [\mathbf{P} + \mathbf{Q}] \quad (\text{D.67})$$

onde

$$\mathbf{P} = \begin{bmatrix} p_0(0) & p_0(1) & \cdots & p_0(M-1) \\ p_0(1) & p_0(0) & \cdots & p_0(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ p_0(M-1) & p_0(M-2) & \cdots & p_0(0) \end{bmatrix} \quad (\text{D.68})$$

e

$$\mathbf{Q} = \begin{bmatrix} q_0(2) & q_0(3) & \cdots & q_0(M+1) \\ q_0(3) & \tilde{q}_0(4) & \cdots & q_0(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ q_0(M+1) & q_0(M+2) & \cdots & q_0(2M) \end{bmatrix} \quad (\text{D.69})$$

com

$$p_0(k) = r_0(k), \quad k = 0, 1, \dots, M-1 \quad (\text{D.70})$$

e

$$q_0(k) = r_0(k), \quad k = 2, 3, \dots, 2M \quad (\text{D.71})$$

onde

$$r_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} e^{j\omega k} d\omega \quad (\text{expressão (D.61)})$$

### Observações relacionadas à computação de $\nabla_{\mathbf{b}} \cdot \boldsymbol{\varepsilon}$ e $\mathbf{H}$

Quando  $\alpha = 0$ , ou seja, para as análises cepstral e cepstral generalizada, as componentes de  $\nabla_{\mathbf{b}} \cdot \boldsymbol{\varepsilon}$  e os elementos de  $\mathbf{H}$  podem ser calculados via FFT. Isso é possível, uma vez que  $r_\gamma(k)$ ,  $p_\gamma(k)$  e  $q_\gamma(k)$  (caso cepstral generalizado) e  $r_0(k)$  (caso cepstral) correspondem a transformadas inversas de Fourier. Isto não é verdade para as formas aqui apresentadas para  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$ ,  $\tilde{q}_\gamma(k)$  (caso mel-cepstral generalizado) e  $\tilde{r}_0(k)$  (caso mel-cepstral). A computação das integrais para tais elementos é tratada pelos métodos descritos no Apêndice E e no Apêndice F.

## Apêndice E

# Computação do Gradiente e da Hessiana de $\varepsilon$ via Fórmulas Recursivas ( $\alpha \neq 0$ )

Este Apêndice toma como base o Apêndice D para as deduções das fórmulas associadas aos métodos de [2] e [7] para o cálculo computacional do vetor  $\nabla_{b'}\varepsilon$  e da matriz Hessiana  $\mathbf{H}$  do critério  $\varepsilon$  para os casos MGC e Mel-Cepstrais  $\alpha \neq 0$ . Os métodos consistem no cálculo dos elementos  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$ ,  $\tilde{q}_\gamma(k)$  (caso Mel-Cepstral Generalizado) e  $\tilde{r}_0(k)$  (caso Mel-Cepstral) por meio de fórmulas recursivas. Uma fórmula para o cálculo econômico do critério  $\varepsilon$ , necessário a cada iteração do método de Newton Raphson, também é deduzida.

Será útil que  $\Phi_n(z)$  seja explicitado na forma

$$\Phi_n(z) = \begin{cases} 1 & n = 0 \\ \Phi_1(z)[\Psi(z)]^{n-1} & n > 0 \end{cases} \quad (\text{E.1})$$

com

$$\Phi_1(z) = \frac{(1 - \alpha^2)z^{-1}}{1 - \alpha z^{-1}} \quad (\text{E.2})$$

e

$$\Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (\text{E.3})$$

A expressão

$$\Lambda(z) = \frac{1}{1 - \alpha z^{-1}} \quad (\text{E.4})$$

também se mostrará útil.

## E.1 Análise Mel-Cepstral Generalizada

Para o cálculo de  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$ , são consideradas integrais com formas equivalentes àquelas dadas pelas expressões (D.50), (D.58) e (D.59) (utilizadas no cálculo do vetor gradiente  $\nabla_{\mathbf{b}'} \varepsilon$  e da Hessiana  $\mathbf{H}$  do caso cepstral-generalizado ( $\alpha = 0, \gamma \neq 0$ )), dadas por

$$r_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}^{1+\gamma}(e^{j\omega})|^2} \bar{D}^\gamma(e^{j\omega}) e^{j\omega k} d\omega, \quad k = 0, 1, \dots, N' \quad (\text{E.5})$$

$$p_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}^{1+\gamma}(e^{j\omega})|^2} e^{j\omega k} d\omega, \quad k = 0, 1, \dots, N' \quad (\text{E.6})$$

$$q_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}^{1+2\gamma}(e^{j\omega})|^2} \bar{D}^{2\gamma}(e^{j\omega}) e^{j\omega k} d\omega, \quad k = 1, 2, \dots, N' \quad (\text{E.7})$$

com

$$\bar{D}(e^{j\omega}) = \left( 1 + \gamma \sum_{n=1}^{N'} c'_\gamma(n) e^{-j\omega n} \right)^{1/\gamma} \quad (\text{E.8})$$

onde os  $N'$  coeficientes  $c'_\gamma(n)$  correspondem ao cepstrum generalizado associado aos  $M$  coeficientes MGC transformados  $b'_\gamma(n)$  (calculados na etapa anterior do método de Newton-Raphson), ou seja,  $\bar{D}(e^{j\omega})$  é tal que

$$\bar{D}(e^{j\omega}) = D(e^{j\omega}) \quad (\text{E.9})$$

onde  $D(e^{j\omega})$  é o modelo MGC expresso por

$$D(e^{j\omega}) = \left( 1 + \gamma \sum_{n=1}^M b'_\gamma(n) \Phi_n(e^{j\omega}) \right)^{1/\gamma} \quad (\text{E.10})$$

O propósito de se expressar  $D(e^{j\omega})$  através de  $\bar{D}(e^{j\omega})$  é que  $\bar{D}^\gamma(e^{j\omega})$  pode ser calculada via FFT. Uma especificação completa do modelo MGC  $D(e^{j\omega})$  através de  $\bar{D}(e^{j\omega})$  exigiria, idealmente,  $N' = \infty$  amostras do cepstrum generalizado  $c'_\gamma(n)$ . É suficiente, porém, que se adote um valor grande  $N' \gg 2M$ , sob o teto de  $N' = N_F - 1$  ( $N_F$  é o comprimento do periodograma amostrado). Aconselha-se a adoção de, no mínimo,  $N' \approx \frac{1+|\alpha|}{1-|\alpha|}(2M)$ <sup>§</sup>.

As expressões dos elementos  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  do gradiente e da Hessiana de  $\varepsilon$  para o caso mel-cepstral generalizado ( $\alpha \neq 0$ ) são dadas por

$$\tilde{r}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) \Phi_k^*(e^{j\omega}) d\omega, \quad k = 1, 2, \dots, M \quad (\text{expressão (D.16)})$$

$$\tilde{p}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_1(e^{j\omega}) \cdot \Phi_{k+1}^*(e^{j\omega}) d\omega, \quad k = 0, 1, \dots, M-1 \quad (\text{expressão (D.32)})$$

e

$$\tilde{q}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_1^*(e^{j\omega}) \cdot \Phi_{k-1}(e^{j\omega}) d\omega, \quad k = 2, 3, \dots, 2M \quad (\text{expressão (D.33)})$$

Como tais integrais produzem resultados reais, pode-se conjugá-las. Conjugando-as e considerando que elas correspondem ao caso  $i = 0$  das seguintes transformadas inversas de Fourier

$$\tilde{r}_\gamma(k, i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{*\gamma}(e^{j\omega}) \Phi_k(e^{j\omega}) e^{j\omega i} d\omega, \quad k = 1, 2, \dots, M \quad (\text{E.11})$$

---

<sup>§</sup> Como é assumido um único valor de  $N'$  para todas as transformações em frequência aqui efetuadas, adota-se o mínimo de  $N' + 1 \approx \frac{1+|\alpha|}{1-|\alpha|}(2M + 1)$  amostras, com base na aplicação do critério de suavização da seção 3.4.3 à transformação (E.33), que gera o maior número de amostras dentre todas as transformações aqui envolvidas ( $2M + 1$  amostras) e conta, portanto, com o maior número  $N' + 1$  de amostras de entrada.



$$\tilde{p}_\gamma(k, i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_1^*(e^{j\omega}) \cdot \Phi_{k+1}(e^{j\omega}) e^{j\omega i} d\omega, \quad k = 0, 1, \dots, M-1 \quad (\text{E.12})$$

$$\tilde{q}_\gamma(k, i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{*2\gamma}(e^{j\omega}) \Phi_1(e^{j\omega}) \cdot \Phi_{k-1}(e^{j\omega}) e^{j\omega i} d\omega, \quad k = 2, 3, \dots, 2M \quad (\text{E.13})$$

temos

$$\tilde{r}_\gamma(k, 0) = \tilde{r}_\gamma(k), \quad \tilde{p}_\gamma(k, 0) = \tilde{p}_\gamma(k) \quad \text{e} \quad \tilde{q}_\gamma(k, 0) = \tilde{q}_\gamma(k) \quad (\text{E.14})$$

Comparando (E.5), (E.6) e (E.7) com (E.11), (E.12) e (E.13), respectivamente, concluímos que  $\tilde{r}_\gamma(k, i)$  é a saída do filtro  $\Phi_k(z)$ , com entrada  $r_\gamma(-i)$ . Já,  $\tilde{p}_\gamma(k, i)$  é a saída do filtro  $\Phi_1^*(z)\Phi_{k+1}(z)$  com entrada  $p_\gamma(-i)$  e  $\tilde{q}_\gamma(k, i)$ , por sua vez, é a saída do filtro  $\Phi_1(z)\Phi_{k-1}(z)$  cuja entrada é  $q_\gamma(-i)$ . Tais casos serão tratados a seguir. Para tal, deve-se levar em conta que o intervalo de  $k$  considerado em (E.5), (E.6) e (E.7) vai além daqueles dos utilizados em  $r_\gamma(k)$ ,  $p_\gamma(k)$  e  $q_\gamma(k)$  como elementos de  $\mathbf{V}_{b'}\boldsymbol{\varepsilon}$  e  $\mathbf{H}$  para a análise cepstral generalizada ( $\alpha = 0$ ) de ordem  $M$ . Para a análise mel-cepstral generalizada de ordem  $M$ , o cálculo das componentes  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$ , que têm os comprimentos dados em (D.16), (D.32) e (D.33) (expressões replicadas no presente apêndice), conta com  $N' + 1$  componentes de cada uma das seqüências  $r_\gamma(k)$ ,  $p_\gamma(k)$  e  $q_\gamma(k)$ . As expressões (E.5), (E.6) e (E.7) para tais seqüências, assim como  $\bar{D}^\gamma(e^{j\omega})$  (com  $\bar{D}(e^{j\omega})$  expresso por (E.8)) são calculadas via FFT.

### E.1.1 Cálculo de $\tilde{r}_\gamma(k)$

Considerando  $\Phi_n(e^{j\omega})$  na forma (E.1), a expressão (E.11) é contemplada pela filtragem em cascata representada na Figura E.1.

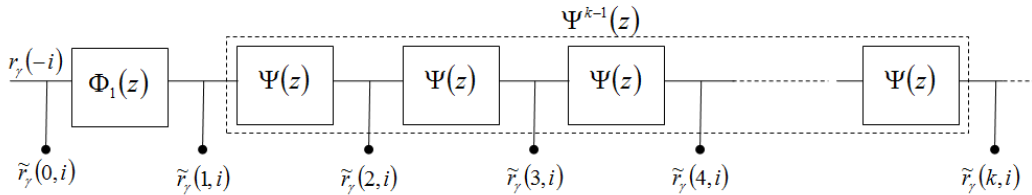


Figura E.1: Banco de filtros em cascata com entrada  $r_\gamma(-i)$  e saída  $\tilde{r}_\gamma(k, i)$  em sua  $k$ -ésima seção. Os valores  $\tilde{r}_\gamma(k, 0)$  das seqüências de saída constituem a seqüência  $\tilde{r}_\gamma(k)$ .

A filtragem da Figura E.1 pode ser computada pela fórmula recursiva

$$\tilde{r}_\gamma(k, i) = \begin{cases} r_\gamma(-i) & k = 0 \\ (1 - \alpha^2)\tilde{r}_\gamma(0, i - 1) + \alpha\tilde{r}_\gamma(1, i - 1), & k = 1 \\ \tilde{r}_\gamma(k - 1, i - 1) + \alpha(\tilde{r}_\gamma(k, i - 1) - \tilde{r}_\gamma(k - 1, i)), & k = 2, 3, \dots, M \end{cases} \quad (\text{E.15})$$

$$i = -N', \dots, -2, -1, 0$$

$$\tilde{r}_\gamma(k) = \tilde{r}_\gamma(k, 0)$$

que pode ser deduzida com base em (E.2) e (E.3)<sup>§</sup>. Assumindo  $\mathbf{r}_{\gamma, N'} = [r_\gamma(0) \ r_\gamma(1) \ \dots \ r_\gamma(N')]^T$  e  $\tilde{\mathbf{r}}_{\gamma, M} = [\tilde{r}_\gamma(0) \ \tilde{r}_\gamma(1) \ \dots \ \tilde{r}_\gamma(M)]^T$ , a transformação que leva  $\mathbf{r}_{\gamma, N'}$  em  $\tilde{\mathbf{r}}_{\gamma, M}$  através de (E.15) será representada por

$$\tilde{\mathbf{r}}_{\gamma, M} = T_1(\mathbf{r}_{\gamma, N'}, N', M, \alpha) \quad (\text{E.16})$$

A partir de  $\tilde{\mathbf{r}}_{\gamma, M}$  é possível o cálculo de  $\varepsilon$ , cuja taxa de variação fornece um critério de parada para as iterações do método de Newton-Raphson. Tal cálculo é abordado na seção E.1.4.

### E.1.2 Cálculo de $\tilde{p}_\gamma(k)$

Na integral (E.12), o fator  $\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2}$  é real e par, de maneira que, se não se impuserem limitações aos índices, a seqüência no tempo associada a tal fator é dada por  $p_\gamma(k)$  real e não-causal com simetria par (ver (E.6)). Se dividirmos, porém,  $p_\gamma(k)$  em uma parte causal  $p'_\gamma(k)$  e outra não-causal  $p'_\gamma(-k)$ , de maneira que  $p_\gamma(k) = p'_\gamma(k) + p'_\gamma(-k)$ , temos

$$p'_\gamma(k) = \begin{cases} \frac{1}{2}p_\gamma(0), & k = 0 \\ p_\gamma(k), & k = 1, 2, \dots, N' \end{cases} \quad (\text{E.17})$$

Considerando que  $P'_\gamma(e^{j\omega}) \leftrightarrow p'_\gamma(k)$ , no lugar de se representar  $\tilde{p}_\gamma(k)$  por (D.32), pode-se representá-la por

<sup>§</sup> A dedução é feita nos mesmos moldes daquela para a fórmula (3.29) na seção 3.4.1.

$$\tilde{p}'_{\gamma}(k) = \begin{cases} 2 \cdot \tilde{p}'_{\gamma}(0), & k = 0 \\ \tilde{p}'_{\gamma}(k), & k = 1, 2, \dots, M-1 \end{cases} \quad (\text{E.18})$$

onde

$$\tilde{p}'_{\gamma}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_{\gamma}(e^{j\omega}) \Phi_1(e^{j\omega}) \cdot \Phi_{k+1}^*(e^{j\omega}) d\omega, \quad k = 0, 1, \dots, M-1 \quad (\text{E.19})$$

o que é demonstrado no final deste tópic. Uma vez que  $p'_{\gamma}(k)$  é causal,  $\tilde{p}'_{\gamma}(k)$  pode ser obtida por

$$\tilde{p}'_{\gamma}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_{\gamma}(e^{j\omega}) (1 - \alpha^2) \Lambda^*(e^{j\omega}) \Phi_k^*(e^{j\omega}) d\omega, \quad k = 0, 1, \dots, M-1 \quad (\text{E.20})$$

onde  $\Lambda(e^{j\omega})$  é dada por (E.4). A validade da expressão (E.20) também é demonstrada no final do tópic. Como  $\tilde{p}'_{\gamma}(k)$  é real, (E.20) pode ser conjugada. Uma vez conjugada, ela corresponde ao caso  $i = 0$  da seguinte transformada inversa de Fourier

$$\tilde{p}'_{\gamma}(k, i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_{\gamma}{}^*(e^{j\omega}) (1 - \alpha^2) \Lambda(e^{j\omega}) \Phi_k(e^{j\omega}) e^{j\omega i} d\omega, \quad k = 0, 1, \dots, M-1 \quad (\text{E.21})$$

A expressão (E.21) é contemplada pela filtragem em cascata representada na Figura E.2.

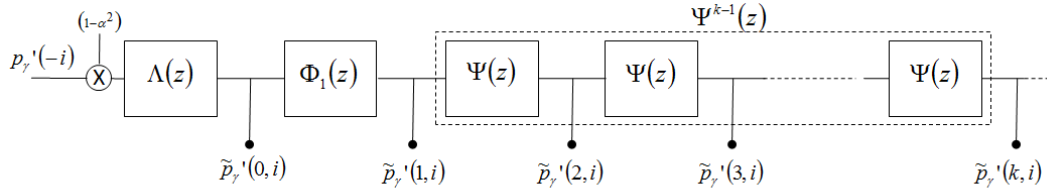


Figura E.2: Banco de filtros em cascata com entrada  $p'_{\gamma}(-i)$  e saída  $\tilde{p}'_{\gamma}(k, i)$  em sua  $(k+1)$ -ésima seção. Os valores  $\tilde{p}'_{\gamma}(k, 0)$  das seqüências de saída constituem a seqüência  $\tilde{p}'_{\gamma}(k)$ .

A filtragem da Figura E.2 pode ser computada pela fórmula recursiva

$$\tilde{p}'_{\gamma}(k, i) = \begin{cases} (1 - \alpha^2)p'_{\gamma}(-i) + \alpha\tilde{p}'_{\gamma}(0, i - 1), & k = 0 \\ (1 - \alpha^2)\tilde{p}'_{\gamma}(0, i - 1) + \alpha\tilde{p}'_{\gamma}(1, i - 1), & k = 1 \\ \tilde{p}'_{\gamma}(k - 1, i - 1) + \alpha(\tilde{p}'_{\gamma}(k, i - 1) - \tilde{p}'_{\gamma}(k - 1, i)), & k = 2, 3, \dots, M - 1 \end{cases} \quad (\text{E.22})$$

$$i = -N', \dots, -2, -1, 0$$

$$\tilde{p}'_{\gamma}(k) = \tilde{p}'_{\gamma}(k, 0)$$

onde a seqüência  $p'_{\gamma}(k)$  é dada por (E.17).  $\tilde{p}'_{\gamma}(k)$  é determinada a partir de  $\tilde{p}'_{\gamma}(k)$  por (E.18). A fórmula recursiva (E.22) pode ser deduzida com base em (E.2), (E.3) e (E.4)<sup>§</sup>. Assumindo  $\mathbf{p}'_{\gamma, N'} = [p'_{\gamma}(0) p'_{\gamma}(1) \dots p'_{\gamma}(N')]^T$  e  $\tilde{\mathbf{p}}'_{\gamma, M-1} = [\tilde{p}'_{\gamma}(0) \tilde{p}'_{\gamma}(1) \dots \tilde{p}'_{\gamma}(M-1)]^T$ , a transformação que leva  $\mathbf{p}'_{\gamma, N'}$  em  $\tilde{\mathbf{p}}'_{\gamma, M-1}$  através de (E.22) é dada por

$$\tilde{\mathbf{p}}'_{\gamma, M-1} = (1 - \alpha^2) \cdot T_{\Lambda}(\mathbf{p}'_{\gamma, N'}, N', M - 1, \alpha) \quad (\text{E.23})$$

onde  $T_{\Lambda}(\cdot)$  corresponde à transformação em freqüência cujas propriedades são abordadas na seção 3.4.3.

### Demonstração de (E.18)/(E.19)

Considere

$$\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} = P'_{\gamma}(e^{j\omega}) + P'_{\gamma*}(e^{j\omega}) \quad (\text{E.24})$$

onde o primeiro termo do lado direito da igualdade está associado à parte causal de  $\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2}$  e o segundo termo, à sua parte não-causal.

Com base em (D.32), temos, para  $k = 0$

$$\tilde{p}'_{\gamma}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} |\Phi_1(e^{j\omega})|^2 d\omega \quad (\text{E.25})$$

onde, substituindo (E.24), obtém-se

$$\tilde{p}'_{\gamma}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_{\gamma}(e^{j\omega}) |\Phi_1(e^{j\omega})|^2 d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_{\gamma*}(e^{j\omega}) |\Phi_1(e^{j\omega})|^2 d\omega \quad (\text{E.26})$$

como as integrais são reais e o fator  $|\Phi_1(e^{j\omega})|^2$  também é real, os termos podem ser conjugados sem que sejam alterados. Conjugando o segundo termo e efetuando a soma, temos

<sup>§</sup> A dedução é feita nos mesmos moldes daquela para a fórmula (3.29) na seção 3.4.1.

$$\tilde{p}_\gamma(0) = 2 \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_\gamma(e^{j\omega}) |\Phi_1(e^{j\omega})|^2 d\omega \right] \quad (\text{E.27})$$

o que demonstra (E.18) para  $k = 0$ .

Na expressão (D.32), dada por

$$\tilde{p}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Phi_1(e^{j\omega}) \Phi_{k+1}^*(e^{j\omega}) d\omega \quad (\text{expressão (D.32)})$$

o fator  $\Phi_1(e^{j\omega}) \Phi_{k+1}^*(e^{j\omega})$ , para  $k > 0$ , pode ser expresso por

$$\Phi_1(e^{j\omega}) \Phi_{k+1}^*(e^{j\omega}) = \Phi_k^*(e^{j\omega}) + \alpha \Phi_{k+1}^*(e^{j\omega}), \quad k > 1 \quad (\text{E.28})$$

(obtido substituindo  $m = 1$  e  $l = k + 1$  na expressão (B.15)) onde se constata que  $\Phi_1(e^{j\omega}) \Phi_{k+1}^*(e^{j\omega})$  é anti-causal, já que  $\Phi_k^*(e^{j\omega})$  e  $\Phi_{k+1}^*(e^{j\omega})$  são anti-causais. Assim, como  $\tilde{p}_\gamma(k)$  (expressão (D.32)) corresponde à amostra zero da convolução entre uma seqüência não-causal par (associada a  $\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2}$ ) e uma seqüência anti-causal, recebe contribuições somente da parte causal  $P'_\gamma(e^{j\omega})$  de  $\frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2}$  e pode, portanto, ser expressa por

$$\tilde{p}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_\gamma(e^{j\omega}) \Phi_1(e^{j\omega}) \Phi_{k+1}^*(e^{j\omega}) d\omega, \quad k > 0 \quad (\text{E.29})$$

o que demonstra (E.18) para  $k > 0$ .

### Demonstração de (E.20)

Substituindo (B.21), para  $k = 0$ , na expressão (E.19), obtém-se

$$\tilde{p}'_\gamma(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_\gamma(e^{j\omega}) \alpha \Phi_1(e^{j\omega}) d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_\gamma(e^{j\omega}) (1 - \alpha^2) \Lambda^*(e^{j\omega}) d\omega \quad (\text{E.30})$$

onde a primeira integral corresponde à amostra zero da convolução entre duas seqüências causais, sendo que aquela associada a  $\Phi_1(z)$  tem amostra zero nula (pela presença do atraso  $z^{-1}$ ). Tal integral é, portanto, nula e  $\tilde{p}'_\gamma(0)$  é dado por

$$\tilde{p}'_\gamma(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P'_\gamma(e^{j\omega}) (1 - \alpha^2) \Lambda^*(e^{j\omega}) d\omega \quad (\text{E.31})$$

que corresponde a (E.20) para  $k = 0$ . Chega-se a (E.20) para  $k > 0$ , substituindo a expressão (B.21) para  $k > 0$ , na expressão (E.19).

### E.1.3 Cálculo de $\tilde{q}_\gamma(k)$

A expressão (E.13) é contemplada pela filtragem em cascata representada na Figura E.3.

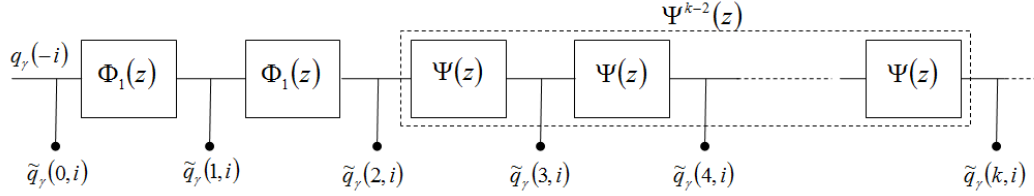


Figura E.3: Banco de filtros em cascata com entrada  $q_\gamma(-i)$  e saída  $\tilde{q}_\gamma(k, i)$  em sua  $k$ -ésima seção. Os valores  $\tilde{q}_\gamma(k, 0)$  das seqüências de saída constituem a seqüência  $\tilde{q}_\gamma(k)$ .

A filtragem da Figura E.3 pode ser computada pela fórmula recursiva

$$\tilde{q}_\gamma(k, i) = \begin{cases} q_\gamma(-i) & k = 0 \\ (1 - \alpha^2) \cdot \tilde{q}_\gamma(0, i - 1) + \alpha \tilde{q}_\gamma(1, i - 1), & k = 1 \\ (1 - \alpha^2) \cdot \tilde{q}_\gamma(1, i - 1) + \alpha \tilde{q}_\gamma(2, i - 1), & k = 2 \\ \tilde{q}_\gamma(k - 1, i - 1) + \alpha (\tilde{q}_\gamma(k, i - 1) - \tilde{q}_\gamma(k - 1, i)), & k = 3, 4, \dots, 2M \end{cases} \quad (\text{E.32})$$

$$i = -N', \dots, -2, -1, 0$$

$$\tilde{q}_\gamma(k) = \tilde{q}_\gamma(k, 0)$$

que pode ser deduzida com base em (E.2) e (E.3)<sup>§</sup>. Assumindo  $\mathbf{q}_{\gamma, N'} = [q_\gamma(0) \ q_\gamma(1) \ \dots \ q_\gamma(N')]^T$  e  $\tilde{\mathbf{q}}_{\gamma, 2M} = [\tilde{q}_\gamma(0) \ \tilde{q}_\gamma(1) \ \dots \ \tilde{q}_\gamma(2M)]^T$ , a transformação que leva  $\mathbf{q}_{\gamma, N'}$  em  $\tilde{\mathbf{q}}_{\gamma, 2M}$  através de (E.32) será representada por

$$\tilde{\mathbf{q}}_{\gamma, 2M} = T_{1, \Phi_1}(\mathbf{q}_{\gamma, N'}, N, 2M, \alpha) \quad (\text{E.33})$$

### E.1.4 Cálculo de $\varepsilon$

A cada iteração do método de Newton-Raphson, é necessário que se determine o valor de  $\varepsilon$  já que sua taxa de variação é utilizada como critério de parada na execução do método. Com o objetivo de que seja minimizado o custo para seu cálculo,  $\varepsilon$  pode ser

<sup>§</sup> A dedução é feita nos mesmos moldes daquela para a fórmula (3.29) na seção 3.4.1.

computado com base em  $\tilde{r}_\gamma(0)$ ,  $\mathbf{r} = [\tilde{r}_\gamma(1) \ \tilde{r}_\gamma(2) \ \dots \ \tilde{r}_\gamma(M)]^T$  e  $\mathbf{b}'_\gamma = [b'_\gamma(1) \ b'_\gamma(2) \ \dots \ b'_\gamma(M)]^T$  através de

$$\varepsilon = \tilde{r}_\gamma(0) + \gamma \mathbf{r}^T \cdot \mathbf{b}'_\gamma \quad (\text{E.34})$$

expressão cuja validade é demonstrada a seguir

### Demonstração de (E.34)

A expressão (E.34) pode ser representada por

$$\tilde{r}_\gamma(0) + \gamma \sum_{k=1}^M b'_\gamma(k) \tilde{r}_\gamma(k) \quad (\text{E.35})$$

onde explicitando-se  $\tilde{r}_\gamma(k)$  e  $\tilde{r}_\gamma(0)$  (dados por (D.16) e (D.17)) temos

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) d\omega \\ & + \gamma \sum_{k=1}^M b'_\gamma(k) \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) \Phi_k^*(e^{j\omega}) d\omega \right) \end{aligned} \quad (\text{E.36})$$

portanto

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^\gamma(e^{j\omega}) \left( 1 + \gamma \sum_{k=1}^M b'_\gamma(k) \Phi_k^*(e^{j\omega}) \right) d\omega \quad (\text{E.37})$$

O fator entre parênteses no integrando de (E.37) corresponde a  $(D^\gamma(e^{j\omega}))^*$ , de maneira que ela se reduz à expressão (D.1), o que confirma (E.34).

## E.2 Análise Mel-Cepstral ( $\gamma = 0$ )

Para o cálculo de  $\tilde{r}_0(k)$ ,  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$ , é considerada uma integral com forma equivalente àquelas dada pela expressão (utilizada no cálculo do vetor gradiente  $\nabla_{\mathbf{b}'} \varepsilon$  e da Hessiana  $\mathbf{H}$  do caso cepstral ( $\alpha = 0$ )), dada por

$$r_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\bar{D}(e^{j\omega})|^2} e^{j\omega k} d\omega \quad l = 0, 1, \dots, N' \quad (\text{E.38})$$

com

$$\bar{D}(e^{j\omega}) = \exp\left(\sum_{n=1}^{N'} c_0(n) e^{-j\omega n}\right) \quad (\text{E.39})$$

onde os  $N'$  coeficientes  $c_0'(n)$  correspondem ao cepstrum associado aos  $M$  coeficientes Mel-Cepstrais transformados  $b_0(n)$  (calculados na etapa anterior do método de Newton-Raphson), ou seja,  $\bar{D}(e^{j\omega})$  é tal que

$$\bar{D}(e^{j\omega}) = D(e^{j\omega}) \quad (\text{E.40})$$

onde  $D(e^{j\omega})$  é o modelo Mel-Cepstral expresso por

$$D(e^{j\omega}) = \exp\left(\sum_{n=1}^M b_0(n) \Phi_n(e^{j\omega})\right) \quad (\text{E.41})$$

O propósito de se expressar  $D(e^{j\omega})$  através de  $\bar{D}(e^{j\omega})$  é que  $\ln(\bar{D}(e^{j\omega}))$  pode ser calculada via FFT. Uma especificação completa do modelo Mel-Cepstral  $D(e^{j\omega})$  através de  $\bar{D}(e^{j\omega})$  exigiria, idealmente,  $N' = \infty$  amostras do cepstrum  $c_0'(n)$ . É suficiente, porém, que se adote um valor grande  $N' \gg 2M$ , sob o teto de  $N' = N_F - 1$  ( $N_F$  é o comprimento do periodograma amostrado). Aconselha-se a adoção de, no mínimo,  $N' \approx \frac{1+|\alpha|}{1-|\alpha|}(2M)$ <sup>§</sup>. A expressão para  $\tilde{r}_0(k)$ , utilizada no cálculo do gradiente e da Hessiana de  $\varepsilon$  para o caso mel-cepstral ( $\alpha \neq 0$ ) é dada por

$$\tilde{r}_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_k^*(e^{j\omega}) d\omega \quad k = 0, 1, \dots, 2M \quad (\text{E.42})$$

Como a integral é real, pode-se conjugá-la. Conjugando-a e considerando que ela corresponde ao caso  $i = 0$  da seguinte transformada inversa de Fourier

$$\tilde{r}_0(k, i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_k(e^{j\omega}) e^{j\omega i} d\omega \quad k = 0, 1, \dots, 2M \quad (\text{E.43})$$

temos, com  $i = 0$ , temos

$$\tilde{r}_0(k, 0) = \tilde{r}_0(k) \quad (\text{E.44})$$

---

<sup>§</sup> Como é assumido um único valor de  $N'$  para as transformações em frequência aqui envolvidas, adota-se o mínimo de  $N' + 1 \approx \frac{1+|\alpha|}{1-|\alpha|}(2M + 1)$  amostras, com base na aplicação do critério de suavização da seção 3.4.3 à transformação (E.46), que gera o maior número de amostras dentre as transformações ( $2M + 1$  amostras) e conta, portanto, com o maior número  $N' + 1$  de amostras de entrada.



Comparando (E.43) com (E.38), concluímos que  $\tilde{r}_0(k, i)$  é a saída do filtro  $\Phi_l(z)$ , com entrada  $r_0(-i)$ . A análise mel-cepstral generalizada de ordem  $M$ , exige o cálculo de  $2M + 1$  componentes  $\tilde{r}_0(k)$  (ver (E.42)) e conta com  $N' + 1$  componentes da seqüência  $r_\gamma(k)$ . A expressão (E.38) para tal seqüência, assim como o somatório da expressão (E.39) (para  $\bar{D}(e^{j\omega})$ ) são calculadas via FFT.

Considerando  $\Phi_n(e^{j\omega})$  na forma (E.1), a expressão (E.11) é contemplada pela filtragem em cascata representada na Figura E.4.

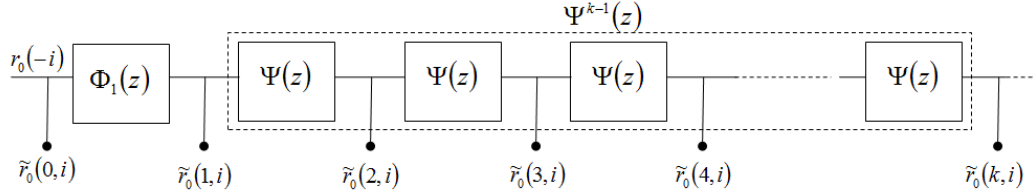


Figura E.4: Banco de filtros em cascata com entrada  $r_0(-i)$  e saída  $\tilde{r}_0(k, i)$  em sua  $k$ -ésima seção. Os valores  $\tilde{r}_0(k, 0)$  das seqüências de saída constituem a seqüência  $\tilde{r}_0(k)$ .

A filtragem da Figura E.1 pode ser computada pela fórmula recursiva

$$\tilde{r}_0(k, i) = \begin{cases} r_0(-i) & k = 0 \\ (1 - \alpha^2)\tilde{r}_0(0, i - 1) + \alpha\tilde{r}_0(1, i - 1), & k = 1 \\ \tilde{r}_0(k - 1, i - 1) + \alpha(\tilde{r}_0(k, i - 1) - \tilde{r}_0(k - 1, i)), & k = 2, 3, \dots, 2M \end{cases} \quad (\text{E.45})$$

$$i = -N', \dots, -2, -1, 0$$

$$\tilde{r}_0(k) = \tilde{r}_0(k, 0)$$

que pode ser deduzida com base em (E.2) e (E.3)<sup>§</sup>. Assumindo  $\mathbf{r}_{0,N'} = [r_0(0) r_0(1) \dots r_0(N')]^T$  e  $\tilde{\mathbf{r}}_{0,2M} = [\tilde{r}_0(0) \tilde{r}_0(1) \dots \tilde{r}_0(2M)]^T$ , a transformação que leva  $\mathbf{r}_{0,N'}$  em  $\tilde{\mathbf{r}}_{0,2M}$  é dada por

$$\tilde{\mathbf{r}}_{0,2M} = T_1(\mathbf{r}_{0,N'}, N', 2M, \alpha) \quad (\text{E.46})$$

que equivale à transformação (E.16) utilizada para o cálculo de  $\tilde{r}_\gamma(k)$ .

Como já considerado no Apêndice D,  $\tilde{p}_0(k)$  e  $\tilde{q}_0(k)$  são calculados, a partir de  $\tilde{r}_0(k)$ , por

$$\tilde{p}_0(k) = \begin{cases} (1 - \alpha^2)\tilde{r}_0(0) + 2\alpha\tilde{r}_0(1), & k = 0 \\ \tilde{r}_0(k) + \alpha\tilde{r}_0(k + 1), & k = 1, 2, \dots, M - 1 \end{cases} \quad \text{expressão (D.46)}$$

e

<sup>§</sup> A dedução é feita nos mesmos moldes daquela para a fórmula (3.29) na seção 3.4.1.

$$\tilde{q}_0(k) = \tilde{r}_0(k) + \alpha \tilde{r}_0(k-1), \quad k = 2, 3, \dots, 2M \quad \text{expressão (D.47)}$$

## Apêndice F

# Expressões para a Computação do Gradiente e da Hessiana de $\varepsilon$ via FFTs ( $\alpha \neq 0$ )

As integrais obtidas no Apêndice D para o cálculo dos elementos do vetor  $\nabla_{\mathbf{b}'} \varepsilon$  e da matriz Hessiana  $\mathbf{H}$  têm base na variável  $\omega$ . Na variável  $\tilde{\omega}$  (eixo não-linear de frequências) as integrais  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$ ,  $\tilde{q}_\gamma(k)$  (caso MGC) e  $\tilde{r}_0(k)$  (caso Mel-Cepstral) têm suas formas dadas por transformadas de Fourier, da mesma maneira que para os casos GC e Cepstral. Isso permite que sejam computadas via FFT. As formas das integrais  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$ ,  $\tilde{q}_\gamma(k)$  e  $\tilde{r}_0(k)$ , aqui dadas, resultam da aplicação da mudança de variável  $\omega = \omega(\tilde{\omega})$  às respectivas integrais correspondentes dadas no Apêndice D. A computação das FFTs exige que se disponha de  $I_N(\omega(\tilde{\omega}))$ , cujo método para obtenção é tratado na seção 4.8. A mudança de variável  $\omega = \omega(\tilde{\omega})$  aplicada ao critério  $\varepsilon$ , dado por (D.1), fornece

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \left( \frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} \right) d\tilde{\omega} \quad (\text{F.1})$$

onde  $\tilde{I}_N(\tilde{\omega}) = I_N(\omega(\tilde{\omega}))$  e  $\tilde{D}(e^{j\tilde{\omega}}) = D(e^{j\omega(\tilde{\omega})})$ . Com base em tal critério, podem ser obtidas as expressões para  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$ ,  $\tilde{q}_\gamma(k)$  e  $\tilde{r}_0(k)$  na variável  $\tilde{\omega}$ . Para auxílio em sua determinação as seguintes expressões serão úteis:

$$\frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} = \frac{(1 - \alpha^2)}{(e^{j\tilde{\omega}} + \alpha)(e^{-j\tilde{\omega}} + \alpha)} \quad (\text{F.2})$$

e

$$e^{-j\omega(\tilde{\omega})} = \frac{e^{-j\tilde{\omega}} + \alpha}{1 + \alpha e^{-j\tilde{\omega}}} \quad (\text{F.3})$$

obtida invertendo-se a expressão (B.2). Com base na expressão (B.1), assumindo ali  $\Phi_n(e^{j\omega(\tilde{\omega})}) = \tilde{\Phi}_n(e^{j\tilde{\omega}})$ , temos

$$\tilde{\Phi}_n(e^{j\tilde{\omega}}) = \begin{cases} 1 & n = 0 \\ \frac{(1 - \alpha^2)e^{-j\omega(\tilde{\omega})}}{1 - \alpha e^{-j\omega(\tilde{\omega})}} e^{-j\tilde{\omega}(n-1)} & n > 0 \end{cases} \quad (\text{F.4})$$

Substituindo (F.3) em (F.4), obtém-se

$$\tilde{\Phi}_n(e^{j\tilde{\omega}}) = \begin{cases} 1, & n = 0 \\ (e^{-j\tilde{\omega}} + \alpha)e^{-j\tilde{\omega}(n-1)}, & n > 0 \end{cases} \quad (\text{F.5})$$

As expressões a seguir podem ser deduzidas a partir de (F.2) e (F.5) e serão utilizadas nas expressões do Apêndice D para a obtenção das integrais associadas aos elementos de  $\mathbf{\nabla}_b \boldsymbol{\varepsilon}$  e  $\mathbf{H}$  na variável  $\tilde{\omega}$ . São elas:

$$\tilde{\Phi}_k(e^{j\tilde{\omega}}) = \begin{cases} 1, & k = 0 \\ \left(\frac{d\omega(\tilde{\omega})}{d\tilde{\omega}}\right)^{-1} \frac{(1 - \alpha^2)}{(1 + \alpha e^{-j\tilde{\omega}})} e^{-j\tilde{\omega}k}, & k > 0 \end{cases} \quad (\text{F.6})$$

$$\tilde{\Phi}_1^*(e^{j\tilde{\omega}})\tilde{\Phi}_{k+1}(e^{j\tilde{\omega}}) = \left(\frac{d\omega(\tilde{\omega})}{d\tilde{\omega}}\right)^{-1} (1 - \alpha^2)e^{-j\tilde{\omega}k}, \quad k \geq 0 \quad (\text{F.7})$$

$$\tilde{\Phi}_1(e^{j\tilde{\omega}})\tilde{\Phi}_{k-1}(e^{j\tilde{\omega}}) = \left(\frac{d\omega(\tilde{\omega})}{d\tilde{\omega}}\right)^{-1} (1 - \alpha^2) \left(\frac{e^{-j\tilde{\omega}} + \alpha}{1 + \alpha e^{-j\tilde{\omega}}}\right) e^{-j\tilde{\omega}(k-1)}, \quad k > 1 \quad (\text{F.8})$$

## F.1 Análise Mel-Cepstral Generalizada

Para o caso mel-cepstral generalizado, fazendo a substituição de variáveis  $\omega = \omega(\tilde{\omega})$  em (D.16), (D.32) e (D.33), obtém-se

$$\tilde{r}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \tilde{\Phi}_k^*(e^{j\tilde{\omega}}) \left( \frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} \right) d\tilde{\omega} \quad k = 0, 1, \dots, M \quad (\text{F.9})$$

$$\tilde{p}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{\Phi}_1(e^{j\tilde{\omega}}) \cdot \tilde{\Phi}_{k+1}^*(e^{j\tilde{\omega}}) \left( \frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} \right) d\tilde{\omega} \quad k = 0, 1, \dots, M-1 \quad (\text{F.10})$$

e

$$\tilde{q}_\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+2\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^{2\gamma}(e^{j\tilde{\omega}}) \tilde{\Phi}_1^*(e^{j\tilde{\omega}}) \cdot \tilde{\Phi}_{k-1}^*(e^{j\tilde{\omega}}) \left( \frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} \right) d\tilde{\omega} \quad (\text{F.11})$$

$$k = 2, 3, \dots, 2M$$

onde em (F.9), inclui-se  $\tilde{r}_\gamma(0)$ , necessário para o cálculo de  $\varepsilon$ . Substituindo (F.6), (F.7) e (F.8) em (F.9), (F.10) e (F.11), respectivamente, são obtidos  $\tilde{r}_\gamma(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  nas formas

$$\tilde{r}_\gamma(k) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \frac{(1-\alpha^2)}{(e^{j\tilde{\omega}} + \alpha)(e^{-j\tilde{\omega}} + \alpha)} d\tilde{\omega}, & k = 0 \\ (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \frac{1}{(1 + \alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}k} d\tilde{\omega}, & k = 1, 2, \dots, M \end{cases} \quad (\text{F.12})$$

$$\tilde{p}_\gamma(k) = (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} e^{j\tilde{\omega}k} d\tilde{\omega} \quad k = 0, 1, \dots, M-1 \quad (\text{F.13})$$

e

$$\tilde{q}_\gamma(k) = (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+2\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^{2\gamma}(e^{j\tilde{\omega}}) \left( \frac{e^{-j\tilde{\omega}} + \alpha}{1 + \alpha e^{-j\tilde{\omega}}} \right) e^{-j\tilde{\omega}(k-1)} d\tilde{\omega} \quad (\text{F.14})$$

$$k = 2, 3, \dots, 2M$$

As integrais (F.13) e (F.14) são calculadas via FFT. Para  $k > 0$ , a integral em (F.12), se calculada via FFT, não contemplaria  $\tilde{r}_\gamma(0)$ .  $\tilde{r}_\gamma(k)$  para  $k \geq 0$  pode ser obtida calculando, via FFT, a seguinte integral auxiliar

$$a(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) e^{j\tilde{\omega}k} d\tilde{\omega} \quad k = -\frac{N}{2}, \dots, -1, 0, 1, \frac{N}{2} - 1 \quad (\text{F.15})$$

de onde determina-se recursivamente

$$a_1(n) = \begin{cases} a\left(\frac{N}{2} - 1\right), & n = \frac{N}{2} - 1 \\ a(n) - \alpha a_1(n+1), & 0 \leq n < \frac{N}{2} - 1 \end{cases} \quad (\text{F.16})$$

e

$$a_2(n) = \begin{cases} a\left(-\frac{N}{2}\right), & n = -\frac{N}{2} \\ a(n) - \alpha a_2(n-1), & -\frac{N}{2} < n \leq -1 \end{cases} \quad (\text{F.17})$$

Com base em  $a_1(n)$  e  $a_2(n)$ ,  $\tilde{r}_\gamma(k)$  é dado por

$$\tilde{r}_\gamma(k) = \begin{cases} a_1(0) - \alpha a_2(-1), & k = 0 \\ (1 - \alpha^2) a_1(k), & k = 1, 2, \dots, M \end{cases} \quad (\text{F.18})$$

o que é demonstrado no final do tópic.

Nas integrais para  $a(k)$ ,  $\tilde{p}_\gamma(k)$  e  $\tilde{q}_\gamma(k)$  ((F.15), (F.13) e (F.14)), o modelo  $\tilde{D}(e^{j\tilde{\omega}})$  na forma

$$\tilde{D}(e^{j\tilde{\omega}}) = \left( 1 + \gamma(1 + \alpha e^{j\tilde{\omega}}) \sum_{n=1}^M b'_\gamma(n) e^{-j\tilde{\omega}n} \right)^{1/\gamma} \quad (\text{F.19})$$

permite o cálculo via FFT do fator do somatório. A expressão (F.19) é obtida substituindo-se (F.5) em (E.10).

Uma vez calculado  $\tilde{r}_\gamma(k)$ , pode-se calcular  $\varepsilon$  através de

$$\varepsilon = \tilde{r}_\gamma(0) + \gamma \mathbf{r}^T \cdot \mathbf{b}'_\gamma \quad (\text{F.20})$$

onde  $\mathbf{r} = [\tilde{r}_\gamma(1) \ \tilde{r}_\gamma(2) \ \dots \ \tilde{r}_\gamma(M)]^T$  e  $\mathbf{b}'_\gamma = [b'_\gamma(1) \ b'_\gamma(2) \ \dots \ b'_\gamma(M)]^T$  ( $\mathbf{b}'_\gamma$  é o vetor dos coeficientes calculados na iteração anterior e  $\tilde{r}_\gamma(0)$  e  $\mathbf{r}$  são calculados com base no mesmo vetor  $\mathbf{b}'_\gamma$ ). A validade de (F.20) é demonstrada no Apêndice E (seção E.1.4). Os elementos associados ao vetor gradiente  $\nabla_{\mathbf{b}'_\gamma} \varepsilon$  são dados por

$$\tilde{r}_\gamma(k), \quad k = 1, 2, \dots, M$$

e aqueles associados à matriz Hessiana  $\mathbf{H}$ , por (F.13) e (F.14).

### Demonstração de (F.18)

A fórmula (F.16) contempla  $a_1(n)$  como saída do filtro  $\frac{1}{(1+\alpha e^{j\tilde{\omega}})}$  cuja entrada é  $a(n)$ . Assim  $a_1(n)$  é representada por

$$a_1(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \frac{1}{(1+\alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}n} d\tilde{\omega} \quad (\text{F.21})$$

Da mesma forma, (F.17) contempla  $a_2(n)$  como saída do filtro  $\frac{1}{(1+\alpha e^{-j\tilde{\omega}})}$  cuja entrada é  $a(n)$ . Assim  $a_2(n)$  é representada por

$$a_2(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}^{1+\gamma}(e^{j\tilde{\omega}})|^2} \tilde{D}^\gamma(e^{j\tilde{\omega}}) \frac{1}{(1+\alpha e^{-j\tilde{\omega}})} e^{j\tilde{\omega}n} d\tilde{\omega} \quad (\text{F.22})$$

Substituindo (F.21) e (F.22) em (F.18), verifica-se que (F.18) equivale a (F.12).

## F.2 Análise Mel-Cepstral

Para o caso mel-cepstral, fazendo a substituição de variáveis  $\omega = \omega(\tilde{\omega})$  em (D.35), obtém-se

$$\tilde{r}_0(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \tilde{\Phi}_k^*(e^{j\tilde{\omega}}) \left( \frac{d\omega(\tilde{\omega})}{d\tilde{\omega}} \right) d\tilde{\omega} \quad k = 0, 1, \dots, 2M \quad (\text{F.23})$$

Substituindo (F.6) em (F.23), obtém-se  $\tilde{r}_0(k)$  na forma

$$\tilde{r}_0(k) = \begin{cases} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \frac{(1-\alpha^2)}{(e^{j\tilde{\omega}} + \alpha)(e^{-j\tilde{\omega}} + \alpha)} d\tilde{\omega} & k = 0 \\ (1-\alpha^2) \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \frac{1}{(1+\alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}k} d\tilde{\omega}, & k = 1, 2, \dots, 2M \end{cases} \quad (\text{F.24})$$

Assim, para o cálculo de  $\tilde{r}_0(k)$ , computa-se, via FFT, a integral auxiliar

$$a_0(l) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{I}_N(\tilde{\omega})}{|\tilde{D}(e^{j\tilde{\omega}})|^2} \frac{1}{(1 + \alpha e^{j\tilde{\omega}})} e^{j\tilde{\omega}l} d\tilde{\omega} \quad l = 0, 1, \dots, 2M \quad (\text{F.25})$$

e  $\tilde{r}_0(k)$  é obtido por

$$\tilde{r}_0(k) = \begin{cases} a_0(0) - \alpha a_0(1), & k = 0 \\ (1 - \alpha^2) a_0(k), & k = 1, 2, \dots, 2M \end{cases} \quad (\text{F.26})$$

expressão cuja consistência com (F.24) (para  $k = 0$ ) pode ser facilmente verificada.

Para o cálculo de  $a_0(l)$ , utiliza-se

$$\tilde{D}(e^{j\tilde{\omega}}) = \exp \left( (1 + \alpha e^{j\tilde{\omega}}) \sum_{n=1}^M b'_\gamma(n) e^{-j\tilde{\omega}n} \right) \quad (\text{F.27})$$

expressão na qual o fator do somatório pode ser calculado via FFT.

O critério espectral  $\varepsilon$  é dado por  $\tilde{r}_0(0)$  (o que é verificado em (F.24) para  $k = 0$ , que equilibra a (F.1)). Como já considerado no Apêndice D, os elementos associados ao vetor gradiente  $\nabla_{\mathbf{b}'_\gamma} \varepsilon$  são dados por

$$\tilde{r}_0(k), \quad k = 1, 2, \dots, M$$

e aqueles associados à matriz Hessiana  $\mathbf{H}$ , por

$$\tilde{p}_0(k) = \begin{cases} (1 - \alpha^2) \tilde{r}_0(0) + 2\alpha \tilde{r}_0(1), & k = 0 \\ \tilde{r}_0(k) + \alpha \tilde{r}_0(k + 1), & k = 1, 2, \dots, M - 1 \end{cases} \quad \text{expressão (D.46)}$$

e

$$\tilde{q}_0(k) = \tilde{r}_0(k) + \alpha \tilde{r}_0(k - 1), \quad k = 2, 3, \dots, 2M \quad \text{expressão (D.47)}$$



# Referências Bibliográficas

- [1] Keiichi Tokuda, Takao Kobayashi, Takashi Masuko, and Satoshi Imai, "Mel-Generalized Cepstral Analysis - A Unified Approach to Speech Spectral Estimation," *Proc. ICSLP-94*, vol. 3, pp. 1043-1046, Sep. 1994.
- [2] Keiichi Tokuda, Takao Kobayashi, Satoshi Imai, and Takeshi Chiba, "Spectral Estimation of Speech by Mel-Generalized Cepstral Analysis," *Electronics and Communications in Japan, Part 3*, vol. 76, no. 2, pp. 30-43, 1993.
- [3] B. S. Atal and Suzanne L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *The Journal of the Acoustic Society of America*, vol. 50, no. 2B, pp. 637-655, Apr. 1971.
- [4] S. Imai and C. Furuichi, "Unbiased Estimator of Log Spectrum and Its Application to Speech Signal Processing," *Proc. 1988 Eurasip*, pp. 203-206, 1988.
- [5] Keiichi Tokuda, Takao Kobayashi, and Satochi Imai, "Adaptive Cepstral Analysis of Speech," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 6, pp. 481-489, Nov. 1995.
- [6] Toshiaki Fukada, Keiichi Tokuda, Takao Kobayashi, and Satoshi Imai, "An Adaptive Algorithm for Mel-Cepstral Analysis of Speech," *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92.*, vol. 1, pp. 137-140, Mar. 1992.
- [7] K. Tokuda, T. Kobayashi, T. Fukada, H. Saito, and S. Imai, "Spectral estimation of speech based on mel-cepstral representation," *Trans. (A) I.E.I.C.E., Japan*, vol. J74-A, no. 8, pp. 1240-1248, Aug. 1991. (em Japonês).
- [8] E. Kruger and H. W. Strube, "Linear Prediction on a Warped Frequency Scale," *IEEE Trans., Acoust., Speech, and Signal Processing*, vol. ASSP-36, pp. 1529-1531, 1988.
- [9] Keiichi Tokuda, Takao Kobayashi, and Satochi Imai, "Generalized Cepstral Analysis of Speech - Unified Approach to LPC and Cepstral Method," *Proc. ICSLP-90*, pp. 37-40, 1990.
- [10] Kazuhito Koishida, Keiichi Tokuda, Takao Kobayashi, and Satoshi Imai, "Spectral Representation of Speech Based on Mel-Generalized Cepstral Coefficients and Its Properties," *Electronics and Communications in Japan, Part 3*, vol. 83, no. 3, pp. 50-59, Mar. 2000.
- [11] Kazuhito Koishida, Keiichi Tokuda, Takao Kobayashi, and Satoshi Imai, "CELP Coding System Based on Mel-Generalized Cepstral Analysis," *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*,

vol. 1, pp. 318-321, Oct. 1996.

- [12] Kazuhito Koishida, Keiichi Tokuda, Takao Kobayashi, and Satoshi Imai, "Efficient Encoding of Mel-Generalized Cepstrum for CELP Coders," *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 2, pp. 1355-1358, Apr. 1997.
- [13] Kazuhito Koishida, Gou HIRABAYASHI, Keiichi Tokuda, and Takao Kobayashi, "A 16 kb/s Wideband CELP-Based Speech Coder Using Mel-Generalized Cepstral Analysis," *IEICE Trans. Inf. & Syst.*, vol. E83-D, no. 4, pp. 876-883, Apr. 2000.
- [14] Heiga Zen, Tomoki Toda, and Keiichi Tokuda, "The Nitech-NAIST HMM-Based Speech Synthesis System for the Blizzard Challenge 2006," *IEICE TRANS. INF. & SYST., VOL.E91-D, NO.6 JUNE*, vol. E91-D, no. 6, pp. 1764-1773, June 2008.
- [15] T. KOBAYASHI and S. IMAI, "Spectral Analysis Using Generalized Cepstrum," *IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING*, vol. ASSP-3, no. 5, pp. 1087-1089, Aug. 1984.
- [16] Jae S. Lim, "Spectral Root Homomorphic Deconvolution System," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. AASP2-7, no. 3, pp. 223-233, Junho 1979.
- [17] A. V. Oppenheim and R. W. Shafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975, ch. 10.
- [18] Satoshi Imai, "Adaptive mel cepstral analysis based on UELS method," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, vol. AS-SPCC. The IEEE 2000, Lake Louise, Alta., Canada., 2000, pp. 304 - 309.
- [19] Murray R. Spiegel, *Schaum's Outlines: Complex Variables With an Introduction to Conformal Mapping and Its Applications*, 1st ed. USA: McGraw-Hill, 1964.
- [20] Takao Kobayashi and Satoshi Imai, "Speech Synthesis Using Generalized Cepstrum," *Electronics and Communications in Japan*, vol. 65-A, no. 3, pp. 28-36, 1982.
- [21] A. V. Oppenheim and D. H. Johnson, "Discrete Representation of signals," *Proceedings of the IEEE*, vol. 60, no. 6, pp. 681-691, June 1972.
- [22] Keiichi Tokuda, Takao Kobayashi, and Satoshi Imai. (1994, Apr.) Recursive Calculation of Mel-Cepstrum from LP Coefficients. [Online]. [http://www.sp.nitech.ac.jp/~tokuda/tips/mgceptr\\_sa2.pdf](http://www.sp.nitech.ac.jp/~tokuda/tips/mgceptr_sa2.pdf)
- [23] K. Tokuda, T. Kobayashi, and S. Imai, "Recursion Formula for Calculation of Mel Generalized Cepstrum Coefficients," *Trans. IEICE*, vol. J71-A, pp. 128-131, Jan. 1988. (em Japonês).

- [24] F. Itakura and S. Saito, "Speech Information Compression based on the Maximum Likelihood Spectral Estimation," *J. Acous. Soc. Japan*, vol. 27, no. 9, pp. 463-472, Sep. 1971.
- [25] M. Pinsker, *Information and Information Stability of Random Variables*. San Francisco: Holden-Day, 1963.
- [26] I. Gohberg and I. Koltracht, "Efficient algorithm for Toeplitz plus Hankel matrices," *Integral Equations and Operator Theory*, vol. 12, pp. 136-142, 1989.
- [27] A. E. Yagle, "New analogs of split algorithms for arbitrary Toeplitz-plus-Hankel matrices," *IEEE Trans. Acoust., Speech & Signal Process.*, vol. ASSP-39, no. 11, pp. 2457-2463, Nov. 1991.
- [28] Ramiro R. A. Barreira and Fábio Violaro, "Avaliação da Análise Cepstral Generalizada Aplicada à Modelagem de Vogais," in *Anais do 7o Congresso da AES Brasil 2009*, São Paulo, 2009, pp. 95-101.
- [29] Ramiro R. A. Barreira and Fábio Violaro, "Uma Avaliação da Análise Cepstral Generalizada Aplicada ao Espectro de Vogais," in *Anais do XXVII Simpósio Brasileiro de Telecomunicações, 2009*, Blumenau, 2009.
- [30] Satoshi Imai, Kazuo Sumita, and Chieko Furuichi, "Mel Log Spectrum Approximation (MLSA) Filter for Speech Synthesis," *Electronics and Communications in Japan*, vol. 66-A, no. 2, pp. 10-18, 1983.
- [31] T. Kobayashi, S. Imai, and Y. Fukuda, "Mel generalized-log spectrum approximation (MGLSA) filter," *Journal of IEICE (Japanese Edition)*, vol. J68-A, no. 6, pp. 610-611, 1985.
- [32] Heiga Zen, Keiichi Tokuda, and Alan W. Black, "Statistical Parametric Speech Synthesis," *Speech Communication*, vol. 51, no. 11, pp. 1039-1064, Nov. 2009.
- [33] Hideki Kawahara, Ikuyo Masuda-Katsuse, and Alain de Cheveigné, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," *Speech Communication*, vol. 27, pp. 187-207, 1999.
- [34] Heiga Zen, Tomoki Toda, Masaru Nakamura, and Keiichi Tokuda, "Details of the Nitech HMM-Based Speech Synthesis System for the Blizzard Challenge 2005," *IEICE Trans, Inf. & Sust.*, vol. E90-D, no. 1, pp. 325-333, Jan. 2007.
- [35] Hideki Kawahara, Jo Estill, and Osamu Fujimura, "Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT," in *2nd MAVEBA*, Firenze, Italy, 2001.
- [36] R. Maia, T. Toda, H. Zen, Y. Nankaku, and K. Tokuda, "An excitation model for HMM-based speech synthesis based on residual modeling," in *Proc. ISCA SSW6*,

2007, pp. 131-136.

- [37] Thomas Drugman, Geoffrey Wilfart, and Thierry Dutoit, "A Deterministic plus Stochastic Model of the Residual Signal for Improved Parametric Speech Synthesis," in *Proc. Interspeech 2009*, Brighton, UK, 2009.
- [38] Thomas Drugman, Geoffrey Wilfart, and Thierry Dutoit, "EIGENRESIDUALS FOR IMPROVED PARAMETRIC SPEECH SYNTHESIS," in *17th European Signal Processing Conference (EUSIPCO 2009)*, Glasgow, Scotland, 2009, pp. 2176-2180.
- [39] T. Toda and K. Tokuda, "A speech parameter generation algorithm considering global variance for HMM-based speech synthesis," *IEICE Trans. Inf. Syst.*, vol. E90-D, no. 5, pp. 816–824, 2007.
- [40] Chiyomi Miyajima, Hideyuki Watanabe, Keiichi Tokuda, Tadashi Kitamura, and Shigeru Katagiri, "A new approach to designing a feature extractor in speaker identification based on discriminative feature extraction," *Speech Communication*, vol. 35, pp. 203-218, 2001.