

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**DIEGO ECKHARD**

**FERRAMENTAS PARA MELHORIA DA  
CONVERGÊNCIA DOS MÉTODOS DE  
IDENTIFICAÇÃO POR ERRO DE  
PREDIÇÃO**

Porto Alegre  
2012

**DIEGO ECKHARD**

**FERRAMENTAS PARA MELHORIA DA  
CONVERGÊNCIA DOS MÉTODOS DE  
IDENTIFICAÇÃO POR ERRO DE  
PREDIÇÃO**

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Controle e Automação

ORIENTADOR: Prof. Dr. Alexandre Sanfelice  
Bazanella

Porto Alegre  
2012

**DIEGO ECKHARD**

**FERRAMENTAS PARA MELHORIA DA  
CONVERGÊNCIA DOS MÉTODOS DE  
IDENTIFICAÇÃO POR ERRO DE  
PREDIÇÃO**

Esta tese foi julgada adequada para a obtenção do título de Doutor em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: \_\_\_\_\_  
Prof. Dr. Alexandre Sanfelice Bazanella, UFRGS  
Doutor pela Universidade Federal de Santa Catarina – Florianópolis, Brasil

Banca Examinadora:

Prof. Dr. Pericles Rezende Barros, UFCG  
Ph.D. pela University of Newcastle – Newcastle, Austrália.

Prof. Dr. Liu Hsu, UFRJ  
Doutor pela Université de Toulouse III (Paul Sabatier) – Toulouse, França

Prof. Dr. João Manoel Gomes da Silva Jr., UFRGS  
Doutor pela Université de Toulouse III (Paul Sabatier) – Toulouse, França

Prof. Dr. Luís Fernando Alves Pereira, UFRGS  
Doutor pelo Instituto Tecnológico de Aeronáutica – São José dos Campos, Brasil

Prof. Dr. Sérgio Luís Haffner, UFRGS  
Doutor pela Universidade Estadual de Campinas – Campinas, Brasil.

Coordenador do PPGEE: \_\_\_\_\_  
Prof. Dr. João Manoel Gomes da Silva Jr.

Porto Alegre, dezembro de 2012.

## **AGRADECIMENTOS**

Gostaria de agradecer à Priscila por todo amor, apoio e compreensão.

Aos colegas Luciola, Jeferson, Maurício, Manoel e Rafael pela amizade e pelas inúmeras contribuições à este trabalho.

Ao Prof. Alexandre Bazanella pela excelente orientação na pós-graduação.

Ao supervisor de estágio sanduíche Prof. Håkan Hjalmarsson e ao Prof. Cristian Rojas pela recepção na KTH na Suécia.

Ao PPGEE-UFRGS pelo ensino de qualidade, ao CNPq pela concessão da bolsa de doutorado e à CAPES pela concessão da bolsa de estágio sanduíche.

## RESUMO

O método de identificação por minimização do erro de predição está relacionado com um problema de otimização não convexo. É comum utilizar algoritmos iterativos para resolver o problema de otimização. Contudo, os algoritmos iterativos podem ficar presos em mínimos locais da função custo ou convergir para a borda do domínio de busca. Uma análise da função custo e condições suficientes para garantir a convergência dos algoritmos iterativos para o mínimo global são apresentadas neste trabalho. Observa-se que estas condições dependem do espectro do sinal de entrada utilizado no experimento. Este trabalho apresenta ferramentas para melhorar a convergência dos algoritmos para o mínimo global, as quais são baseadas na manipulação do espectro do sinal de entrada.

**Palavras-chave: Identificação de Sistemas, Convergência, Projeto de Experimentos.**

## **ABSTRACT**

The Prediction Error Method is related to a non-convex optimization problem. It is usual to apply iterative algorithms to solve this optimization problem. However, iterative algorithms can get stuck at a local minimum of the cost function or converge to the border of the searching space. An analysis of the cost function and sufficient conditions to ensure the convergence of the iterative algorithms to the global minimum are presented in this work. It is observed that these conditions depend on the spectrum of the input signal used in the experiment. This work presents tools to improve the convergence of the algorithms to the global minimum, which are based on the manipulation of the input spectrum.

**Keywords: System Identification, Convergence, Experiment Design.**

# SUMÁRIO

<b>LISTA DE ILUSTRAÇÕES</b> . . . . .	8
<b>LISTA DE TABELAS</b> . . . . .	10
<b>LISTA DE ABREVIATURAS</b> . . . . .	11
<b>LISTA DE SÍMBOLOS</b> . . . . .	12
<b>1 INTRODUÇÃO</b> . . . . .	13
<b>2 IDENTIFICAÇÃO DE SISTEMAS</b> . . . . .	16
2.1 O sistema . . . . .	16
2.2 O modelo de predição . . . . .	17
2.3 A estrutura do modelo . . . . .	19
2.4 O método da minimização do erro de predição . . . . .	20
2.4.1 Propriedades assintóticas da estimativa . . . . .	21
2.4.2 Obtendo a estimativa . . . . .	23
2.5 O método dos mínimos quadrados . . . . .	24
2.6 O método das variáveis instrumentais . . . . .	25
2.7 Considerações finais . . . . .	27
<b>3 OTIMIZAÇÃO</b> . . . . .	28
3.1 Definições . . . . .	28
3.2 Algoritmos de otimização . . . . .	30
3.3 Alguns algoritmos e propriedades de convergências . . . . .	33
3.3.1 O método do gradiente . . . . .	33
3.3.2 O método de Newton-Raphson . . . . .	37
3.3.3 Outros algoritmos de busca . . . . .	39
3.4 Robustez . . . . .	41
3.5 Qual algoritmo escolher . . . . .	42
3.6 Considerações finais . . . . .	43
<b>4 CONVERGÊNCIA PARA O MÍNIMO GLOBAL</b> . . . . .	45
4.1 Função semi-convexa . . . . .	45
4.2 Propriedades da função $\bar{V}(\theta)$ . . . . .	49
4.3 Parametrização genérica . . . . .	50
4.4 Parametrizações usuais . . . . .	53
4.4.1 Estrutura ARX generalizada . . . . .	53
4.4.2 Estrutura ARMAX generalizada . . . . .	54

4.4.3	Estrutura OE generalizada . . . . .	55
<b>4.5</b>	<b>Parametrização independente . . . . .</b>	<b>56</b>
<b>4.6</b>	<b>Considerações finais . . . . .</b>	<b>58</b>
<b>5</b>	<b>PROJETO DE EXPERIMENTOS . . . . .</b>	<b>59</b>
<b>5.1</b>	<b>Projeto do sinal de entrada . . . . .</b>	<b>59</b>
<b>5.2</b>	<b>Espectro discreto . . . . .</b>	<b>62</b>
5.2.1	Positividade do espectro . . . . .	63
5.2.2	Restrição em potência média . . . . .	63
5.2.3	Restrição na amplitude do espectro . . . . .	64
5.2.4	Restrição na matriz de covariância . . . . .	64
<b>5.3</b>	<b>Espectro contínuo . . . . .</b>	<b>64</b>
5.3.1	Positividade do espectro . . . . .	65
5.3.2	Restrição em potência média . . . . .	66
5.3.3	Restrição na amplitude do espectro . . . . .	66
5.3.4	Restrição na matriz de covariância . . . . .	67
<b>5.4</b>	<b>Exemplos . . . . .</b>	<b>67</b>
5.4.1	Sistema tipo FIR . . . . .	67
5.4.2	Sistema tipo ARX . . . . .	69
<b>5.5</b>	<b>Considerações finais . . . . .</b>	<b>70</b>
<b>6</b>	<b>RESTRICÃO DE CONVERGÊNCIA . . . . .</b>	<b>72</b>
<b>6.1</b>	<b>Restrições infinitas . . . . .</b>	<b>72</b>
<b>6.2</b>	<b>Scenario Approach . . . . .</b>	<b>74</b>
6.2.1	Exemplo . . . . .	76
<b>6.3</b>	<b>S-procedure . . . . .</b>	<b>82</b>
6.3.1	Estrutura Output-Error generalizada . . . . .	84
6.3.2	Utilização da abordagem . . . . .	85
6.3.3	Exemplo . . . . .	85
6.3.4	Projeto de experimento sem restrições de convergência . . . . .	87
6.3.5	Projeto de experimento com restrições de convergência . . . . .	88
<b>6.4</b>	<b>Considerações finais . . . . .</b>	<b>90</b>
<b>7</b>	<b>FILTRAGEM ITERATIVA . . . . .</b>	<b>91</b>
<b>7.1</b>	<b>Estrutura Output-Error generalizada . . . . .</b>	<b>92</b>
<b>7.2</b>	<b>Estrutura ARMAX generalizada . . . . .</b>	<b>94</b>
<b>7.3</b>	<b>Estrutura Box-Jenkins generalizada . . . . .</b>	<b>95</b>
<b>7.4</b>	<b>Exemplo . . . . .</b>	<b>96</b>
<b>7.5</b>	<b>Considerações finais . . . . .</b>	<b>99</b>
<b>8</b>	<b>CONCLUSÃO . . . . .</b>	<b>101</b>
<b>APÊNDICE A</b>	<b>. . . . .</b>	<b>103</b>
<b>A.1</b>	<b>Teorema KYP . . . . .</b>	<b>103</b>
<b>A.2</b>	<b>Teorema de Parseval . . . . .</b>	<b>103</b>
<b>REFERÊNCIAS</b>	<b>. . . . .</b>	<b>104</b>

## LISTA DE ILUSTRAÇÕES

Figura 3.1:	Uma função de Lyapunov e seus conjuntos de nível. . . . .	33
Figura 3.2:	Função custo $V(\theta) = e^\theta(\theta - 1)^2$ . . . . .	35
Figura 3.3:	Convergência do método do gradiente com a condição inicial $\theta^1 = -0,5$ e passos $\gamma(\theta^i) = 0,2 \forall i$ . . . . .	35
Figura 3.4:	Divergência do método do gradiente para a condição inicial $\theta^1 = 1,1$ e passos $\gamma(\theta^i) = 0,7 \forall i$ . . . . .	36
Figura 3.5:	Divergência do método do gradiente para a condição inicial $\theta^1 = -1,5$ e passos $\gamma(\theta^i) = 0,5 \forall i$ . . . . .	36
Figura 3.6:	Convergência do método do gradiente com a condição inicial $\theta^1 = -0,8$ e tamanho de passos $\gamma(\theta^i) = 0,2 \forall i$ . . . . .	40
Figura 3.7:	Divergência do método de Newton-Raphson com a condição inicial $\theta^1 = 0,2$ . . . . .	40
Figura 3.8:	Convergência do método de Newton-Raphson com a condição inicial $\theta^1 = 0,5$ . . . . .	41
Figura 4.1:	Direção do gradiente. . . . .	46
Figura 4.2:	Exemplo de conjuntos de nível de uma função semi-convexa. . . . .	48
Figura 4.3:	Exemplo de conjuntos de nível de funções semi-convexas (à direita) e não semi-convexas (à esquerda). . . . .	48
Figura 5.1:	Simulação Monte-Carlo e Intervalo de Confiança de 95%. . . . .	69
Figura 5.2:	Simulação Monte-Carlo e Intervalo de Confiança de 95%. . . . .	71
Figura 5.3:	Espectro do sinal de entrada . . . . .	71
Figura 5.4:	Realização do sinal de entrada $u(t)$ . . . . .	71
Figura 6.1:	Exemplo de conjunto $\mathcal{C}$ . . . . .	74
Figura 6.2:	Exemplo de alguns pontos do conjunto $\mathcal{C}$ . . . . .	75
Figura 6.3:	Exemplo dos conjuntos. . . . .	75
Figura 6.4:	Diagrama de Bode de $G_0(z)$ . . . . .	77
Figura 6.5:	Espectros. . . . .	79
Figura 6.6:	Rodadas de Monte-Carlo para $\Phi_1(\omega)$ . . . . .	80
Figura 6.7:	Boxplot das rodadas de Monte-Carlo para $\Phi_1(\omega)$ . . . . .	80
Figura 6.8:	Rodadas de Monte-Carlo para $\Phi_2(\omega)$ . . . . .	81
Figura 6.9:	Boxplot das rodadas de Monte-Carlo para $\Phi_2(\omega)$ . . . . .	81
Figura 6.10:	Resposta ao degrau do sistema. . . . .	86
Figura 6.11:	Diagrama de Bode do sistema. . . . .	86
Figura 6.12:	Sinal de entrada projetado sem as restrições de convergência. . . . .	88
Figura 6.13:	Elipsoide de restrições $\mathcal{C}$ projetado no subespaço $\theta = [0 \ x \ x \ x]^T$ . . . . .	89

Figura 6.14: Sinal de entrada calculado com as restrições de convergência. . . . .	89
Figura 6.15: Resposta ao degrau do sistema real e do modelo. . . . .	90
Figura 7.1: Rodadas de Monte-Carlo com a plataforma <code>unit</code> . . . . .	98
Figura 7.2: Rodadas de Monte-Carlo com a plataforma <code>ident</code> . . . . .	99
Figura 7.3: Rodadas de Monte-Carlo com o algoritmo proposto. . . . .	100

## LISTA DE TABELAS

Tabela 2.1: Distribuição $\chi^2_\alpha$ para alguns níveis de confiança $\alpha$ e diferentes graus de liberdade $n$ . . . . .	22
---	----

## LISTA DE ABREVIATURAS

AR	Auto-regressivo
ARMAX	Auto-regressivo com média móvel e entrada exógena
ARX	Auto-regressivo com entrada exógena
BJ	Box-Jenkins
FIR	Finite Impulse Response
KYP	Lema Kalman-Yakubovich-Popov
LMI	Desigualdade matricial linear
OE	Output Error
PRBS	Sinal binário pseudo-randômico
SISO	Monovariável
SPR	Strictly Positive Real

## LISTA DE SÍMBOLOS

$\mathcal{S}$	Sistema
$\mathcal{M}(\theta)$	Modelo
$\mathcal{M}^*$	Conjunto de modelos
$\mathcal{M}$	Estrutura do modelo
$\theta$	Parâmetros do modelo
$n$	Dimensão de $\theta$
$y(t)$	Sinal de saída
$u(t)$	Sinal de entrada
$\Phi_u(\omega)$	Espectro do sinal de entrada
$e(t)$	Ruído branco
$\Phi_e(\omega)$	Espectro do ruído
$\varepsilon(t, \theta)$	Erro de predição
$\hat{y}(t, \theta)$	Predição um passo a frente do modelo
$\mathcal{Z}^N$	Conjunto de dados
$N$	Quantidade de dados coletados
$V(\theta)$	Função custo
$V_N(\theta, \mathcal{Z}^N)$	Função custo para $N$ dados
$A^T$	Transposto de $A$
$\mathbb{R}$	Conjunto dos números reais
$\mathbb{R}^+$	Conjunto dos números reais não negativos
$\mathbb{N}$	Conjunto dos números naturais
$I_n$	Matriz identidade de dimensão $n \times n$
$\mathbf{0}_n$	Matriz de zeros de dimensão $n \times n$
$\mathbf{0}$	Vetor de zeros

# 1 INTRODUÇÃO

Modelos matemáticos de sistemas dinâmicos são fundamentais em várias áreas do conhecimento. Basicamente, existem três maneiras de se obter um modelo matemático. A primeira é conhecida como *modelagem matemática*, a qual utiliza leis básicas da física para descrever o comportamento dinâmico de um fenômeno ou processo. A segunda, chamada de *identificação de sistemas caixa-preta*, utiliza uma abordagem experimental, onde os parâmetros de um modelo são ajustados aos dados coletados do processo. A terceira chamada de *identificação de sistemas caixa-cinza*, une as duas abordagens anteriores; a estrutura do modelo é escolhida utilizando leis da física, e os parâmetros do modelo são ajustados a partir de dados coletados em um experimento.

A identificação de sistemas pode ser vista como a interface entre o mundo real das aplicações e o mundo matemático dos modelos abstratos. Ela é baseada em três elementos: um conjunto de dados obtido do sistema em operação, uma classe de modelos candidatos e uma regra para escolher um modelo dentro da classe, baseada no conjunto de dados coletado. (LJUNG, 2010).

As técnicas de identificação por minimização do erro de predição são uma ampla família de métodos que são baseados na resolução de um problema de otimização (LJUNG, 2002). Nestes métodos, é escolhida como função custo a ser minimizada, o erro médio quadrático entre a predição do modelo e os dados coletados do sistema. Utilizando hipóteses razoáveis, estes métodos geram estimativas não-polarizadas para os parâmetros do modelo.

É amplamente sabido que para muitas escolhas das estruturas do modelo, a função custo possui um formato bastante complicado, o que dificulta a resolução do problema de otimização (LJUNG, 1987). Os artigos (ÅSTRÖM; SÖDERSTRÖM, 1974; SÖDERSTRÖM, 1975; SÖDERSTRÖM; STOICA, 1982) apresentam alguns casos específicos que a função custo possui apenas um mínimo, mas mesmo nestes casos não é fácil obter a solução ótima do problema de otimização, dada a complexidade da função custo.

Para resolver o problema de otimização, muitas vezes são utilizados algoritmos iterativos baseados no gradiente da função custo. Quando as condições iniciais destes algoritmos estão bastante próximas dos parâmetros ótimos, os algoritmos conseguem alcançar o mínimo global com bastante facilidade. Entretanto, nos casos em que as condições iniciais estão longes dos parâmetros ótimos, muitas vezes acontece a convergência para mínimos locais ou para o limite do domínio de busca, o que bastante é indesejável. Para evitar estes problemas, a literatura sugere que sejam utilizadas condições iniciais para os algoritmos que estejam o mais próximas possíveis do mínimo global. Para obter as condições iniciais são utilizados problemas de otimização secundários, mas estas técnicas falham inúmeras vezes mesmo com níveis de ruído bastante baixos (LJUNG, 2010).

Uma das contribuições inéditas deste trabalho é obtenção de condições suficientes

sobre a convergência dos algoritmos de otimização para o mínimo global do critério de otimização. Estas condições expressam como os parâmetros que o usuário pode modificar estão relacionadas com a convergência dos algoritmos. Foi observado que se as condições iniciais dos algoritmos pertencem a uma região em que a função custo é semi-convexa então os algoritmos convergem com mais facilidade para o mínimo global do critério. Foi também observado que o formato da função custo e a região em que esta função é semi-convexa estão intimamente relacionados aos dados coletados do sistema. Desta maneira, é proposto nesta tese que sejam modificados os sinais de entrada do experimento para tornar a função custo mais “comportada” evitando que existam mínimos locais próximos do mínimo global.

Nesta tese são propostas duas ferramentas para melhorar a convergência dos métodos de identificação por erro de predição, as quais também são contribuições inéditas. Estas ferramentas modificam o espectro dos sinais de entrada com o objetivo de modificar o formato da função custo e melhorar a convergência dos algoritmos de otimização. A primeira ferramenta é baseada no projeto de experimentos enquanto que a segunda é baseada na filtragem dos dados.

A escolha dos sinais de entrada de um experimento é realizada na etapa do projeto do experimento. Esta escolha é feita de maneira que os dados apresentem características desejadas (GEVERS; BOMBOIS, 2006). Características desejadas típicas são que os dados sejam tais que os modelos identificados tenham a melhor precisão possível para um certo nível de esforço do experimento (medido pela energia dos sinais de entrada-saída ou pela duração dos experimentos). Outra possibilidade consiste em escolher o sinal de menor energia que gera uma estimativa do modelo com um valor máximo de incerteza especificado. O projeto de experimentos foi inicialmente proposto como um problema de otimização não convexo e de dimensão infinita, o que torna sua solução impraticável. Na última década, Håkan Hjalmarsson<sup>1</sup> tem trabalhado intensivamente no problema de projeto de experimentos, obtendo resultados importantes que tornaram a solução do problema de otimização viável. A partir de uma parametrização finita e linear do espectro do sinal de entrada, o problema tornou-se convexo e então pode ser resolvido por meio de desigualdades matriciais lineares (LMIs) (LINDQVIST; HJALMARSSON, 2001; JANSSON; HJALMARSSON, 2005). A partir destes resultados foi possível estender o método para obter sinais adaptativos (GERENCSEK; HJALMARSSON, 2005), utilizar dados obtidos com o sistema operando em malha-fechada (HJALMARSSON, 2005; HJALMARSSON; JANSSON, 2008), projetar sinais para identificação de sistemas não-lineares (HJALMARSSON; MARTENSSON, 2007), estender para sistemas com múltiplas entradas e múltiplas saídas (BARENTHIN et al., 2008; ROJAS; HJALMARSSON; HILDEBRAND, 2009) e para o problema de filtragem  $H_2$  ótima.

A inclusão de restrições ao projeto de experimentos parece ser um caminho natural para melhorar a convergência dos algoritmos de identificação para os parâmetros ótimos. A segunda contribuição desta tese é a obtenção de restrições convexas, que podem ser utilizadas no contexto de projeto de experimentos, as quais expandem a região de atração do mínimo global da função custo. Neste trabalho são apresentados dois tipos diferentes de restrições para o projeto de experimento, ambas baseadas nas condições previamente obtidas sobre a convergência dos algoritmos.

Em muitas aplicações, os dados coletados do sistema são obtidos a partir de um experimento que não pode ser escolhido pelo usuário. Podemos citar aqui sistemas econômicos

---

<sup>1</sup>O professor Håkan Hjalmarsson foi o supervisor do estágio sanduíche realizado no Royal Institute of Technology - KTH na Suécia.

e biológicos dos quais existem séries históricas de dados. Como não existe liberdade para a escolha do experimento, o projeto de experimentos não pode ser utilizado. Neste caso, é necessário utilizar outra ferramenta para melhorar a convergência para o mínimo global da função custo. A terceira contribuição deste trabalho é um método não baseado em projeto de experimentos, que também melhora a convergência dos algoritmos para o mínimo global. Este método é baseado na filtragem dos dados obtidos de um experimento. Os dados filtrados possuem espectro distinto do espectro dos dados originais. Este trabalho apresenta uma forma de escolher os filtros para melhorar a convergência dos algoritmos de otimização.

A literatura já apresentava, no contexto de controle adaptativo e controle baseado em dados, a ideia de modificar os dados de um experimento para melhorar a convergência de algoritmos (RIEDLE; PRALY; KOKOTOVIC, 1986; LEE et al., 1995; BAZANELLA et al., 2008). Os projetos de controladores baseado em dados e adaptativo consistem na escolha dos parâmetros de um controlador que minimizam um critério diretamente relacionado com dados de entrada e saída do processo. O problema de otimização é bastante similar ao apresentado para identificação de sistemas, e sofre da mesma dificuldade na resolução do problema devido ao formato da função custo. Estas técnicas haviam sido profundamente estudadas em (ECKHARD; BAZANELLA, 2009, 2010, 2011a, 2012) resultando em métodos para melhorar a convergência dos algoritmos para os parâmetros ótimos para os controladores. Este estudo resultou na publicação do livro (BAZANELLA; CAMPESTRINI; ECKHARD, 2012) no qual são apresentadas técnicas para o projeto de controladores que são muito similares às técnicas utilizadas para a identificação de sistemas. Nesta tese estas técnicas são adaptadas e melhoradas para a utilização no contexto de identificação de sistemas.

Esta tese é organizada da seguinte forma. O Capítulo 2 apresenta o método de identificação por minimização do erro de predição, suas propriedades e características. No Capítulo 3 são introduzidos os algoritmos de otimização e é feita uma análise da estabilidade destes algoritmos utilizando a teoria de sistemas não-lineares. O Capítulo 4 apresenta condições suficientes sobre a convergência dos algoritmos de otimização. Estas condições expressam como a convergência dos algoritmos está relacionada com o espectro dos sinais utilizados na identificação dos sistemas. O projeto de experimentos é descrito no Capítulo 5, onde são apresentados métodos e restrições utilizadas para a escolha do espectro do sinal de entrada. O Capítulo 6 descreve dois tipos de restrições para o projeto de experimentos que melhoram a convergência dos algoritmos de otimização para o mínimo global do critério de otimização. Finalmente, no Capítulo 7 é apresentado o método baseado na filtragem dos dados, que pode ser utilizado sem a necessidade de realizar um experimento específico no sistema, o qual também melhora a convergência dos algoritmos de otimização. A conclusão e trabalhos futuros são descritos no Capítulo 8.

## 2 IDENTIFICAÇÃO DE SISTEMAS

Um modelo é a descrição de algumas propriedades do sistema que são úteis por algum motivo. A identificação de sistemas consiste em métodos sistemáticos para obtenção desses modelos.

A *identificação de sistemas* se diferencia da *modelagem matemática*, pois a modelagem matemática se baseia em um profundo conhecimento das relações matemáticas que descrevem os fenômenos envolvidos no processo do qual se deseja obter o modelo, enquanto que a identificação de sistemas não utiliza este conhecimento prévio. A identificação de sistemas é baseada em três elementos: um sistema do qual pode-se coletar dados de entrada e saída, um conjunto de modelos e um critério para a escolha do modelo dentro do conjunto que melhor representa o sistema.

### 2.1 O sistema

O principal elemento da identificação de sistemas consiste no sistema do qual se deseja obter um modelo matemático. Neste trabalho assume-se que este processo é linear e invariante no tempo, que possui apenas uma entrada e uma saída e que pode ser descrito pela seguinte relação:

$$\mathcal{S} : \quad y(t) = G_0(q)u(t) + H_0(q)e_0(t). \quad (2.1)$$

O sinal  $u(t)$  é a entrada, o sinal  $y(t)$  é a saída e o sinal  $e_0(t)$  é ruído branco com variância  $\lambda_0$ . Além disso,  $G_0(q)$  e  $H_0(q)$  são funções de transferência racionais, próprias e estáveis que descrevem a dinâmica do processo. Aqui foi introduzido o operador de avanço  $q$

$$qu(t) = u(t + 1).$$

Assume-se que  $G_0(\infty) = 0$ , ou seja, que o sistema possui pelo menos 1 atraso de tempo entre sua entrada e saída. Sem perda de generalidade considera-se que  $H_0(\infty) = 1$ , ou seja,  $H_0(q)$  é mônica e portanto a resposta impulsiva  $h_0(t)$  do filtro  $H_0(q)$  satisfaz  $h_0(0) = 1$ . Além disso o filtro  $H_0(q)$  é de fase mínima.

O sinal  $u(t)$  é *quasi-estacionário* (LJUNG, 1987), ou seja,  $\exists C$  tal que

- $|E[u(t)]| \leq C \quad \forall t$
- $|E[u(t)u(r)]| \leq C \quad \forall t, r$
- $\bar{E}[u(t)u(t - \tau)] = R_u(\tau)$

onde

$$\bar{E}[\cdot] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N E[\cdot]$$

e  $E[\cdot]$  denota o *valor esperado*.

O espectro de  $u(t)$  é definido como

$$\Phi_u(\omega) = \sum_{\tau=-\infty}^{\infty} R_u(\tau) e^{-j\tau\omega}.$$

Assume-se também que os sinais  $u(t)$  e  $e_0(t)$  respeitam

$$\bar{E}[u(t)e_0(s)] = 0 \quad \forall t, s,$$

o que ocorre quando os dados são coletados em malha-aberta.

Os dados de entrada e saída de um experimento realizado no processo são coletados, armazenados e formam o seguinte conjunto

$$\mathcal{Z}^N = \{u(1), y(1), u(2), y(2), u(3), y(3), \dots, u(N), y(N)\}$$

onde  $N$  é a quantidade de dados coletados do processo.

Deseja-se com a identificação de sistemas obter um modelo matemático que descreva a dinâmica do processo  $\mathcal{S}$ . Para tanto, são utilizados os dados coletados, visto que eles contêm informações sobre esta dinâmica.

Existem vários métodos diferentes para obter um modelo para o processo, e estes métodos podem ser divididos em dois grandes grupos: os métodos paramétricos e os métodos não-paramétricos. Os métodos não-paramétricos resultam em uma representação gráfica que caracteriza a dinâmica do sistema, como a resposta impulsiva ou a resposta em frequência. Já os métodos paramétricos resultam em um modelo matemático que pode ser descrito por funções de transferência. Este trabalho será focado nos métodos paramétricos para estimação de modelos matemáticos.

## 2.2 O modelo de predição

Um *modelo probabilístico completo* (LJUNG, 1987) pode ser especificado por

$$y(t, \theta) = G(q, \theta)u(t) + H(q, \theta)e(t) \quad (2.2)$$

onde  $G(q, \theta)$  e  $H(q, \theta)$  descrevem a dinâmica do modelo e  $e(t)$  é ruído branco com variância  $\lambda$ . Logo, para caracterizar um modelo probabilístico completo deve-se especificar  $G(q, \theta)$ ,  $H(q, \theta)$  e  $\lambda$ .

Note que as funções de transferência foram parametrizadas em  $\theta \in \mathbb{R}^n$ . Este vetor  $\theta$  contém todos os parâmetros desconhecidos das funções de transferência, os quais se deseja estimar.

Uma das principais aplicações de um modelo é prever a saída futura do sistema. Para fazer a predição são usados valores presentes e passados dos sinais de entrada e saída. Um modelo que prediz a saída do sistema um passo a frente é caracterizado por

$$\hat{y}(t|t-1, \theta) = W_u(q, \theta)u(t) + W_y(q, \theta)y(t) \quad (2.3)$$

onde  $W_u(q, \theta)$  e  $W_y(q, \theta)$  são funções de transferência que também são parametrizadas por  $\theta$ .

Para simplificar a notação, o *modelo de predição* será escrito como

$$\hat{y}(t, \theta) = W(q, \theta)z(t) \quad (2.4)$$

onde

$$W(q, \theta) = [W_u(q, \theta) \quad W_y(q, \theta)] \quad \text{e} \quad z(t) = [u(t) \quad y(t)]^T.$$

Quando  $W(e^{j\omega}, \theta^1) = W(e^{j\omega}, \theta^2)$  para quase todo  $\omega$  vamos simplificar a notação e dizer que

$$W(e^{j\omega}, \theta^1) \equiv W(e^{j\omega}, \theta^2).$$

Um *modelo de predição ótimo* consiste em um sistema que prevê “da melhor maneira possível” a saída do processo no instante  $t$ , utilizando os sinais  $u(t)$  e  $y(t)$  apenas até o instante de tempo  $t - 1$ .

Note que não é possível estimar de maneira exata qual será o valor da saída no instante  $t$  pois este sinal é corrompido por ruído com característica estocástica. O melhor preditor é definido como aquele que minimiza a variância da diferença entre a saída do processo e a predição do modelo.

O processo pode ser reescrito como (SÖDERSTRÖM; STOICA, 1989):

$$y(t) = G_0(q)u(t) + (H_0(q) - 1)e_0(t) + e_0(t). \quad (2.5)$$

Isolando  $e_0(t)$  em (2.1), substituindo em (2.5) e reorganizando os termos obtém-se

$$\begin{aligned} y(t) &= [H_0^{-1}(q)G_0(q)u(t) + (1 - H_0^{-1}(q))y(t)] + e_0(t) \\ y(t) &= \phi(t) + e_0(t). \end{aligned}$$

Seja  $y^*(t)$  um preditor arbitrário, que utiliza os sinais de entrada e saída até o instante  $t - 1$  para estimar a saída do processo. Lembre que o preditor ótimo é aquele que minimiza a variância entre a saída e a predição. A variância do preditor arbitrário é definida por

$$\text{Variância} = E [(y(t) - y^*(t))^2] = E [(\phi(t) + e_0(t) - y^*(t))^2].$$

Note que o sinal  $e_0(t)$  não é correlacionado com o sinal  $\phi(t)$  pois por definição  $H_0(q)$  é mônico. Note também que  $e_0(t)$  não é correlacionado com  $y^*(t)$  pois o preditor utiliza apenas dados da saída até o instante  $t - 1$ . Logo, pode-se escrever a variância do preditor como

$$\begin{aligned} \text{Variância} &= E [(\phi(t) - y^*(t))^2] + E [e_0(t)^2] \\ &= E [(\phi(t) - y^*(t))^2] + \lambda_0 \geq \lambda_0. \end{aligned}$$

Portanto a menor variância possível é obtida quando  $y^*(t) = \phi(t)$ , logo o *preditor ótimo* é dado por

$$\hat{y}(t) = H_0^{-1}(q)G_0(q)u(t) + (1 - H_0^{-1}(q))y(t). \quad (2.6)$$

Como  $H_0(q)$  é mônico, este preditor utiliza dados do sinal de saída apenas até o instante  $t - 1$ , o que está de acordo com as hipóteses utilizadas.

Uma propriedade deste *preditor ótimo* é que o erro de predição é exatamente o sinal do ruído

$$y(t) - \hat{y}(t) = e_0(t).$$

O único erro que este preditor comete, ao estimar o valor da saída do processo no instante  $t$ , é exatamente o ruído neste instante, cujo valor é impossível de prever.

Ao estimar um modelo cujo objetivo é prever a saída do sistema, deve-se procurar obter o modelo mais próximo possível do preditor ótimo. Comparando (2.4) com (2.6) observa-se que uma maneira natural de parametrizar o *modelo de predição* é dada por

$$W_u(q, \theta) = H^{-1}(q, \theta)G(q, \theta) \quad \text{e} \quad W_y(q, \theta) = 1 - H^{-1}(q, \theta).$$

### 2.3 A estrutura do modelo

Os métodos de identificação utilizam o conjunto de dados coletados  $\mathcal{Z}^N$  para escolher o modelo que melhor descreve o sistema. Para tanto, deve-se definir um *conjunto de modelos* dentre os quais o método deve escolher o melhor. Como o método deve fazer uma busca neste conjunto, é interessante parametrizar o modelo e então fazer a busca em um conjunto de parâmetros.

Formalmente, o conjunto de modelos é definido por

$$\mathcal{M} = \{W(q, \theta) \mid \theta \in \mathcal{D}\}$$

onde  $\theta$  representa os parâmetros que se deseja estimar e o conjunto  $\mathcal{D}$  especifica as restrições destes parâmetros. Um *modelo*  $W(q, \theta)$  dentro do conjunto é definido por

$$W(q, \theta) = [W_u(q, \theta) \quad W_y(q, \theta)].$$

A *estrutura do modelo* define a maneira como o vetor de parâmetros  $\theta$  afeta um modelo. Uma estrutura de modelo  $M(\theta)$  é um mapa diferenciável de um conjunto  $\mathcal{D}$  para um *conjunto de modelos*. Em notação matemática tem-se que

$$M(\theta) : \mathcal{D} \rightarrow \mathcal{M}$$

A estrutura do modelo é um elemento muito importante na identificação de sistemas pois ela relaciona os parâmetros que serão identificados com as funções de transferência que descrevem a dinâmica do modelo.

Diz-se que uma estrutura de modelo  $M(\theta)$  é *uniformemente estável* se os filtros  $W(q, \theta)$ ,  $\frac{\partial}{\partial \theta} W(q, \theta)$  e  $\frac{\partial^2}{\partial \theta^2} W(q, \theta)$  são BIBO-estáveis para todo  $\theta$  pertencente a  $\mathcal{D}$ .

É importante saber se o mapa  $M(\theta)$  é injetor ou não, ou seja, se um mesmo modelo  $W(q, \theta)$  pode ser mapeado por diferentes valores de  $\theta$ . Quando um certo modelo  $W(q, \theta^1)$  pode ser mapeado por apenas um certo  $\theta^1$  diz-se que  $M(\theta)$  é *globalmente identificável* em  $\theta^1$ . Uma definição formal é dada a seguir.

**Definição 2.1** Uma estrutura de modelo  $M(\theta)$  é *globalmente identificável em*  $\theta^1$  se

$$W(q, \theta^1) \equiv W(q, \theta^2)$$

para  $\theta^1, \theta^2 \in \mathcal{D}$  implica

$$\theta^1 = \theta^2.$$

◇

Quando a estrutura do modelo é escolhida de tal maneira que o processo pode exatamente ser descrito pelo modelo, diz-se que  $\mathcal{S} \in \mathcal{M}$ . Vamos formalizar esta hipótese.

**Hipótese 2.1** *O sistema  $\mathcal{S}$  pertence ao conjunto de modelos  $\mathcal{M}$  ( $\mathcal{S} \in \mathcal{M}$ ), isto é, existe  $\theta^0 \in \mathcal{D}$  tal que*

$$G(q, \theta^0) = G_0(q) \text{ e } H(q, \theta^0) = H_0(q).$$

◆

Vamos definir aqui também o conjunto dos parâmetros que verificam a Hipótese 2.1.

$$\mathcal{D}_T = \{\theta \in \mathcal{D} | G(q, \theta) = G_0(q); H(q, \theta) = H_0(q)\}.$$

Quando  $\mathcal{S} \in \mathcal{M}$  o conjunto  $\mathcal{D}_T$  possui um ou mais elementos. Se além disso, a estrutura de modelo  $M(\theta)$  for globalmente identificável em  $\theta^0$ , então o único elemento no conjunto é  $\theta^0$ , ou seja,

$$\mathcal{D}_T = \{\theta^0\}.$$

A escolha da estrutura do modelo é uma etapa crítica pois afeta diretamente a qualidade do modelo obtido. Deseja-se muitas vezes que a estrutura do modelo possa descrever o sistema, e que  $M(\theta)$  seja globalmente identificável em  $\theta^0$ , mas fazer a escolha da estrutura muitas vezes não é uma tarefa fácil. Como  $\mathcal{S}$  é desconhecido pelo usuário, esta escolha tipicamente envolve testes com diferentes estruturas  $M(\theta)$ . Avalia-se os modelos obtidos com diferentes estruturas e escolhe-se aquele que melhor representa o sistema.

## 2.4 O método da minimização do erro de predição

Os métodos de identificação de sistemas consistem em regras para determinar qual o modelo  $W(q, \theta)$  do conjunto  $\mathcal{M}$  que melhor representa o sistema  $\mathcal{S}$  baseando-se no conjunto de dados coletados do processo  $\mathcal{Z}^N$ . Como o modelo  $W(q, \theta)$  é parametrizado por  $\theta$ , um método é um mapa que leva do conjunto de dados  $\mathcal{Z}^N$  para um  $\hat{\theta}_N$  do conjunto  $\mathcal{D}$

$$\mathcal{Z}^N \rightarrow \hat{\theta}_N \in \mathcal{D}$$

onde  $\hat{\theta}_N$  é a estimativa dos parâmetros obtida pelo método.

A diferença entre cada método de identificação está na maneira em que  $\hat{\theta}_N$  é estimado. O método da *minimização do erro de predição* procura pelos parâmetros do modelo que minimizam o erro entre a saída do processo e o modelo de predição.

O *erro de predição* é definido como

$$\varepsilon(t, \theta) \triangleq y(t) - \hat{y}(t, \theta) \tag{2.7}$$

$$= H^{-1}(q, \theta) [y(t) - G(q, \theta)u(t)], \tag{2.8}$$

e a estimativa dos parâmetros do modelo é obtida como a solução do seguinte problema de otimização:

$$\hat{\theta}_N = \arg \min_{\theta \in \mathcal{D}} V_N(\theta, \mathcal{Z}^N) \tag{2.9}$$

$$V_N(\theta, \mathcal{Z}^N) = \frac{1}{N} \sum_{t=1}^N \varepsilon^2(t, \theta). \tag{2.10}$$

Note que a função custo depende dos parâmetros do modelo  $\theta$  e do conjunto de dados  $\mathcal{Z}^N$  coletado do processo.

### 2.4.1 Propriedades assintóticas da estimativa

A estimativa do modelo sempre vai conter erros devidos à quantidade limitada de dados e ao ruído presente no sinal de saída. Neste caso é comum realizar uma análise *assintótica*, a qual considera que a quantidade de dados tende ao infinito.

Considere que o conjunto de dados  $\mathcal{Z}^\infty$  é obtido sob as mesmas condições do conjunto  $\mathcal{Z}^N$  quando  $N \rightarrow \infty$ . Diz-se que o conjunto  $\mathcal{Z}^\infty$  é informativo se ele permite distinguir entre dois modelos distintos.

**Definição 2.2** *Um conjunto de dados quasi-estacionário  $\mathcal{Z}^\infty$  é suficientemente informativo com respeito ao conjunto de modelos  $\mathcal{M}$  se, para dois modelos  $W(q, \theta^1)$  e  $W(q, \theta^2)$  nesse conjunto*

$$\bar{E} \left[ \left( (W(q, \theta^1) - W(q, \theta^2)) z(t) \right)^2 \right] = 0$$

implica  $W(q, \theta^1) \equiv W(q, \theta^2)$ . ◇

Vamos definir também a função limite  $\bar{V}(\theta)$  para a qual a função custo  $V_N(\theta, \mathcal{Z}^N)$  converge com probabilidade 1 quando  $N \rightarrow \infty$ :

$$\bar{V}(\theta) = \bar{E} [\varepsilon^2(t, \theta)].$$

Quando  $\varepsilon(t, \theta)$  é quasi-estacionário pode-se utilizar o teorema de Parseval (A.2) para escrever a função custo na forma integral

$$\begin{aligned} \bar{V}(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) \left| H^{-1}(e^{j\omega}, \theta) (G_0(e^{j\omega}) - G(e^{j\omega}, \theta)) \right|^2 d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_0 \left| H^{-1}(e^{j\omega}, \theta) H_0(e^{j\omega}) \right|^2 d\omega. \end{aligned}$$

onde  $\Phi_u(\omega)$  é o espectro de  $u(t)$  (LJUNG, 1987).

Pode acontecer da função  $\bar{V}(\theta)$  não possuir apenas um mínimo global. Neste caso pode-se definir o conjunto dos valores que minimizam a função como

$$\mathcal{D}_c = \arg \min_{\theta \in \mathcal{D}} \bar{V}(\theta). \quad (2.11)$$

Um dos principais resultados da teoria de identificação pela minimização do erro de predição pode então ser apresentado.

**Lema 2.1** (LJUNG, 1987) *Seja  $\hat{\theta}_N$  definido por (2.9),  $\mathcal{D}_c$  definido por (2.11),  $M(\theta)$  uma estrutura de modelo uniformemente estável e o conjunto  $\mathcal{Z}^\infty$  definido anteriormente. Então*

$$\hat{\theta}_N \rightarrow \mathcal{D}_c, \text{ com probabilidade 1 quando } N \rightarrow \infty.$$

□

Outro importante resultado é dado a seguir.

**Lema 2.2** (LJUNG, 1987) *Seja  $\mathcal{Z}^\infty$  um conjunto infinito de dados como definido anteriormente,  $M(\theta)$  uma estrutura de modelo uniformemente estável tal que  $S \in \mathcal{M}$ . Assuma que  $\mathcal{Z}^\infty$  é suficientemente informativo com relação a  $M(\theta)$ . Então*

$$\mathcal{D}_c = \mathcal{D}_T.$$

Se além disso  $M(\theta)$  for globalmente identificável em  $\theta^0 \in \mathcal{D}_T$ , então

$$\mathcal{D}_c = \{\theta^0\}.$$

□

Quando estes dois importantes resultados são aplicados conjuntamente temos que

$$G(q, \hat{\theta}_N) \rightarrow G_0(q) \quad \text{e} \quad H(q, \hat{\theta}_N) \rightarrow H_0(q)$$

com probabilidade 1 quando  $N \rightarrow \infty$ , e então dizemos que o estimador é *consistente*. Além disso, neste caso pode-se mostrar (LJUNG, 1987) que quando  $N \rightarrow \infty$

$$\sqrt{N}(\hat{\theta}_N - \theta^0) \rightarrow \mathcal{N}(0, P_\theta) \quad (2.12)$$

onde

$$P_\theta = \lambda_0 \left( \bar{E} [\psi(t, \theta^0) \psi^T(t, \theta^0)] \right)^{-1} \quad (2.13)$$

$$\psi(t, \theta^0) = \frac{\partial}{\partial \theta} \hat{y}(t, \theta) |_{\theta=\theta^0}.$$

O termo  $\mathcal{N}(0, P_\theta)$  representa uma distribuição normal com média 0 e a matriz  $P_\theta$  é então a matriz de covariância da distribuição assintótica.

Da estimativa assintótica normal dos parâmetros, segue que a variável aleatória

$$(\hat{\theta}_N - \theta^0)^T \left( \frac{P_\theta}{N} \right)^{-1} (\hat{\theta}_N - \theta^0) \quad (2.14)$$

tende a uma distribuição Qui-quadrado quando  $N \rightarrow \infty$ . Portanto, o parâmetro  $\hat{\theta}_N$  pertence assintoticamente ao conjunto

$$U_\theta = \left\{ \theta \mid (\hat{\theta}_N - \theta^0)^T \left( \frac{P_\theta}{N} \right)^{-1} (\hat{\theta}_N - \theta^0) \leq \chi_\alpha^2 \right\} \quad (2.15)$$

com probabilidade  $\alpha$ , onde  $\chi_\alpha^2$  é o valor da distribuição de probabilidade para o nível  $\alpha$  de confiança. A Tabela 2.1 mostra alguns valores de  $\chi_\alpha^2$  para alguns níveis de confiança  $\alpha$  e diferentes graus de liberdade, ou seja, quantidade de parâmetros que serão identificados  $n$  (SPIEGEL, 1992).

Tabela 2.1: Distribuição  $\chi_\alpha^2$  para alguns níveis de confiança  $\alpha$  e diferentes graus de liberdade  $n$ .

$n$	$\alpha$	90%	95%	97,5%	99%	99,5%
1		2,7055	3,8415	5,0239	6,6349	7,8794
2		4,6052	5,9915	7,3778	9,2104	10,5965
3		6,2514	7,8147	9,3484	11,3449	12,8381
4		7,7794	9,4877	11,1433	13,2767	14,8602
5		9,2363	11,0705	12,8325	15,0863	16,7496
6		10,6446	12,5916	14,4494	16,8119	18,5475
7		12,0170	14,0671	16,0128	18,4753	20,2777

Deste resultado podemos chegar a uma aproximação bastante útil para a covariância de  $\hat{\theta}_N$ :

$$Cov \left[ \hat{\theta}_N \right] \approx \frac{P_\theta}{N}$$

quando  $N$  é um valor elevado. Esta aproximação revela que o erro da estimativa decai na proporção de  $1/N$ , logo estimativas melhores são obtidas quando são coletados mais dados.

Contudo, além do tamanho do experimento  $N$ , o espectro do sinal de entrada também influencia a covariância da estimativa. A maneira pela qual o espectro atua em  $P_\theta$  é dada no próximo lema.

**Lema 2.3** O inverso da matriz de covariância,  $P_\theta^{-1}$  é dado por

$$P_\theta^{-1} = \frac{1}{2\pi\lambda_0} \int_{-\pi}^{\pi} \mathcal{F}_u(e^{j\omega}, \theta_0) \Phi_u(\omega) \mathcal{F}_u^*(e^{j\omega}, \theta_0) d\omega + R_0 \quad (2.16)$$

onde

$$R_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathcal{F}_e(e^{j\omega}, \theta_0) \mathcal{F}_e^*(e^{j\omega}, \theta_0) d\omega$$

e

$$\begin{aligned} \mathcal{F}_u(e^{j\omega}, \theta_0) &= H_0^{-1}(e^{j\omega}) \left. \frac{\partial G(e^{j\omega}, \theta)}{\partial \theta} \right|_{\theta=\theta_0} \\ \mathcal{F}_e(e^{j\omega}, \theta_0) &= H_0^{-1}(e^{j\omega}) \left. \frac{\partial H(e^{j\omega}, \theta)}{\partial \theta} \right|_{\theta=\theta_0}. \end{aligned}$$

□

*Prova:* Insira (2.6) em (2.13) e utilize o teorema de Parseval (A.2) para obter a expressão na forma integral. ■

A matriz  $P_\theta$  é uma medida do tamanho dos erros nos parâmetros. O Lema 2.3 mostra como espectro  $\Phi_u(\omega)$  afeta a matriz  $P_\theta$ .

#### 2.4.2 Obtendo a estimativa

Foi mostrado na seção anterior que quando algumas condições são satisfeitas o método de identificação pela minimização do erro de predição é um estimador consistente para o sistema  $\mathcal{S}$ . Foi mostrado na equação (2.9), que o método consiste em resolver um problema de otimização, que procura pelo vetor de parâmetros  $\theta$ , o qual minimiza a função custo  $V_N(\theta, \mathcal{Z}^N)$ . A solução deste problema de otimização foi denotada  $\hat{\theta}_N$ .

Contudo, nada foi dito até agora sobre como resolver o problema de otimização. Para encontrar uma maneira de resolver este problema é importante estudar a função custo  $V_N(\theta, \mathcal{Z}^N)$  e o conjunto  $\mathcal{D}$ . Infelizmente, muitas vezes a função custo  $V_N(\theta, \mathcal{Z}^N)$  pode apresentar vários extremos<sup>1</sup>. Para dificultar ainda mais o problema, o conjunto  $\mathcal{D}$  também não é convexo, o que torna a solução deste problema de otimização ainda mais difícil.

O método de identificação pela minimização do erro de predição propõe um estimador com propriedades estatísticas muito boas, mas que só faz sentido se for possível encontrar  $\hat{\theta}_N$ . Todos os resultados assintóticos apresentados anteriormente dependem da correta solução de (2.9). Logo, é importante utilizar as ferramentas corretas na resolução do problema de otimização.

Tipicamente são utilizados algoritmos iterativos para tentar solucionar este problema de otimização. Estes algoritmos utilizam informações coletadas do problema para tentar a cada iteração obter uma melhor estimativa de  $\hat{\theta}_N$ . Após um certo número de iterações o algoritmo termina e resulta em uma estimativa para o mínimo global da função custo. Esta estimativa será denotada  $\bar{\theta}$ . Estes algoritmos funcionam muito bem em muitas classes de problemas, mas também é sabido que eles falham em muitas outras. O próximo exemplo apresenta um caso bastante simples em que os algoritmos utilizados costumam falhar.

**Exemplo 2.1** Considere a identificação do sistema  $\mathcal{S}$

$$G_0(q) = \frac{0,033157}{(q-0,7)(q-0,8)(q-0,9)} \quad H_0(q) = 1 \quad (2.17)$$

<sup>1</sup>Uma definição formal de extremos será apresentada no próximo capítulo.

onde o ruído possui variância  $\lambda_0 = 0, 1$ .

Foi escolhida a seguinte estrutura de modelo  $M(\theta)$

$$G(q, \theta) = \frac{\theta_1}{q^3 + \theta_2 q^2 + \theta_3 q + \theta_4} \quad H_0(q, \theta) = 1 \quad (2.18)$$

onde  $\theta = [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4]^T$ , tal que  $\mathcal{S} \in \mathcal{M}$ . Note que a Hipótese 2.1 é satisfeita.

O conjunto de dados  $\mathcal{Z}^N$  será coletado em um experimento onde o sinal de entrada  $u(t)$  consiste em  $N = 1.000$  amostras de ruído branco com variância 1 e o conjunto  $\mathcal{D}$  é tal que o modelo de predição deve ser estável.

Foi utilizado o comando `oe` do toolbox de identificação “System Identification Toolbox” do MATLAB (LJUNG, 2000) para estimar os parâmetros do modelo. A solução obtida foi a seguinte

$$\bar{\theta} = [-0,0005708 \ 1,5911 \ 0,2946 \ -0,36012]^T;$$

a qual gera a seguinte função de transferência

$$G(q, \bar{\theta}) = \frac{-0,0005708}{(q - 0,3607)(q^2 + 1,952q + 0,9985)}.$$

O toolbox retornou o valor  $\bar{\theta}$ , mas isto não significa que o mínimo global da função custo  $V_N(\theta, \mathcal{Z}^N)$  foi alcançado. Observe que a estimativa do modelo é bastante diferente do sistema real  $\mathcal{S}$ . O módulo do número complexo que representa o par de pólos complexos é 0,9992, ou seja, o algoritmo parou muito próximo da fronteira do espaço de busca  $\mathcal{D}$  e portanto não convergiu para o mínimo global  $\hat{\theta}_N$ .

A correta estimativa de  $\hat{\theta}_N$  depende de vários fatores dos quais podemos citar: o algoritmo utilizado, a função custo  $V_N(\theta, \mathcal{Z}^N)$  e o conjunto  $\mathcal{D}$ . Neste trabalho serão analisados diversos destes fatores que influenciam a resolução do problema de otimização (2.9) para entender como eles afetam a estimativa  $\hat{\theta}_N$ . Pretende-se então apresentar novas ferramentas que auxiliarão a resolução do problema de otimização.

## 2.5 O método dos mínimos quadrados

Apesar do método da minimização do erro de predição apresentar propriedades estatísticas muito boas, ele depende da solução de um problema de otimização complexo. Na literatura são apresentados outros métodos para a identificação de sistemas cuja estimativa é mais simples de ser obtida. O método dos mínimos quadrados também é baseado na solução de um problema de otimização, mas utiliza uma função custo que é quadrática, cujo mínimo pode ser encontrado com eficiência.

Considere que o modelo do sistema é parametrizado da seguinte forma:

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + A^T(q)\theta} \quad H(q, \theta) = \frac{1}{1 + A^T(q)\theta} \quad (2.19)$$

onde os elementos de  $A(q) \in \mathbb{R}^n$  e  $B(q) \in \mathbb{R}^n$  são funções em  $q$ . Esta parametrização possui as seguintes propriedades. Primeiramente, tanto o numerador como o denominador de  $G(q, \theta)$  são funções afins em  $\theta$ . Além disso,  $G(q, \theta)$  e  $H(q, \theta)$  possuem o mesmo denominador.

Utilizando esta parametrização, o modelo preditor é dado por

$$\hat{y}(t, \theta) = B^T(q)\theta u(t) - A^T(q)\theta y(t).$$

Vamos definir o vetor regressor como

$$\varphi(t) = B(q)u(t) - A(q)y(t)$$

de maneira que o erro de predição é dado por

$$\varepsilon(t, \theta) = y(t) - \varphi^T(t)\theta.$$

Logo, a função custo torna-se

$$V_N(\theta, \mathcal{Z}^N) = \frac{1}{N} \sum_{t=1}^N (y(t) - \varphi^T(t)\theta)^2$$

a qual é quadrática em  $\theta$  e possui um mínimo que pode ser calculado por

$$\hat{\theta}_N^{MQ} = \left[ \frac{1}{N} \sum_{t=1}^N \varphi(t)\varphi^T(t) \right]^{-1} \left[ \frac{1}{N} \sum_{t=1}^N \varphi^T(t)y(t) \right]. \quad (2.20)$$

Nota-se que neste caso a solução do problema de otimização é facilmente obtida pela resolução de (2.20). Apesar da parametrização de  $G(q, \theta)$  ser bastante flexível em (2.19), a parametrização de  $H(q, \theta)$  não é. Logo, em poucos casos a hipótese  $\mathcal{S} \in \mathcal{M}$  será satisfeita, a qual foi fundamental na seção anterior para obter as características assintóticas do estimador.

O método dos mínimos quadrados propõe um estimador para os parâmetros do modelo que é fácil de ser calculado, mas em contrapartida, tal estimador não é consistente considerando um caso geral. Nos casos em que o sistema real é da forma (2.19), o estimador possui as mesmas propriedades do método baseado na minimização do erro de predição. Logo, este método pode ser então visto como um caso particular do método da minimização do erro de predição.

## 2.6 O método das variáveis instrumentais

Foi mostrado anteriormente que quando o preditor ótimo é considerado, então o erro de predição se torna ruído branco. Por consequência, um modelo bom é aquele cujo erro de predição é descorrelacionado dos valores passados da entrada e da saída do processo. O método das variáveis instrumentais utiliza tal raciocínio na identificação dos parâmetros: ele procura pelo modelo cujo erro de predição é menos correlacionado com uma *variável instrumental*. Esta variável é escolhida de maneira que seja correlacionada com os dados do processo, mas descorrelacionada do ruído.

O método propõe a seguinte função custo

$$f_N(\theta, \mathcal{Z}^N) = \frac{1}{N} \sum_{t=1}^N \zeta(t)\varepsilon(t, \theta)$$

onde  $\zeta(t)$  é chamada de variável instrumental e é escolhida de maneira a ser correlacionada com os dados do processo, mas não correlacionada com o ruído.

Considerando novamente a estrutura do modelo (2.19) temos que

$$f_N(\theta, \mathcal{Z}^N) = \frac{1}{N} \sum_{t=1}^N \zeta(t)y(t) - \frac{1}{N} \sum_{t=1}^N \zeta(t)\varphi(t)\theta.$$

A melhor estimativa para o vetor de parâmetros é aquela que faz com que o erro de predição seja descorrelacionado com a variável instrumental. Fazendo  $f_N(\theta, \mathcal{Z}^N) = 0$  temos que

$$\hat{\theta}_N^{VI} = \left[ \frac{1}{N} \sum_{t=1}^N \zeta(t)\varphi^T(t) \right]^{-1} \left[ \frac{1}{N} \sum_{t=1}^N \zeta^T(t)y(t) \right]. \quad (2.21)$$

Note que o estimador de variáveis instrumentais é bastante similar ao estimador de mínimos quadrados. Contudo, suas propriedades estatísticas são distintas. Antes de enunciar uma das principais propriedades deste método é necessário fazer algumas hipóteses.

**Hipótese 2.2** *A estrutura de modelo  $M(\theta)$  é escolhida de maneira que*

$$\exists \theta_G \in \mathcal{D} \mid G(q, \theta_G) = G_0(q).$$



Esta hipótese assume que é possível descrever a dinâmica entre a entrada  $u(t)$  e a saída  $y(t)$  do sistema com a estrutura do modelo escolhida. Contudo, não é necessário fazer hipótese sobre a dinâmica do ruído, diferentemente do que foi feito na Hipótese 2.1.

Agora podemos enunciar uma importante propriedade do método das variáveis instrumentais.

**Lema 2.4** (LJUNG, 1987) *Seja  $\hat{\theta}_N^{VI}$  definido por (2.21),  $M(\theta)$  uma estrutura de modelo uniformemente estável que respeita a Hipótese 2.2 e que é globalmente identificável em  $\theta_G$ ,  $\mathcal{Z}^\infty$  o conjunto definido anteriormente,  $\zeta(t)$  uma variável instrumental escolhida de maneira a não ser correlacionada com o ruído tal que  $\bar{E} [\zeta(t)\varphi^T(t)]$  não seja singular. Então*

$$G(q, \hat{\theta}_N^{VI}) \rightarrow G_0(q) \text{ com probabilidade 1 quando } N \rightarrow \infty.$$



Quando as condições do Lema 2.4 são satisfeitas o método das variáveis instrumentais é um estimador consistente de  $G_0(q)$ .

Este método apresenta duas desvantagens. Primeiramente, o método não estima a função de transferência do ruído  $H(q, \theta)$ . Em muitas aplicações é necessário o conhecimento do modelo do ruído, como é o caso do projeto de controladores de mínima variância. A segunda desvantagem está relacionada com a variância da estimativa  $\hat{\theta}_N^{VI}$ . Pode-se mostrar que a variância da estimativa obtida com o método das variáveis instrumentais é maior que a variância da estimativa obtida com o método da minimização do erro de predição.

Este método é largamente utilizado como condição inicial para os métodos baseados na minimização do erro de predição. O toolbox “System Identification Toolbox” do MATLAB é um dos aplicativos que utiliza esta técnica. Contudo, isto não é suficiente para garantir que os métodos alcancem o mínimo global  $\hat{\theta}_N$  da função custo. O Exemplo 2.1 mostrou um caso típico em que não foi possível obter uma boa estimativa para o modelo. Surge então a necessidade de entender melhor a convergência dos métodos para propor formas de melhorar a convergência.

## 2.7 Considerações finais

Neste capítulo foram apresentados conceitos básicos sobre o método de identificação por minimização do erro de predição. O método consiste na solução de um problema de otimização cujo critério está relacionado ao erro do modelo de predição.

Devido ao ruído e à quantidade limitada de dados os parâmetros estimados sempre contêm erros. Uma análise das propriedades estatísticas deste estimador foi apresentada considerando que a quantidade de dados coletados tende ao infinito.

Foi apontado que em muitas casos não é fácil resolver o problema de otimização. Algoritmos iterativos são muitas vezes escolhidos para resolver o problema, mas sabe-se que não é incomum a falha em encontrar o mínimo global da função custo.

Apresentou-se também outros dois métodos: o método dos mínimos quadrados e o método das variáveis instrumentais. A estimativa deste métodos pode ser computada de maneira fechada, o que é uma grande vantagem em relação ao método da minimização do erro de predição. Contudo, as propriedades estatísticas destes métodos restringem suas aplicações. Os dois métodos são largamente empregados para obter uma estimativa aproximada do modelo o qual será utilizada como condição inicial do método de identificação pela minimização do erro de predição.

Pretende-se neste trabalho estudar o método da minimização do erro de predição, os algoritmos, a função custo e o domínio de busca para entender como afetam a solução do problema de otimização. Pretende-se então apresentar novas ferramentas que auxiliem na tarefa de determinar  $\hat{\theta}_N$ , o mínimo global do critério utilizado.

## 3 OTIMIZAÇÃO

Neste capítulo serão introduzidos conceitos centrais sobre otimização, os quais são úteis para a identificação de sistemas. Serão apresentados conceitos básicos de cálculo multivariável, conceitos da teoria de sistemas não-lineares e da estabilidade de sistemas discretos utilizando a teoria de Lyapunov. Desta forma vamos fixar a nomenclatura utilizada no trabalho e introduzir conceitos necessários para a análise dos algoritmos de otimização. Uma análise da convergência de dois algoritmos fundamentais será feita: do método do gradiente e do método de Newton-Raphson. Em seguida será apresentado um resultado sobre a robustez dos algoritmos utilizados. Por fim serão feitas considerações sobre a escolha de que algoritmo é mais indicado em cada situação.

### 3.1 Definições

Seja  $V(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^+$  uma função analítica no domínio  $\mathcal{D} \subseteq \mathbb{R}^n$ . O gradiente desta função é um vetor coluna definido como

$$\nabla V(\theta) \triangleq \frac{\partial V(\theta)}{\partial \theta}.$$

A Hessiana desta função é definida como a seguinte matriz simétrica

$$\nabla^2 V(\theta) \triangleq \frac{\partial^2 V(\theta)}{\partial \theta^2}.$$

A Hessiana é simétrica e portanto seus autovalores são reais. Quando  $\nabla^2 V(\theta) > 0$  diz-se que a matriz Hessiana é definida positiva.

#### Definição 3.1 Extremo

Um ponto  $\theta^* \in \mathcal{D} \subseteq \mathbb{R}^n$  é um **mínimo** de  $V(\cdot)$  se

$$\begin{aligned} \exists \epsilon > 0 \mid V(\theta) \geq V(\theta^*) \quad \forall \theta \in \mathcal{B}_\epsilon(\theta^*) \\ \mathcal{B}_\epsilon(\theta^*) = \{ \theta \mid (\theta - \theta^*)^T (\theta - \theta^*) < \epsilon \} \end{aligned}$$

Ele é um **máximo** de  $V(\cdot)$  se

$$\exists \epsilon > 0 \mid V(\theta) \leq V(\theta^*) \quad \forall \theta \in \mathcal{B}_\epsilon(\theta^*).$$

Ele é um **extremo** se ele é um mínimo ou um máximo.

Um mínimo  $\theta^*$  é dito<sup>1</sup>

---

<sup>1</sup>Definições similares existem para os máximos.

- um **mínimo isolado** se  $\exists \epsilon > 0$  tal que  $V(\theta) > V(\theta^*) \quad \forall \theta \in \mathcal{B}_\epsilon(\theta^*)$ ;
- um **mínimo global** em  $\mathcal{D}$  se  $V(\theta) \geq V(\theta^*) \quad \forall \theta \in \mathcal{D}$ ;
- o **mínimo global único** em  $\mathcal{D}$  se  $V(\theta) > V(\theta^*) \quad \forall \theta \in \mathcal{D}$ ;
- um **mínimo local** se é um mínimo mas não é um mínimo global.

◇

Vamos também definir:

**Definição 3.2** Um dado ponto  $\theta^*$  é **ponto crítico** de  $V(\cdot)$  se  $\nabla V(\theta^*) = \mathbf{0}$ .

◇

Uma propriedade dos extremos é

**Lema 3.1** Qualquer extremo é um ponto crítico.

□

*Prova:* Considere que  $\theta^*$  não é um ponto crítico de  $V(\cdot)$  – ou seja,  $\nabla V(\theta^*) \neq \mathbf{0}$ . Vamos provar que  $\theta^*$  não pode ser um extremo.

Seja  $\mathbf{d}$  um vetor unitário tal que  $\mathbf{d}^T \nabla V(\theta^*) \neq \mathbf{0}$ . Como  $V(\cdot)$  é analítica e  $\mathbf{d}$  é unitário, podemos escrever utilizando a Série de Taylor

$$V(\theta^* + \delta \mathbf{d}) - V(\theta^*) = \nabla V(\theta^*) \delta \mathbf{d} + \text{hot}(\delta)$$

para qualquer  $\delta \in \mathbb{R}$ , onde  $\text{hot}(\delta)$  representa os termos de alta ordem, neste caso sendo os termos de ordem maior que um, o que significa que

$$\lim_{\delta \rightarrow 0} \frac{\text{hot}(\delta)}{\delta} = 0.$$

Logo,  $\exists \delta_0$  tal que  $\|\nabla V(\theta^*) \delta \mathbf{d}\| > \|\text{hot}(\delta)\|$  para todo  $\delta$  tal que  $|\delta| < \delta_0$ . Portanto, para  $|\delta| < \delta_0$ , o sinal de  $V(\theta^* + \delta \mathbf{d}) - V(\theta^*)$  é determinado somente pelo produto  $\nabla V(\theta^*) \delta \mathbf{d}$ :

$$\text{sign}[V(\theta^* + \delta \mathbf{d}) - V(\theta^*)] = \text{sign}[\nabla V(\theta^*) \delta \mathbf{d}] \quad \forall \delta : |\delta| < \delta_0.$$

Fica claro que quando  $\delta$  cruza o zero,  $V(\theta^* + \delta \mathbf{d}) - V(\theta^*)$  muda de sinal. Então existem pontos arbitrariamente próximos de  $\theta^*$  para os quais  $V(\theta) > V(\theta^*)$  (o que implica que  $\theta^*$  não é um máximo) e outros pontos tais que  $V(\theta) < V(\theta^*)$  (o que implica que  $\theta^*$  não é um mínimo). Portanto,  $\theta^*$  não é um extremo. ■

É importante notar que esta condição é necessária mas não é suficiente, ou seja, nem todos os pontos críticos são extremos. Pontos de inflexão também são pontos críticos, mas não são extremos.

O Lema 3.1 estabelece que o gradiente é sempre zero nos máximos e mínimos. O próximo lema vai apresentar outras propriedades dos mínimos.

**Lema 3.2** Seja  $\theta^*$  um ponto crítico de  $V(\cdot)$ . Então os seguintes fatos são verdade.

- $\theta^*$  é um mínimo isolado de  $V(\cdot)$  se  $\nabla^2 V(\theta^*) > 0$ ;
- $\theta^*$  é um máximo isolado de  $V(\cdot)$  se  $\nabla^2 V(\theta^*) < 0$ .

□

*Prova:* O primeiro item será provado, enquanto que a prova do segundo é a mesma *mutatis mutandis*. Como  $\nabla V(\theta^*) = \mathbf{0}$  podemos escrever

$$V(\theta^* + \delta \mathbf{d}) - V(\theta^*) = \frac{\delta^2 \mathbf{d}^T \nabla^2 V(\theta^*) \mathbf{d}}{2} + \text{hot}(\delta)$$

onde  $\mathbf{d}$  é um vetor unitário e  $\text{hot}(\delta)$  agora satisfaz

$$\lim_{\delta \rightarrow 0} \frac{\text{hot}(\delta)}{\delta^2} = 0.$$

Pelo mesmo argumento do Lema 3.1, o sinal de  $V(\theta^* + \delta \mathbf{d}) - V(\theta^*)$  é determinado exclusivamente pelo primeiro termo da expressão para valor suficientemente pequenos de  $\delta$ , ou seja:

$$\text{ sinal}[V(\theta^* + \delta \mathbf{d}) - V(\theta^*)] = \text{ sinal}[\mathbf{d}^T \nabla^2 V(\theta^*) \mathbf{d}] \quad \forall \delta : |\delta| < \delta_0 \quad (3.1)$$

para algum  $\delta_0 > 0$  e para todo vetor unitário  $\mathbf{d}$ . Se a Hessiana é positiva definida, então o lado direito de (3.1) é sempre positivo, e portanto

$$V(\theta^* + \delta \mathbf{d}) > V(\theta^*) \quad \forall \delta : |\delta| < \delta_0$$

Ou seja,  $\theta^*$  é um mínimo isolado. ■

Um fato óbvio derivado do teorema acima é que uma função analítica sempre apresenta um comportamento convexo na vizinhança de um mínimo isolado.

**Definição 3.3** Uma função  $V(\cdot)$  é dita convexa no domínio  $\Omega$  se  $\nabla^2 V(\theta) > 0 \quad \forall \theta \in \Omega$  e então o conjunto  $\Omega$  é chamado de um **domínio de convexidade**. Ele é dito côncavo em  $\Omega$  se  $\nabla^2 V(\theta) < 0 \quad \forall \theta \in \Omega$  e então o conjunto  $\Omega$  é chamado de um **domínio de concavidade**. ◇

Dada esta nomenclatura, fica claro da definição de extremo que existe um conjunto de convexidade não vazio em torno de qualquer mínimo isolado de uma função analítica  $V(\cdot)$ . Similarmente, existe um conjunto de concavidade não vazio em torno de qualquer máximo isolado de uma função analítica  $V(\cdot)$ .

## 3.2 Algoritmos de otimização

Seja  $V(\theta) : \mathbb{R}^n \rightarrow \mathbb{R}^+$  uma função custo diferenciável com um mínimo global  $\theta^*$ , o qual deseja-se encontrar. Um algoritmo de otimização iterativo é uma recursão, o qual gera sucessivas aproximações para o mínimo global de  $V(\theta)$ :

$$\theta^{i+1} = f(\theta^i) \quad (3.2)$$

onde  $f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Note que a cada iteração a função  $f(\cdot)$  é calculada na aproximação atual  $\theta^i$ , ou seja, somente a informação “local” (informação na aproximação atual) é utilizada na otimização.

Para o algoritmo ser útil, a função  $f(\cdot)$  deve, enquanto as iterações passam, gerar aproximações  $\theta^i$  que tendem a um mínimo, eventualmente convergindo para ele. Preferencialmente, o algoritmo deve convergir para um mínimo global. Portanto, o estudo desses algoritmos e da suas convergências para um mínimo global são de maior relevância.

A equação (3.2) que define o algoritmo, nada mais é que um sistema em tempo discreto, onde  $\theta^i$  é o estado do sistema no “tempo” (iteração)  $i$ . Parece razoável então tirar proveito deste fato e utilizar a teoria clássica de estabilidade de sistemas para estudar a convergência dos algoritmos. O problema de convergência será abordado considerando que o algoritmo é um sistema dinâmico não-linear, e que a convergência depende da estabilidade deste sistema. Vamos fazer então uma breve revisão sobre a teoria de sistemas não-lineares.

**Definição 3.4** *Ponto fixo.*

Um vetor  $\theta^*$  é chamado um **ponto fixo** de um algoritmo (3.2) se  $f(\theta^*) = \theta^*$ . Um ponto fixo é atrativo (ou assintoticamente estável) se o algoritmo converge para este ponto para inicializações suficientemente próximas ao mesmo, ou seja, se existe um escalar positivo  $\epsilon$  tal que para todo  $\theta^1 \in \mathcal{B}_\epsilon(\theta^*)$ ,  $\lim_{i \rightarrow \infty} \theta^i = \theta^*$ .  $\diamond$

O nome “ponto fixo” expressa o fato que para este algoritmo este ponto é fixo. Em outras palavras, uma vez que o algoritmo encontra este ponto, ele não pode mais sair dele. Em termos formais

$$\theta^k = \theta^* \rightarrow \theta^i = \theta^* \quad \forall i > k.$$

Se os pontos fixos são atrativos ou não talvez seja o principal tema da análise de sistemas não-lineares. Este problema pode ser tratado pelos dois métodos de Lyapunov: o direto e o indireto. O método indireto envolve a linearização do sistema não-linear (3.2) no ponto fixo e é útil para analisar se o ponto fixo é atrativo ou não.

**Teorema 3.1** (KHALIL, 1996) *Método Indireto de Lyapunov*

Seja  $\theta^*$  um ponto fixo do algoritmo (3.2) onde  $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}^n$  e  $\theta^* \in \mathcal{D}$ . Seja

$$A = \left. \frac{\partial f(\theta)}{\partial \theta} \right|_{\theta=\theta^*}.$$

Então,

1. O ponto fixo  $\theta^*$  é **atrativo** se todos os autovalores de  $A$  pertencem ao interior do círculo unitário;
2. O ponto fixo  $\theta^*$  é **instável** se algum autovalor de  $A$  pertence ao exterior do círculo unitário.

$\square$

Contudo, é também importante analisar para qual conjunto de inicializações o algoritmo converge ao ponto fixo, o que não é coberto pelo método indireto. Este conjunto de inicializações é chamado de domínio de atração.

**Definição 3.5** *Seja  $\theta^*$  um ponto fixo atrativo do algoritmo  $\theta^{i+1} = f(\theta^i)$ . O conjunto  $\Omega \subset \mathbb{R}^n$  é um **domínio de atração** (DOA) de  $\theta^*$  se  $\lim_{i \rightarrow \infty} \theta^i = \theta^* \quad \forall \theta^1 \in \Omega$ .*  $\diamond$

Note que  $\Omega$  na definição **não** é o conjunto de **todas** as inicializações que convergem para o ponto fixo. Este conjunto pode ser bastante complexo e pode ser muito difícil caracterizá-lo de maneira significativa. Por outro lado, não é realmente importante conhecer todas as inicializações que convergem para o ponto fixo, mas apenas se estas inicializações formam um conjunto “suficientemente grande”, de maneira que o algoritmo converge para este ponto fixo mesmo quando inicializado “longe” dele. Neste trabalho vamos focar em conjuntos que são fáceis de caracterizar matematicamente, como é o caso dos elipsoides:

$$\mathcal{E} = \{ \theta : (\theta - \theta^*)^T P (\theta - \theta^*) < 1 \}, \quad (3.3)$$

onde  $P$  é uma matriz simétrica definida positiva. O elipsoide  $\mathcal{E}$  descrito em (3.3) é centrado em  $\theta^*$ , seu tamanho é determinado pelos autovalores de  $P$  e o formato depende dos autovetores. Quando  $P = aI_n$  para algum escalar positivo  $a$  o conjunto é uma hipersfera (o qual será chama de “bola” neste trabalho).

Agora é possível enunciar o teorema do método direto de Lyapunov.

**Teorema 3.2** (KHALIL, 1996) *Método Direto de Lyapunov*

Considere o algoritmo (3.2) com um ponto fixo  $\theta^* \in \Omega$ . O ponto fixo  $\theta^*$  é atrativo se existe uma função  $W(\cdot) : \Omega \rightarrow \mathbb{R}^+$  tal que:

$$W(\theta^*) = 0 \quad (3.4)$$

$$W(\theta) > 0 \quad \forall \theta : \theta \neq \theta^*, \quad \theta \in \Omega \quad (3.5)$$

$$W(f(\theta)) - W(\theta) < 0 \quad \forall \theta \in \Omega \quad (3.6)$$

□

Uma função  $W(\cdot)$  que satisfaz as condições do teorema é chamada de função de Lyapunov. Contudo encontrar tal função pode ser uma tarefa difícil. A principal utilidade do método direto é determinar domínios de atração para o ponto fixo, visto que os conjuntos de nível servem como domínios de atração. Os conjuntos de nível são definidos na sequência.

**Definição 3.6** *Conjuntos de nível*

Os conjuntos de nível da função de Lyapunov  $W(\cdot)$  são definidos como

$$\mathcal{L}_c \triangleq \{ \theta : W(\theta) \leq c \}$$

para um escalar  $c > 0$ . Note que por construção,  $\theta^* \in \mathcal{L}_c \quad \forall c$  (pois por definição  $W(\theta^*) = 0 < c$ ). ◇

A Figura 3.1 ilustra uma função de Lyapunov típica e seus conjuntos de nível. Um conjunto de nível  $\mathcal{L}_c$  é um domínio de atração de um ponto fixo  $\theta^*$  se  $\mathcal{L}_c \in \Omega$ , onde  $\Omega$  é conjunto do enunciado do Teorema 3.2. Para uma dada função de Lyapunov, o maior  $c$  para o qual esta inclusão é satisfeita proporciona o melhor (maior) domínio de atração que pode ser obtido com esta função de Lyapunov. Na próxima seção, estes conceitos serão utilizados para analisar os algoritmos de otimização mais comuns.

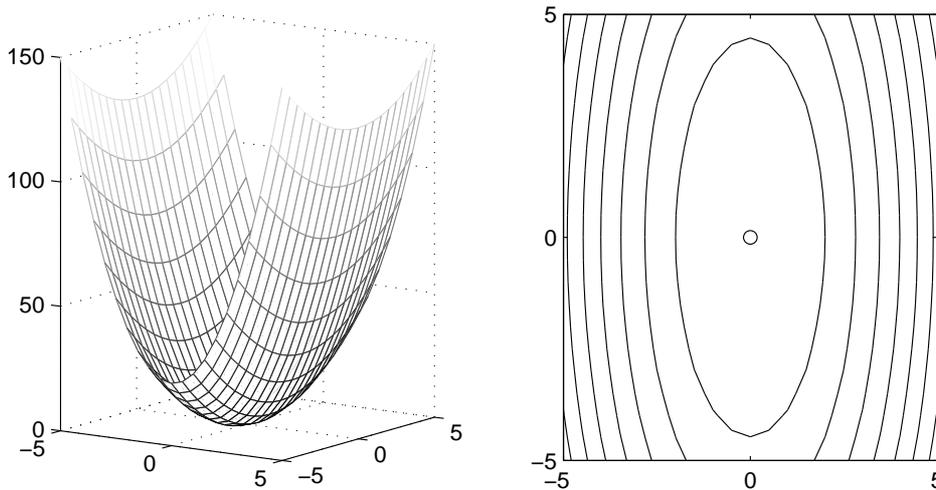


Figura 3.1: Uma função de Lyapunov e seus conjuntos de nível.

### 3.3 Alguns algoritmos e propriedades de convergências

Que informação está disponível no passo  $i$  de uma otimização? Conhecendo  $\theta^i$  é normalmente possível computar informações sobre a função custo em  $\theta^i$ : o valor, o gradiente, a Hessiana, etc. Tipicamente estas são as informações utilizadas pelos algoritmos de otimização. Vários algoritmos podem ser descritos por

$$f(\theta^i) = \theta^i - \gamma(\theta^i)R(\theta^i)\nabla V(\theta^i) \quad (3.7)$$

onde  $R(\theta^i)$  é uma matriz e  $\gamma(\theta^i)$  é um escalar positivo. A matriz  $R(\theta^i)$  define a direção da atualização dos parâmetros no passo  $i$  pois esta matriz projeta o gradiente  $\nabla V(\theta^i)$ . Já o escalar  $\gamma(\theta^i)$  define o tamanho da atualização a cada iteração. Cada escolha destes parâmetros implica diferentes implementações com diferentes propriedades. Algumas destas escolhas receberam um nome específico onde método do gradiente, Newton-Raphson e Gauss-Newton são as mais conhecidas. Vamos analisar cada um destes algoritmos separadamente.

#### 3.3.1 O método do gradiente

Um dos algoritmos de otimização mais simples é o método do gradiente. Para este método  $R(\theta) = I_n$  na equação (3.7), ou seja, a iteração é dada por

$$\theta^{i+1} = \theta^i - \gamma(\theta^i)\nabla V(\theta^i) \quad (3.8)$$

com  $\gamma(\theta^i) > 0 \forall \theta^i$ . O princípio do método é simples: as atualizações são feitas na direção oposta ao gradiente, então, para pequenos valores de  $\gamma(\theta^i)$ , a cada iteração um valor menor da função custo é encontrado. As propriedades deste método serão analisadas na sequência, onde vamos começar com a caracterização dos pontos fixos.

**Teorema 3.3** *Um dado ponto  $\theta^*$  no espaço de parâmetros é um ponto fixo do método do gradiente se e somente se ele for um ponto crítico da função custo  $V(\theta)$ .*  $\square$

*Prova:* Como  $\theta^*$  é um ponto fixo então  $f(\theta^*) = \theta^*$ . A partir de (3.8) fica claro que para o método do gradiente isto é equivalente a  $\nabla V(\theta^*) = \mathbf{0}$ .  $\blacksquare$

Então, todos os mínimos, máximos e pontos de inflexão são pontos fixos do algoritmo do gradiente. Mas quais destes pontos são atrativos? O próximo teorema apresenta algumas respostas.

**Teorema 3.4** *Considere o algoritmo (3.8) onde  $\nabla V(\theta)$  é o gradiente de uma função custo diferenciável. Então existe uma sequência de passos  $\gamma(\theta^i)$  tal que os mínimos da função custo sejam atrativos. Além disso, não existe uma sequência de passos  $\gamma(\theta^i)$  tal que os máximos da função custo sejam atrativos.  $\square$*

*Prova:*

*Vamos utilizar o método indireto de Lyapunov.*

$$A = \left. \frac{\partial f(\theta)}{\partial \theta} \right|_{\theta=\theta^*} = I_n - \gamma(\theta) \nabla^2 V(\theta) \Big|_{\theta=\theta^*}$$

*Considerando que para um ponto de mínimo isolado a Hessiana é positiva então para este ponto os autovalores de  $A$  pertencerão ao interior do círculo unitário se  $\gamma(\theta^i)$  for suficientemente pequeno. Logo, um mínimo isolado é atrativo apenas para uma escolha adequada de  $\gamma(\theta^i)$ .*

*Considerando que para um ponto de máximo isolado a Hessiana é negativa então para este ponto os autovalores de  $A$  pertencerão ao exterior do círculo unitário para qualquer valor positivo de  $\gamma(\theta^i)$ . Logo, os máximos isolados não são atrativos.  $\blacksquare$*

Este teorema mostra que se o tamanho do passo utilizado for suficientemente pequeno, então um mínimo isolado é atrativo para o método do gradiente. Por outro lado, todos os máximos isolados não são atrativos, não importando qual seja a escolha para o tamanho dos passos. Este resultado mostra que apesar dos máximos serem pontos fixos do algoritmo, o algoritmo não converge para eles. Este resultado é ilustrado no próximo exemplo.

**Exemplo 3.1** *Considere a seguinte função custo*

$$V(\theta) = e^\theta(\theta - 1)^2$$

*a qual é apresentada da Figura 3.2. Esta função custo tem dois extremos isolados em  $\theta = -1$  e  $\theta = 1$ . O ponto  $\theta = 1$  é um mínimo e o ponto  $\theta = -1$  é um máximo da função. Vamos aplicar o algoritmo (3.8) para procurar pelo mínimo da função.*

*O gradiente da função custo é dado por*

$$\nabla V(\theta) = e^\theta(2\theta - 2) + e^\theta(\theta - 1)^2.$$

*Seguindo a discussão apresentada na seção anterior, um algoritmo converge para um ponto se este ponto é atrativo e se o algoritmo é iniciado dentro de seu domínio de atração. Foi mostrado que um ponto de mínimo é atrativo se for feita uma escolha adequada para o tamanho dos passos; e que os máximos não são atrativos. Para ilustrar estas propriedades vamos testar 3 condições iniciais diferentes, com regras para o tamanho do passo também diferentes.*

*A primeira condição inicial é  $\theta^1 = -0,5$ . Vamos utilizar um passo constante:  $\gamma(\theta^i) = 0,2 \forall i$ . A evolução do algoritmo para 20 iterações é apresentada na Figura 3.3, onde pode ser visto que o algoritmo converge para o mínimo global  $\theta = 1$ . O mínimo global é um ponto fixo atrativo para o método do gradiente com esta regra de tamanho de passo. Além disso, a condição inicial  $\theta^1 = -0,5$  está dentro do domínio de atração.*

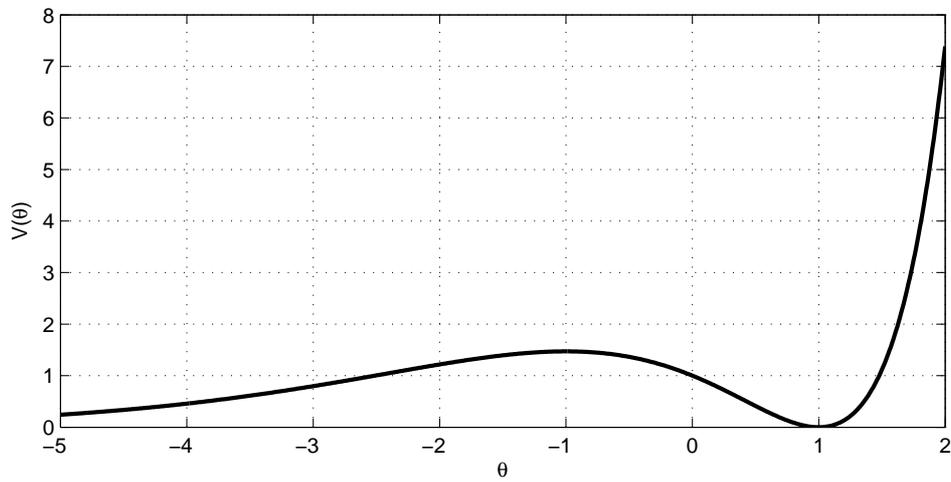


Figura 3.2: Função custo  $V(\theta) = e^\theta(\theta - 1)^2$ .

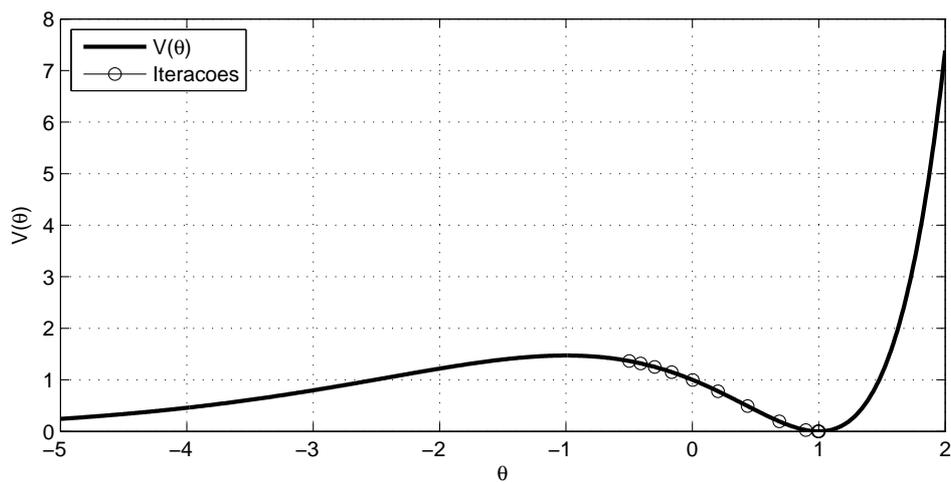


Figura 3.3: Convergência do método do gradiente com a condição inicial  $\theta^1 = -0,5$  e passos  $\gamma(\theta^i) = 0,2 \forall i$ .

Para uma segunda escolha, vamos testar a condição inicial  $\theta^1 = 1,1$  com  $\gamma(\theta^i) = 0,7 \forall i$ . O comportamento para as primeiras 20 iterações pode ser visto na Figura 3.4. O algoritmo divergiu do mínimo global  $\theta = 1$ . Embora a condição inicial estivesse bastante próxima do mínimo, foi utilizada uma regra que não era adequada para o tamanho dos passos. Com essa regra para os passos o mínimo global não é um ponto fixo atrativo.

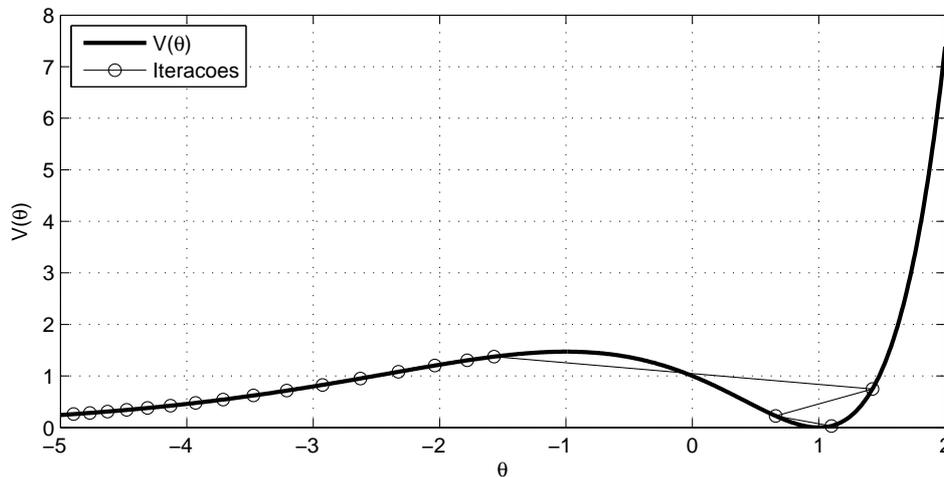


Figura 3.4: Divergência do método do gradiente para a condição inicial  $\theta^1 = 1,1$  e passos  $\gamma(\theta^i) = 0,7 \forall i$ .

A terceira escolha é  $\theta^1 = -1,5$ ,  $\gamma(\theta^i) = 0,5 \forall i$ . Novamente o algoritmo divergiu, como mostrado na Figura 3.5, pois a condição inicial estava muito distante do mínimo global e não pertencia a um domínio de atração.

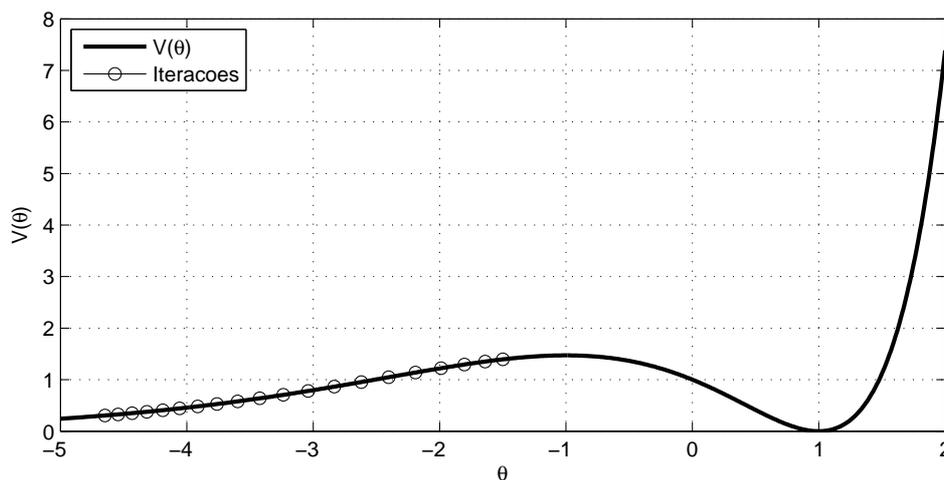


Figura 3.5: Divergência do método do gradiente para a condição inicial  $\theta^1 = -1,5$  e passos  $\gamma(\theta^i) = 0,5 \forall i$ .

Foi mostrado teoricamente e por exemplos que o método do gradiente converge para um mínimo desde que uma escolha apropriada do tamanho dos passos seja feita. Mas esta é somente uma das propriedades do método. Gostaríamos de obter resultados sobre a convergência do método para o mínimo global da função custo, e para o maior conjunto possível de condições iniciais. O próximo resultado avança nestes aspectos.

**Teorema 3.5** *Considere uma função diferenciável  $V(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^+$ . Assuma que esta função possui um mínimo global  $\theta^* \in \mathcal{C}$  e que existe um  $\epsilon$  tal que*

$$\mathcal{B}_\epsilon(\theta^*) \subset \mathcal{C} \quad (3.9)$$

$$(\theta - \theta^*)^T \nabla V(\theta) > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^* \quad (3.10)$$

*Então existe uma regra para o tamanho dos passos  $\gamma(\theta^i)$  tal que  $\mathcal{B}_\epsilon(\theta^*)$  seja um domínio de atração do mínimo global de  $V(\theta)$  para o algoritmo (3.8).  $\square$*

*Prova:* Seja  $V(\theta) = (\theta - \theta^*)^T(\theta - \theta^*)$  uma função de Lyapunov candidata para o sistema em tempo discreto (3.8). Então

$$\begin{aligned} V(\theta^{i+1}) - V(\theta^i) &= (\theta^i - \gamma_i \nabla V(\theta^i) - \theta^*)^T \times \\ &\quad \times (\theta^i - \gamma_i \nabla V(\theta^i) - \theta^*) - (\theta^i - \theta^*)^T (\theta^i - \theta^*) \\ &= -2\gamma_i (\theta^i - \theta^*)^T \nabla V(\theta^i) + \gamma_i^2 \nabla V(\theta^i)^T \nabla V(\theta^i) \end{aligned}$$

o que é negativo desde que

$$\gamma(\theta^i) < 2 \frac{(\theta^i - \theta^*)^T \nabla V(\theta^i)}{\nabla V(\theta^i)^T \nabla V(\theta^i)}. \quad (3.11)$$

Para  $\theta^i \in \mathcal{B}_\epsilon(\theta^*)$  a existência de tal  $\gamma(\theta^i)$  é garantida pela condição (3.9). A prova é completa por notar que  $\mathcal{B}_\epsilon(\theta^*)$  é um conjunto de nível de  $V(\theta)$ .  $\blacksquare$

Note que a prova do Teorema 3.5 é construtiva: ela apresenta uma condição explícita para  $\gamma(\theta^i)$  em (3.11), mas  $\gamma(\theta^i)$  depende do mínimo global  $\theta^*$ , o qual é desconhecido. Embora esta condição não possa ser verificada, ela fornece uma informação importante sobre como deve ser a escolha destes passos. Se o tamanho dos passos for suficientemente pequeno, então  $\mathcal{B}_\epsilon(\theta^*)$  é um domínio de atração do mínimo global de  $V(\theta)$ .

Uma das regras mais utilizadas para a escolha do tamanho dos passos consiste em uma busca linear. De posse do gradiente da função custo em um ponto  $\theta^1$ , avalia-se o valor da função custo para diversos tamanhos diferentes de passo. Escolhe-se então aquele que gera o menor valor para a função custo (POLAK, 1973; ARMIJO, 1966).

No Teorema 3.5 foi assumido que  $\mathcal{B}_\epsilon(\theta^*) \subset \mathcal{C}$ . O conjunto  $\mathcal{C}$  possui diversas propriedades que serão exploradas na Seção 4.1.

### 3.3.2 O método de Newton-Raphson

O método de Newton-Raphson é um pouco mais sofisticado que o método do gradiente, tanto em sua concepção como em sua implementação. Ele é definido como a seguinte recursão:

$$\theta^{i+1} = \theta^i - (\nabla^2 V(\theta^i))^{-1} \nabla V(\theta^i) \quad (3.12)$$

onde  $\nabla^2 V(\theta) \triangleq \frac{\partial^2 V(\theta)}{\partial \theta^2}$  é a Hessiana da função custo. Observe que o método não pode ser utilizado nos pontos que a Hessiana é singular. Observe que este método não exige nenhuma escolha de parâmetros, como é o caso do tamanho dos passos do método do gradiente.

O método foi inicialmente desenvolvido para encontrar as raízes de um sistema de equações não-lineares. Sabendo que  $\nabla V(\theta) = \mathbf{0}$  para os mínimos de uma função custo, podemos tentar resolver tal equação para encontrar um mínimo. Contudo, note que para qualquer ponto crítico de  $V(\theta)$  o gradiente é nulo, e portanto o método pode encontrar mínimos, máximos ou pontos de inflexão. Vamos analisar as propriedades deste método, como foi feito anteriormente.

**Teorema 3.6** Um dado ponto  $\theta^*$  no espaço de parâmetros é um ponto fixo do método de Newton Raphson (3.12) se ele é um ponto crítico da função  $V(\theta)$ .  $\square$

*Prova:* Se  $\theta^*$  é um extremo, então pelo Lema 3.1  $\nabla V(\theta^*) = \mathbf{0}$ , o que implica  $\theta^{i+1} = \theta^i$  na equação (3.12).  $\blacksquare$

Agora, em relação à atratividade dos extremos temos a seguinte propriedade.

**Teorema 3.7** Seja  $\theta^*$  um ponto crítico de  $V(\theta)$ . Então  $\theta^*$  é um ponto-fixo atrativo do método de Newton-Raphson.  $\square$

*Prova:*

Vamos utilizar novamente o método indireto de Lyapunov.

$$\begin{aligned} A = \left. \frac{\partial f(\theta)}{\partial \theta} \right|_{\theta=\theta^*} &= I_n - (\nabla^2 V(\theta))^{-1} (\nabla^2 V(\theta)) - \left. \frac{\partial (\nabla^2 V(\theta))^{-1} \nabla V(\theta)}{\partial \theta} \right|_{\theta=\theta^*} \\ &= I_n - I_n - \mathbf{0}_n = \mathbf{0}_n. \end{aligned}$$

Logo, os autovalores de  $A$  pertencem ao círculo unitário.  $\blacksquare$

Então, infelizmente, a atratividade do método de Newton-Raphson não é restrita aos mínimos. Este fato é um problema para a utilização do método para buscar os mínimos da função custo. Como existem mais pontos fixos atrativos, o domínio de atração de cada mínimo tende a ser menor. Como consequência, o método de Newton-Raphson requer uma melhor inicialização para que ele convirja para um mínimo. Por outro lado, o método apresenta uma taxa de convergência bastante alta para condições iniciais próximas do mínimo. Para obter uma estimativa da taxa de convergência do método de Newton-Raphson, vamos considerar que a função custo é quadrática, ou seja:

$$V(\theta) = a + \mathbf{b}^T \theta + \theta^T C \theta$$

onde  $a$ ,  $\mathbf{b}$  e  $C = C^T$  são constantes,  $C$  não é singular e  $\theta^*$  é o único ponto crítico da função, pois como  $C$  não é singular existe apenas um ponto crítico. O gradiente da função custo é

$$\nabla V(\theta) = \mathbf{b} + 2C\theta \quad (3.13)$$

e  $\theta^*$  é a solução da equação linear  $\nabla V(\theta) = 0$ , ou seja,  $\theta^* = -(2C)^{-1}\mathbf{b}$ . Somando e subtraindo  $2C\theta^*$  em (3.13) gera

$$\nabla V(\theta) = \mathbf{b} + 2C\theta^* + 2C(\theta - \theta^*) = 2C(\theta - \theta^*) \quad (3.14)$$

Finalmente, isolando  $\theta^*$  em (3.14) gera

$$\theta^* = \theta - (2C)^{-1} \nabla V(\theta). \quad (3.15)$$

A equação (3.15) mostra que o ponto crítico da função pode ser exatamente calculado por esta expressão para qualquer  $\theta$ , o qual é exatamente a recursão do método de Newton-Raphson pois  $\nabla^2 V(\theta) = 2C$ . Portanto, para qualquer condição inicial, o método de Newton-Raphson encontra o ponto crítico em apenas um passo.

Concluindo, se a função custo fosse quadrática, então o método iria encontrar o ponto crítico em apenas uma iteração. Logo pode-se esperar que o método de Newton-Raphson convirja em poucas iterações se a função custo for aproximadamente quadrática. De fato isto acontece, e o método de Newton-Raphson converge muito mais rapidamente para

um mínimo que o método do gradiente, desde que a condição inicial do algoritmo esteja próxima do mínimo, pois perto de um mínimo funções duplamente diferenciáveis podem ser aproximadas por quadráticas. Para aumentar a taxa de convergência, o método de Newton-Raphson utiliza uma informação extra sobre a função custo: a Hessiana. Em muitas aplicações pode ser bastante custoso calcular a Hessiana, o que pode inviabilizar a utilização do método.

Estas propriedades serão ilustradas no Exemplo 3.2.

**Exemplo 3.2** *Considere novamente a seguinte função custo*

$$V(\theta) = e^\theta(\theta - 1)^2$$

que é apresentada na Figura 3.2, a qual possui um mínimo isolado em  $\theta = -1$  e um máximo isolado em  $\theta = 1$ . O método do gradiente e o método de Newton-Raphson serão utilizados para procurar o mínimo da função.

O gradiente da função custo é dado por

$$\nabla V(\theta) = e^\theta(2\theta - 2) + e^\theta(\theta - 1)^2.$$

e a Hessiana por

$$\nabla^2 V(\theta) = 2e^\theta + 2e^\theta(2\theta - 2) + e^\theta(\theta - 1)^2.$$

O algoritmo do gradiente converge para o mínimo global se a condição inicial pertence ao conjunto  $\Omega_{SD} = (-1 ; \infty)$ . Para condições iniciais fora deste conjunto o algoritmo diverge.

Com o método de Newton-Raphson, tanto o mínimo como o máximo são pontos fixos atrativos. Utilizando um busca exaustiva, pode-se verificar que o método de Newton-Raphson converge para o mínimo se as condições iniciais estão no conjunto  $\Omega_{NR}^{min} = (-2, 41 ; -1, 93) \cup (0, 25 ; 0, 28) \cup (0, 42 ; \infty)$ . Por outro lado, o conjunto  $\Omega_{NR}^{max} = (-1, 89 ; 0, 24)$  é um domínio de atração para o máximo. Para condições que não pertencem nem a  $\Omega_{NR}^{min}$  nem  $\Omega_{NR}^{max}$  o algoritmo diverge.

Algumas destas simulações ilustrativas são apresentadas a seguir. A Figura 3.6 mostra a convergência do método do gradiente para o mínimo, utilizando a condição inicial  $\theta^1 = -0,8$  e o tamanho de passo  $\gamma = 0,2$ . A Figura 3.7 mostra a convergência do Método de Newton-Raphson para o máximo, com a condição inicial  $\theta^1 = 0,2$ . A Figura 3.8 mostra a convergência do Método de Newton-Raphson para o mínimo com a condição inicial  $\theta^1 = 0,5$ . Quando o método de Newton-Raphson converge, ele o faz muito mais rapidamente (utilizando menor número de iterações) que o método do gradiente.

### 3.3.3 Outros algoritmos de busca

Muitos outros algoritmos de busca podem ser descritos como uma combinação do método do gradiente e do método de Newton-Raphson. O método de Newton-Raphson possui uma excelente taxa de convergência quando o algoritmo está próximo do mínimo, e utiliza em sua formulação uma estimativa da Hessiana. Já o método do gradiente possui um domínio de atração muito maior que o do método de Newton-Raphson, ao custo de ser mais lento. Parece natural então combinar os dois métodos para tentar obter um algoritmo com um grande domínio de atração, alta velocidade e baixo custo computacional.

Vários algoritmos foram criados mas na essência são apenas regras diferentes para a escolha de  $\gamma(\theta^i)$  e  $R(\theta^i)$ . Muitos dos algoritmos combinam estimativas menos custosas para a Hessiana da função custo, mas que se comportam bem próximas de um mínimo,

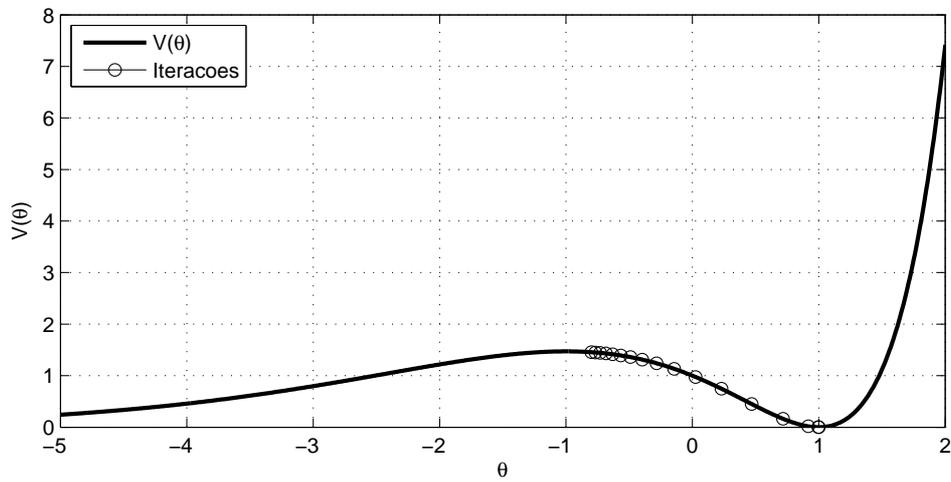


Figura 3.6: Convergência do método do gradiente com a condição inicial  $\theta^1 = -0,8$  e tamanho de passos  $\gamma(\theta^i) = 0,2 \forall i$ .

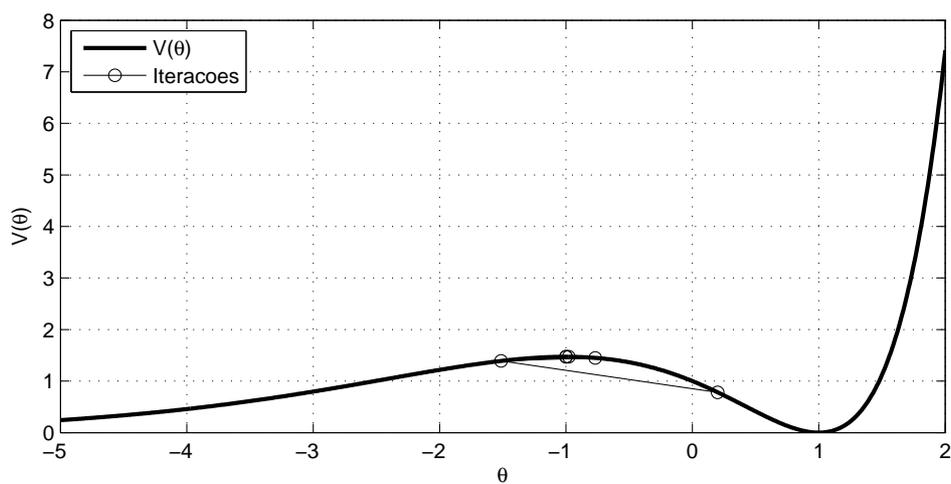


Figura 3.7: Divergência do método de Newton-Raphson com a condição inicial  $\theta^1 = 0,2$ .

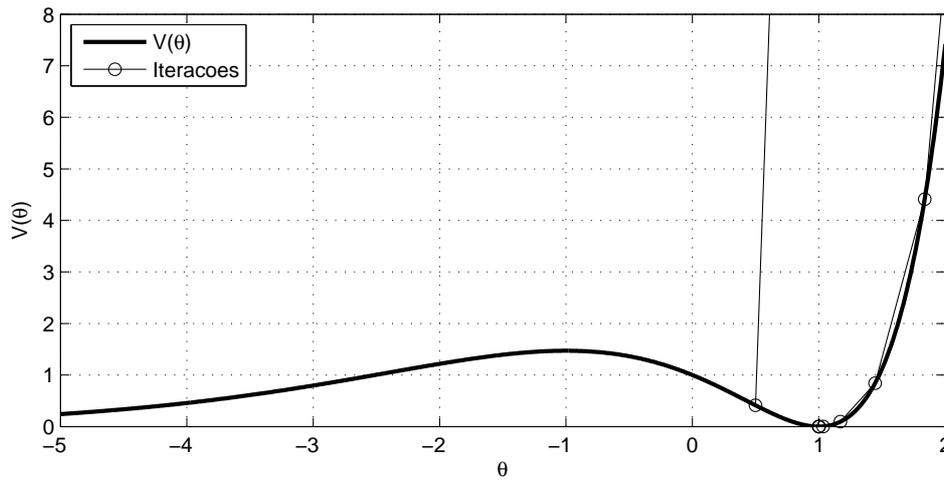


Figura 3.8: Convergência do método de Newton-Raphson com a condição inicial  $\theta^1 = 0,5$ .

com regras de passos mais conservadoras quando os algoritmos estão distantes dos mínimos. Alguns desses algoritmos foram batizados de Gauss-Newton, Levenberg-Marquardt e quasi-Newton.

### 3.4 Robustez

As informações necessárias para utilizar os algoritmos de otimização apresentados anteriormente são o gradiente da função custo e em algumas vezes a Hessiana. Em muitas aplicações, apenas uma estimativa da função custo é conhecida; logo apenas estimativas do gradiente e da Hessiana estão disponíveis para o algoritmo de otimização. Naturalmente uma questão surge: as propriedades dos algoritmos continuam válidas se apenas aproximações do gradiente (e da Hessiana) estão disponíveis? Algumas respostas serão respondidas no próximo teorema.

**Teorema 3.8** *Considere um sistema discreto*

$$\theta^{i+1} = f(\theta^i)$$

*com ponto de equilíbrio  $\theta^*$ . Este sistema admite uma função de Lyapunov quadrática*

$$V(\theta^i) = (\theta^i - \theta^*)^T (\theta^i - \theta^*)$$

*onde  $\exists \beta < 1, \alpha > 0$  tal que*

$$V(f(\theta)) < \beta^2 V(\theta), \forall \theta \in \mathcal{B}_\alpha(\theta^*).$$

*Onde*

$$\mathcal{B}_\alpha(\theta^*) = \{\theta : \|(\theta - \theta^*)\| < \alpha\}.$$

*Considere que este sistema é perturbado por  $g_i$*

$$\theta^{i+1} = f(\theta^i) + g_i \tag{3.16}$$

*onde  $\|g_i\| < \delta \forall i$ .*

Considere ainda o conjunto

$$\mathcal{S}(\theta^*) = \left\{ \theta : \|(\theta - \theta^*)\| < \frac{\delta}{1 - \beta} \right\}.$$

e que  $\mathcal{S}(\theta^*) \subset \mathcal{B}_\alpha(\theta^*)$ .

Então, para toda condição inicial  $\theta^1 \in \mathcal{B}_\alpha(\theta^*)$  o sistema perturbado (3.16) converge para o conjunto invariante  $\mathcal{S}(\theta^*)$ .  $\square$

*Prova:*

Para estudar a convergência do sistema é utilizada uma função de Lyapunov quadrática  $V(\theta^i)$ .

Para todo  $\theta^1 \in \mathcal{B}_\alpha(\theta^*)$  o sistema perturbado (3.16) converge para  $\mathcal{S}(\theta^*)$  se

$$V(\theta^{i+1}) - V(\theta^i) < 0; \quad \forall \theta^i \in \{\mathcal{B}_\alpha(\theta^*) - \mathcal{S}(\theta^*)\}. \quad (3.17)$$

Usando as condições do teorema  $\forall \theta^i \in \mathcal{B}_\alpha(\theta^*)$

$$\begin{aligned} V(\theta^{i+1}) - V(\theta^i) &= \|(f(\theta^i) - \theta^*)\|^2 - \|(\theta^i - \theta^*)\|^2 + 2g_i^T(f(\theta^i) - \theta^*) + g_i^T g_i \\ &< \beta^2 \|(\theta^i - \theta^*)\|^2 - \|(\theta^i - \theta^*)\|^2 + 2\delta\beta \|(\theta^i - \theta^*)\| + \delta^2 \end{aligned}$$

Então uma condição suficiente para garantir (3.17) é

$$(\beta \|(\theta^i - \theta^*)\| + \delta)^2 - \|(\theta^i - \theta^*)\|^2 < 0 \quad \forall \theta^i \in \{\mathcal{B}_\alpha(\theta^*) - \mathcal{S}(\theta^*)\}.$$

Da definição de  $\mathcal{S}(\theta^*)$  é fácil notar que a condição acima é respeitada.

Ainda é necessário mostrar que o conjunto  $\mathcal{S}(\theta^*)$  é invariante. Usando as hipóteses do teorema, se  $\theta^i \in \mathcal{S}(\theta^*)$  então

$$\begin{aligned} \|\theta^{i+1} - \theta^*\| &= \|f(\theta^i) - \theta^* + g_i\| < \|f(\theta^i) - \theta^*\| + \|g_i\| \\ &< \beta \|\theta^i - \theta^*\| + \|g_i\| < \frac{\beta\delta}{1 - \beta} + \delta = \frac{\delta}{1 - \beta} \end{aligned}$$

o que garante que  $\mathcal{S}(\theta^*)$  é invariante.  $\blacksquare$

O teorema mostra que se o sistema possui uma taxa de convergência especificada por  $\beta$ , então quando uma perturbação  $g_i$  age sobre o sistema, ele converge para uma bola cujo raio é  $\frac{\delta}{1 - \beta}$ , a qual é centrada no ponto fixo  $\theta^*$ . Se o sistema é inicializado fora desta bola, então ele converge para este conjunto e permanece nele para sempre.

Note que o tamanho da bola é proporcional ao tamanho do distúrbio e é também função da taxa de convergência  $\beta$ . Portanto, caso um algoritmo convergiria para um ponto fixo se não fosse perturbado, então este algoritmo converge para uma vizinhança deste ponto fixo quando um distúrbio o afeta.

### 3.5 Qual algoritmo escolher

Este capítulo apresentou dois métodos diferentes para a minimização de uma função custo. Primeiramente foi apresentado o método do gradiente e algumas de suas propriedades. Para este método apenas os pontos de mínimo são atrativos, o que é uma vantagem com relação a outros métodos. A convergência do método depende da escolha do tamanho do passo a cada iteração, o que é uma escolha do usuário. Se o tamanho do passo

é demasiadamente grande então não é possível convergir para o mínimo global da função custo. O conjunto dos pontos a partir dos quais o algoritmo converge para o mínimo global é chamado de domínio de atração.

Para o método de Newton-Raphson qualquer ponto crítico é atrativo, e portanto o domínio de atração de um ponto de mínimo de interesse deste método é menor que o do método do gradiente. Para que o método de Newton-Raphson convirja para o mínimo global é necessário que a condição inicial esteja próxima do mínimo global. O método de Newton-Raphson possui uma excelente taxa de convergência quando a função custo se assemelha a uma função quadrática. Como qualquer função duplamente diferenciável se aproxima de uma quadrática para os pontos próximos de um mínimo, o método é bastante eficiente para condições iniciais próximas do mínimo global.

Dadas as propriedades dos métodos apresentados, é aconselhado utilizar inicialmente o método do gradiente, tomando a cautela de utilizar pequenos passos a cada iteração. Desta maneira, após uma série de iterações será obtida uma estimativa muito mais próxima do mínimo global que a estimativa inicialmente utilizada. Pode-se avaliar a taxa de convergência do algoritmo pelo decaimento do valor da função custo de iteração em iteração. Quando a taxa de convergência do método do gradiente se tornar muito baixa aconselha-se trocar para o método de Newton-Raphson. O método de Newton-Raphson possui uma taxa de convergência elevada para condições iniciais próximas do mínimo global, e deve encontrar o mínimo em poucas iterações. Esta abordagem foi utilizada com sucesso em (CAMPESTRINI et al., 2012).

O momento para a troca de um método para o outro é uma escolha que o usuário deve fazer. Idealmente esta troca não é necessária pois o método do gradiente consegue encontrar o mínimo global se utilizado um grande número de iterações. Contudo, na prática, a taxa de convergência do método do gradiente é muito baixa para pontos próximos do mínimo, o que torna o custo computacional muito alto. Uma ideia interessante consiste em avaliar a taxa de convergência do método do gradiente; quando a taxa se tornar muito baixa pode-se então trocar de algoritmo. Deve-se contudo tomar muito cuidado para não utilizar o método de Newton-Raphson muito cedo. Como o domínio de atração do método tende a ser pequeno, a troca precipitada de métodos pode não levar a convergência para o mínimo global. Este cuidado não é tomado por muitos usuários e criadores de aplicativos e vê-se na prática que muitas vezes o mínimo global não é encontrado pela má escolha de algoritmos de otimização. Podemos citar aqui tanto o “System Identification Toolbox” como “University of Newcastle Identification Toolbox” para MATLAB como aplicativos que escolhem de forma não aconselhável os algoritmos de otimização.

### 3.6 Considerações finais

Este capítulo apresentou alguns conceitos importantes sobre algoritmos de otimização e seus pontos de equilíbrio. Foram apresentados os dois métodos de Lyapunov, os quais são ferramentas muito utilizadas para a análise da estabilidade de sistemas não lineares. Na sequência foram apresentados dois algoritmos de otimização muito utilizados na identificação de sistemas. Foi apresentado o método do gradiente e uma análise da estabilidade de seus pontos de equilíbrio. Foi também apresentado o método de Newton-Raphson, que possui duas propriedades relevantes: seu domínio de atração é menor que o do método do gradiente e sua taxa de convergência é mais alta para pontos próximos do mínimo global. Por fim este trabalho propôs que os dois algoritmos apresentados sejam utilizados em sequência para combinar o maior domínio de atração do método do

gradiente com a maior taxa de convergência do método de Newton-Raphson.

Uma das contribuições deste trabalho é a análise de robustez dos métodos para avaliar a sensibilidade aos erros na estimativa do gradiente da função custo. Esta contribuição resultou na seguinte publicação (ECKHARD; BAZANELLA, 2012).

## 4 CONVERGÊNCIA PARA O MÍNIMO GLOBAL

No Capítulo 2 foi apresentado o método de identificação de sistemas baseado na minimização do erro de predição. Foi mostrado que este método, sob certas condições, gera uma estimativa consistente dos parâmetros do sistema cuja variância decai com o tamanho do experimento  $N$  utilizado para coletar os dados.

Foi mostrado também que este método é baseado na resolução de um problema de otimização, onde procuram-se em um conjunto os parâmetros que minimizam o valor médio quadrático do erro de predição. São utilizados algoritmos iterativos na busca pelo mínimo global do critério deste problema de otimização.

No Capítulo 3 foram apresentados algoritmos iterativos muito utilizados na resolução de problemas de otimização. Foi apresentado o algoritmo do gradiente e algumas de suas propriedades. Este algoritmo possui uma região de atração maior que outros algoritmos e portanto é indicado quando não se possui uma boa inicialização. Em contrapartida, o algoritmo possui uma taxa de convergência mais baixa e não é indicado nos casos em que é sabido que se está próximo de um mínimo. Já o algoritmo de Newton-Raphson possui uma taxa de convergência bastante alta, mas somente é indicado para os casos em que a estimativa dos parâmetros está próxima de um mínimo. Logo, foi proposta a utilização dos dois algoritmos em sequência: o algoritmo do gradiente leva a estimativa para próximo do valor do mínimo e o algoritmo de Newton-Raphson finaliza o procedimento de otimização encontrando efetivamente um mínimo do critério.

É importante ressaltar que o objetivo do problema de otimização é encontrar o mínimo **global** do critério, pois é desse mínimo que foram obtidas as propriedades estatísticas no Capítulo 2. O Capítulo 3 mostrou que os algoritmos convergem para o mínimo global da função custo se a condição inicial pertencer ao *domínio de atração* do mínimo global.

Ao estudar o domínio de atração do algoritmo do gradiente, foi introduzido na Seção 3.3.1, o conjunto  $\mathcal{C}$ , no qual  $(\theta - \theta^*)^T \nabla V(\theta) > 0 \forall \theta \in \mathcal{C}; \theta \neq \theta^*$ . Neste capítulo vamos estudar este conjunto para obter condições sobre a convergência dos algoritmos iterativos para o mínimo global da função custo do método de identificação pela minimização do erro de predição.

As condições sobre a convergência dos algoritmos são uma das *contribuições inéditas* desta tese e são a base para a construção das *ferramentas para melhoria da convergência dos métodos de identificação por erro de predição* que são propostas nos capítulos a seguir.

### 4.1 Função semi-convexa

Antes de estudar a convergência dos algoritmos para o mínimo global precisamos definir a região em que a função custo é semi-convexa. As definições a seguir não são

usuais e portanto vamos definir alguns conceitos e apresentar propriedades relativas à estes conceitos.

**Definição 4.1** *Função semi-convexa*

Seja  $\theta^*$  o mínimo global de uma função diferenciável  $V(\theta)$ . Esta função é dita semi-convexa em um conjunto  $\mathcal{C} \ni \theta^*$  se

$$(\theta - \theta^*)^T \nabla V(\theta) > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^*. \quad (4.1)$$

◇

O conjunto  $\mathcal{C}$  no qual a função é semi-convexa possui diversas propriedades relevantes. A primeira delas é que  $\theta^*$  é o único ponto crítico de  $V(\theta)$  no conjunto  $\mathcal{C}$ . Matematicamente temos que

$$\nabla V(\theta) \neq \mathbf{0} \quad \forall \theta \neq \theta^*, \theta \in \mathcal{C}.$$

Neste caso, o único ponto em  $\mathcal{C}$  para o qual o método do gradiente e o método de Newton-Raphson podem convergir é o mínimo global  $\theta^*$  da função custo  $V(\theta)$ . Quanto maior for o conjunto  $\mathcal{C}$ , mais distantes do mínimo global estarão outros pontos críticos da função custo. Portanto é de se esperar que quanto maior for o conjunto  $\mathcal{C}$  mais fácil será para um algoritmo convergir para o mínimo global.

A condição (4.1) possui também algumas interpretações geométricas. Dentro do conjunto  $\mathcal{C}$  o ângulo  $\alpha$  entre  $(\theta - \theta^*)$  e  $\nabla V(\theta)$  é sempre menor que  $\frac{\pi}{2}$  rad – veja a Figura 4.1.

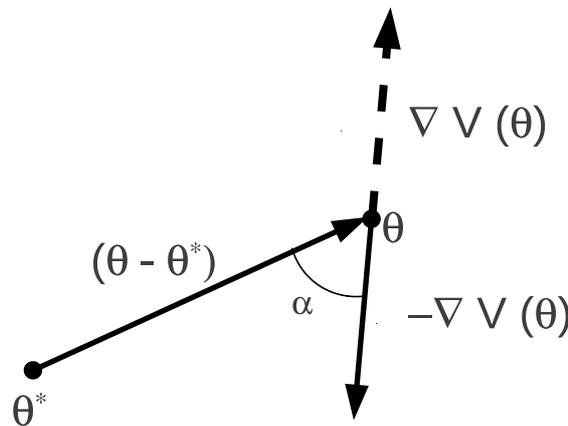


Figura 4.1: Direção do gradiente.

A propriedade apresentada mostra que a convergência do método do gradiente para o mínimo global da função custo é mais fácil para os pontos dentro da região que a função custo é semi-convexa.. Lembre que o método do gradiente se move na direção oposta ao gradiente. Quando este método é utilizado, a cada iteração em que  $\theta^i \in \mathcal{C}$  o algoritmo se aproxima mais um pouco do mínimo global, se for dado um passo suficientemente pequeno. O único cuidado que deve ser dado é para não utilizar um tamanho de passo excessivamente grande.

Outra propriedade interessante das funções semi-convexas é que elas são **radialmente crescentes**. A definição de função radialmente crescente é dada a seguir.

**Definição 4.2** Seja  $\theta^*$  um ponto de mínimo de função  $V(\theta)$ . Esta função é **radialmente crescente** em  $\mathcal{C}$  se para quaisquer dois pontos  $\theta^a, \theta^b \in \mathcal{C}$  onde

$$\theta^a = \theta^* + k(\theta^b - \theta^*) \quad (4.2)$$

$$0 < k < 1 \quad (4.3)$$

o valor da função é tal que

$$V(\theta^a) < V(\theta^b).$$

◇

O teorema a seguir mostra que funções semi-convexas são radialmente crescentes.

**Teorema 4.1** Uma função semi-convexa em  $\mathcal{C}$  é radialmente crescente em  $\mathcal{C}$ . □

*Prova:*

O valor da função no ponto  $\theta_b$  é igual ao valor em  $\theta_a$  somado à integral de linha do gradiente entre os dois pontos.

$$V(\theta^b) = V(\theta^a) + \int_C \nabla V^T(\theta) \cdot d\theta$$

onde  $C$  é a reta que liga o ponto  $\theta_a$  ao ponto  $\theta_b$ .

Vamos parametrizar a reta  $C$  em função de uma variável  $k$ :

$$\theta(k) = \theta^* + k(\theta^b - \theta^*)$$

de maneira que quando  $\theta(1) = \theta^b$  e  $\theta\left(\frac{\|\theta^a - \theta^*\|}{\|\theta^b - \theta^*\|}\right) = \theta^a$ .

Portanto temos que

$$V(\theta^b) = V(\theta^a) + \int_{\frac{\|\theta^a - \theta^*\|}{\|\theta^b - \theta^*\|}}^1 \nabla V^T(\theta(k)) \cdot \frac{d\theta}{dk} dk$$

que pode ser reescrita como

$$V(\theta^b) = V(\theta^a) + \int_{\frac{\|\theta^a - \theta^*\|}{\|\theta^b - \theta^*\|}}^1 \nabla V^T(\theta(k)) \cdot (\theta^b - \theta^*) dk. \quad (4.4)$$

Para qualquer  $\theta \in \mathcal{C}$  tal que  $\theta \neq \theta^*$  temos que

$$\nabla V^T(\theta)(\theta - \theta^*) > 0.$$

Logo, para qualquer  $k$  tal que  $0 < k < 1$  temos que

$$\nabla V^T(\theta(k))(\theta^* + k(\theta^b - \theta^*) - \theta^*) > 0$$

e portanto

$$\nabla V^T(\theta(k))k(\theta^b - \theta^*) > 0.$$

Como  $k > 0$  então

$$\nabla V^T(\theta(k))(\theta^b - \theta^*) > 0.$$

Logo, o integrando de (4.4) é positivo e portanto a integral também é positiva. Conclui-se então que  $V(\theta^b) > V(\theta^a)$ . ■

O fato de uma função semi-convexa ser radialmente crescente é muito conveniente para a otimização desta função, pois os conjuntos de nível das funções radialmente crescentes possuem propriedade importantes.

**Definição 4.3** Um conjunto de nível de uma função  $V(\theta)$  é definido como

$$\mathcal{Q}_c \triangleq \{\theta : V(\theta) \leq c\}.$$

◇

Uma das propriedades dos conjuntos de nível de uma função semi-convexa é

$$a < b \quad \Rightarrow \quad \mathcal{Q}_a \subset \mathcal{Q}_b.$$

Se um algoritmo consegue a cada iteração encontrar um novo ponto cujo valor da função custo é menor que o anterior, então a cada iteração o algoritmo entrará em um novo conjunto de nível, o qual é um subconjunto do conjunto de nível anterior. Desta maneira o algoritmo estará, a cada iteração, mais próximo do mínimo global  $\theta^*$ . A Figura 4.2 mostra um exemplo dos conjuntos de nível de uma função semi-convexa hipotética.

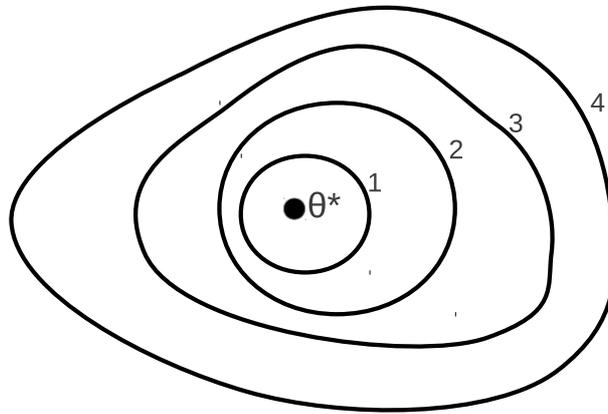


Figura 4.2: Exemplo de conjuntos de nível de uma função semi-convexa.

Outra propriedade dos conjuntos de nível de uma função semi-convexa é que se um ponto pertence a um conjunto de nível então qualquer combinação convexa entre este ponto e o mínimo global  $\theta^*$  também pertence ao conjunto de nível. Matematicamente temos

$$\theta^1 \in \mathcal{Q}_c \quad \Rightarrow \quad (\lambda_1 \theta^* + \lambda_2 \theta_1) \in \mathcal{Q}_c \text{ se } \lambda_1 + \lambda_2 = 1$$

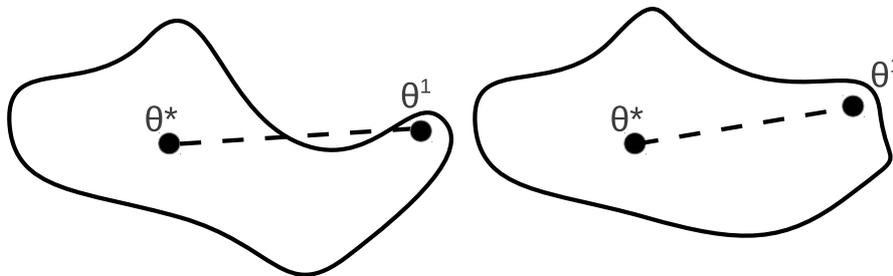


Figura 4.3: Exemplo de conjuntos de nível de funções semi-convexas (à direita) e não semi-convexas (à esquerda).

A Figura 4.3 apresenta dois conjuntos de nível. O conjunto da direita pode representar um conjunto de nível de uma função semi-convexa. Já o conjunto da esquerda não pode

representar um conjunto de nível de uma função semi-convexa, pois a reta pontilhada passa por pontos que não pertencem ao conjunto.

Outra propriedade das funções semi-convexas é que toda função convexa é também semi-convexa. Foi mostrado anteriormente que em torno de todo mínimo isolado de uma função existe um domínio de convexidade. Portanto, em torno de todo mínimo isolado existe uma região em que a função é semi-convexa.

Todas as propriedades apresentadas das funções semi-convexas mostram que é mais fácil encontrar o mínimo global de uma função se estamos procurando apenas em um conjunto que a função é semi-convexa. Portanto, se a função é semi-convexa em todo o domínio de busca, então é mais fácil encontrar o mínimo global desta função utilizando um algoritmo iterativo.

Neste trabalho vamos estudar a função custo relacionada com a minimização do erro de predição para entender em que região esta função é semi-convexa. Pretende-se então criar ferramentas para garantir que a função seja semi-convexa no maior conjunto possível. Desta maneira vamos melhorar a convergência dos métodos de identificação que utilizam a minimização do erro de predição.

## 4.2 Propriedades da função $\bar{V}(\theta)$

Foi apresentado na seção anterior o conceito de função *semi-convexa*. Foi mostrado que é mais fácil convergir para o mínimo global de uma função se a procura é feita apenas em uma região que a função é semi-convexa. Nesta seção vamos estudar a função custo do método de identificação pela minimização do erro de predição para observar que fatores influenciam o conjunto em que a função é semi-convexa. Uma análise das características da função custo para algumas estruturas de modelo é apresentada em (GOODWIN; AGÜERO; SKELTON, 2003; ZOU; HEATH, 2009, 2012). Neste capítulo vamos estudar o conjunto que a a função custo é semi-convexa considerando estruturas mais genéricas para os modelos e portanto estender os resultados encontrados na literatura.

A função custo pode ser descrita na forma integral por

$$\begin{aligned} \bar{V}(\theta) = & \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 d\omega \\ & + \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_0 |H^{-1}(e^{j\omega}, \theta)H_0(e^{j\omega}) - 1|^2 d\omega + \lambda_0. \end{aligned}$$

Ao analisar a função custo na forma integral podemos obter várias características importantes desta função. Observe que o valor da função custo sempre é positivo, pois tanto os integrandos das duas integrais são sempre positivos, assim como o nível do ruído  $\lambda_0$ .

Note que diversos fatores afetam esta função custo: as funções de transferência do sistema  $G_0(q)$  e  $H_0(q)$ , o nível de ruído  $\lambda_0$ , as funções de transferência do modelo  $G(q, \theta)$  e  $H(q, \theta)$  e o espectro do sinal de entrada  $\Phi_u(\omega)$ . Os fatores relacionados ao sistema ( $G_0(q)$ ,  $H_0(q)$  e  $\lambda_0$ ) são desconhecidos e não podem ser modificados pelo usuário. Já os parâmetros do modelo ( $G(q, \theta)$  e  $H(q, \theta)$ ) e o espectro do sinal de entrada  $\Phi_u(\omega)$  podem ser escolhidos pelo usuário.

Neste capítulo vamos fazer algumas hipóteses sobre a estrutura dos modelos e analisar como os fatores que o usuário pode modificar influenciam o formato da função custo e o conjunto em que a função é semi-convexa.

### 4.3 Parametrização genérica

Vamos assumir neste ponto do trabalho que o modelo possui uma estrutura fixa determinada por

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + F^T(q)\theta}, \quad H(q, \theta) = \frac{1 + C^T(q)\theta}{1 + D^T(q)\theta}$$

onde os vetores  $B(q), C(q), D(q), F(q) \in \mathbb{R}^n$  são formados por funções de transferência conhecidas e  $\theta \in \mathbb{R}^n$  é o vetor de parâmetros que se deseja estimar. Esta estrutura é bastante genérica e consegue representar uma grande variedade de modelos incluindo: ARX, ARMAX, Output-Error e Box-Jenkins.

Nesta estrutura tanto o numerador como o denominador são funções afins no vetor de parâmetros  $\theta$  que se deseja estimar. Esta propriedade simplifica o cálculo das derivadas parciais, como será visto na sequência.

$$\nabla G(q, \theta) = \frac{B(q)}{1 + F^T(q)\theta} - \frac{B^T(q)\theta F(q)}{(1 + F^T(q)\theta)^2}$$

e

$$\nabla H^{-1}(q, \theta) = \frac{D(q)}{1 + C^T(q)\theta} - \frac{(1 + D^T(q)\theta)C(q)}{(1 + C^T(q)\theta)^2}.$$

Utilizando esta estrutura de modelo podemos obter condições para que a função custo seja semi-convexa em um conjunto.

**Teorema 4.2** *Seja  $\theta^0$  o mínimo global da função  $\bar{V}(\theta)$ . A função custo é semi-convexa em um conjunto  $\mathcal{C}$  se*

$$\begin{aligned} & \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \\ & \cdot \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} + \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta} - \frac{1 + D^T(e^{j\omega})\theta^0}{1 + D^T(e^{j\omega})\theta} \right\} \\ & + \lambda_0 |H_0(e^{j\omega})(H^{-1}(e^{j\omega}, \theta) - H_0^{-1}(e^{j\omega}))|^2 \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} \right\} d\omega > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^0. \end{aligned} \quad (4.5)$$

□

*Prova:*

*Nesta prova serão omitidas as dependências das funções nas variáveis  $\theta$  e  $\omega$  para simplificar a notação.*

*Primeiramente vamos calcular o gradiente da função custo com relação ao vetor de parâmetros*

$$\begin{aligned} \nabla \bar{V}(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u \nabla (H^{-1}(G_0 - G))^* (H^{-1}(G_0 - G)) d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u (H^{-1}(G_0 - G))^* \nabla (H^{-1}(G_0 - G)) d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_0 \nabla (H_0(H^{-1} - H_0^{-1}))^* (H_0(H^{-1} - H_0^{-1})) d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_0 (H_0(H^{-1} - H_0^{-1}))^* \nabla (H_0(H^{-1} - H_0^{-1})) d\omega \end{aligned}$$

que pode ser rearranjado como

$$\begin{aligned}\nabla \bar{V}(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} 2\Phi_u \Re \left\{ (H^{-1}(G_0 - G))^* \nabla (H^{-1}(G_0 - G)) \right\} d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} 2\lambda_0 \Re \left\{ (H_0(H^{-1} - H_0^{-1}))^* (H_0 \nabla H^{-1}) \right\} d\omega.\end{aligned}$$

A primeira integral depende do espectro do sinal de entrada  $\Phi_u$  enquanto que a segunda depende no nível de ruído  $\lambda_0$ . Vamos utilizar este fato para dividir a expressão em duas parcelas:

$$\nabla \bar{V}(\theta) = \nabla \bar{V}_u(\theta) + \nabla \bar{V}_e(\theta).$$

A primeira parcela pode ser simplificada como

$$\nabla \bar{V}_u(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ (H^{-1}(G_0 - G))^* (\nabla H^{-1}(G_0 - G) - H^{-1} \nabla G) \right\} d\omega$$

enquanto que a segunda como

$$\nabla \bar{V}_e(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 \Re \left\{ (H_0(H^{-1} - H_0^{-1}))^* (H_0 \nabla H^{-1}) \right\} d\omega.$$

Agora vamos utilizar a estrutura do modelo para expandir as duas expressões. O primeiro termo se torna

$$\begin{aligned}\nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \left[ \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right]^* \right. \\ &\quad \cdot \left[ \left( \frac{D}{1 + C^T \theta} - \frac{(1 + D^T \theta)C}{(1 + C^T \theta)^2} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right. \\ &\quad \left. \left. - \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( \frac{B}{1 + F^T \theta} - \frac{B^T \theta F}{(1 + F^T \theta)^2} \right) \right] \right\} d\omega\end{aligned}$$

enquanto que o segundo pode ser escrito como

$$\nabla \bar{V}_e(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 \Re \left\{ (H_0(H^{-1} - H_0^{-1}))^* H_0 \left( \frac{D}{1 + C^T \theta} - \frac{(1 + D^T \theta)C}{(1 + C^T \theta)^2} \right) \right\} d\omega.$$

Relembre que a condição fundamental utilizada para caracterizar uma função semi-convexa é dada por

$$(\theta - \theta^0)^T \nabla \bar{V}(\theta) > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^0$$

Assim como expandimos o gradiente em duas parcelas, vamos também dividir a expressão acima da mesma forma:

$$(\theta - \theta^0)^T \nabla \bar{V}(\theta) = (\theta - \theta^0)^T \nabla \bar{V}_u(\theta) + (\theta - \theta^0)^T \nabla \bar{V}_e(\theta).$$

O primeiro termo pode ser computado como

$$\begin{aligned}(\theta - \theta^0)^T \nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \left[ \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right]^* \right. \\ &\quad \cdot \left[ \left( \frac{D^T(\theta - \theta^0)}{1 + C^T \theta} - \frac{(1 + D^T \theta)C^T(\theta - \theta^0)}{(1 + C^T \theta)^2} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right. \\ &\quad \left. \left. - \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( \frac{B^T(\theta - \theta^0)}{1 + F^T \theta} - \frac{B^T \theta F^T(\theta - \theta^0)}{(1 + F^T \theta)^2} \right) \right] \right\} d\omega\end{aligned}$$

o qual pode ser expandido como

$$\begin{aligned}
(\theta - \theta^0)^T \nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \left[ \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right]^* \right. \\
&\quad \cdot \left[ \left( \frac{D^T(\theta - \theta^0)}{1 + C^T \theta} - \frac{(1 + D^T \theta) C^T (\theta - \theta^0)}{(1 + C^T \theta)^2} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right. \\
&\quad \left. \left. - \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( \frac{(1 + F^T \theta) B^T (\theta - \theta^0) - B^T \theta F^T (\theta - \theta^0)}{(1 + F^T \theta)^2} \right) \right] \right\} d\omega.
\end{aligned}$$

Utilizando a hipótese  $\mathcal{S} \in \mathcal{M}$  sabemos que  $G(q, \theta^0) = G_0$  e portanto podemos reescrever a expressão acima como

$$\begin{aligned}
(\theta - \theta^0)^T \nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \left[ \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right]^* \right. \\
&\quad \cdot \left[ \left( \frac{D^T(\theta - \theta^0)}{1 + C^T \theta} - \frac{(1 + D^T \theta) C^T (\theta - \theta^0)}{(1 + C^T \theta)^2} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right. \\
&\quad \left. \left. + \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( \frac{(1 + F^T \theta^0)}{(1 + F^T \theta)} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right] \right\} d\omega.
\end{aligned}$$

Rearranjando os termos podemos chegar em

$$\begin{aligned}
(\theta - \theta^0)^T \nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \left[ \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right]^* \right. \\
&\quad \cdot \left[ \left( \frac{D^T(\theta - \theta^0)}{1 + D^T \theta} - \frac{C^T(\theta - \theta^0)}{1 + C^T \theta} + \frac{1 + F^T \theta^0}{1 + F^T \theta} \right) \right. \\
&\quad \left. \left. \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right] \right\} d\omega
\end{aligned}$$

a qual pode ser simplificada em

$$\begin{aligned}
(\theta - \theta^0)^T \nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \frac{D^T(\theta - \theta^0)}{1 + D^T \theta} - \frac{C^T(\theta - \theta^0)}{1 + C^T \theta} + \frac{1 + F^T \theta^0}{1 + F^T \theta} \right\} \\
&\quad \cdot \left| \left( \frac{1 + D^T \theta}{1 + C^T \theta} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right|^2 d\omega
\end{aligned}$$

e novamente em

$$\begin{aligned}
(\theta - \theta^0)^T \nabla \bar{V}_u(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ -\frac{1 + D^T \theta^0}{1 + D^T \theta} + \frac{1 + C^T \theta^0}{1 + C^T \theta} + \frac{1 + F^T \theta^0}{1 + F^T \theta} \right\} \\
&\quad \cdot |H^{-1}(G_0 - G)|^2 d\omega.
\end{aligned}$$

Podemos fazer todos os mesmos procedimentos para a parcela  $(\theta - \theta^0)^T \nabla \bar{V}_e$ .

$$\begin{aligned}
(\theta - \theta^0)^T \nabla \bar{V}_e(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 \Re \left\{ (H_0(H^{-1} - H_0^{-1}))^* \right. \\
&\quad \left. \cdot H_0 \left( \frac{D^T(\theta - \theta^0)}{1 + C^T \theta} - \frac{(1 + D^T \theta) C^T (\theta - \theta^0)}{(1 + C^T \theta)^2} \right) \right\} d\omega
\end{aligned}$$

a qual pode ser reescrita como

$$(\theta - \theta^0)^T \nabla \bar{V}_e(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 \Re \{ (H_0(H^{-1} - H_0^{-1}))^* \cdot (H_0(H^{-1} - H_0^{-1})) \Re \left\{ \frac{1 + C^T \theta^0}{1 + C^T \theta} \right\} \} d\omega$$

e simplificada para

$$(\theta - \theta^0)^T \nabla \bar{V}_e(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 |H_0(H^{-1} - H_0^{-1})|^2 \Re \left\{ \frac{1 + C^T \theta^0}{1 + C^T \theta} \right\} d\omega.$$

Agora podemos juntar os dois termos anteriormente separados para obter a seguinte expressão

$$\begin{aligned} (\theta - \theta^0)^T \nabla \bar{V}(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u |H^{-1}(G_0 - G)|^2 \\ &\cdot \Re \left\{ \frac{1 + C^T \theta^0}{1 + C^T \theta} + \frac{1 + F^T \theta^0}{1 + F^T \theta} - \frac{1 + D^T \theta^0}{1 + D^T \theta} \right\} \\ &+ \lambda_0 |H_0(H^{-1} - H_0^{-1})|^2 \Re \left\{ \frac{1 + C^T \theta^0}{1 + C^T \theta} \right\} d\omega. \end{aligned}$$

Para que a função seja semi-convexa no conjunto  $\mathcal{C}$  é necessário que a equação acima seja positiva para todo  $\theta$  dentro do conjunto com exceção do mínimo global  $\theta^0$ . Note que esta é a condição do enunciado e portanto está demonstrado o teorema. ■

O Teorema 4.2 apresenta uma condição que garante que a função custo é semi-convexa em um conjunto. Esta condição é interessante pois apresenta de forma explícita como diversos fatores afetam o formato da função custo. Neste trabalho será dada especial atenção para o espectro do sinal de entrada  $\Phi_u(\omega)$ . Se este espectro for escolhido de forma adequada é possível garantir a positividade da integral para um grande conjunto de parâmetros e conseqüentemente ampliar o conjunto em que a função custo é semi-convexa.

No Teorema 4.2 foi utilizada uma parametrização bastante genérica para a estrutura do modelo. Nas próximas seções serão desenvolvidos resultados para estruturas específicas, o que resultará em algumas condições mais simples sobre o conjunto que a função custo é semi-convexa.

## 4.4 Parametrizações usuais

Vamos agora calcular a expressão (4.1) para algumas estruturas de modelo comumente utilizadas.

### 4.4.1 Estrutura ARX generalizada

Vamos considerar que a estrutura possui a seguinte propriedade  $C(q) = 0$  e  $F(q) = D(q)$ . Então

$$\begin{aligned} (\theta - \theta^0)^T \nabla \bar{V}(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \\ &+ \lambda_0 |H_0(e^{j\omega})(H^{-1}(e^{j\omega}, \theta) - H_0^{-1}(e^{j\omega}))| d\omega > 0 \quad \forall \theta \in \Gamma \end{aligned}$$

Observe que neste caso a condição é sempre satisfeita  $\forall \theta \in \Gamma$ .

Este tipo de estrutura pode representar a estrutura ARX (Auto-regressivo com entrada exógena) se os vetores  $F(q)$  e  $B(q)$  forem parametrizados da seguinte forma

$$B(q) = \begin{bmatrix} q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_b} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \text{ e } F(q) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_f} \end{bmatrix} .$$

É amplamente sabido que para a estrutura ARX a função custo é convexa e portanto pode-se obter de forma eficiente o mínimo da função custo. O resultado acima mostra que mesmo para estruturas mais complexa (com qualquer  $B(q)$  e  $F(q)$ ) a condição (4.1) é satisfeita, o que torna a convergência para o mínimo global mais fácil.

#### 4.4.2 Estrutura ARMAX generalizada

Caso a estrutura seja tal que  $F(q) = D(q)$ ,

$$\begin{aligned} (\theta - \theta^0)^T \nabla \bar{V}(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \left[ \Phi_u(\omega) |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \right. \\ &\quad \left. + \sigma_0^2 |H_0(e^{j\omega})(H^{-1}(e^{j\omega}, \theta) - H_0^{-1}(e^{j\omega}))|^2 \right] \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} \right\} d\omega. \end{aligned}$$

Observe que a positividade da integral depende do termo

$$\Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} \right\}$$

já que se ele for positivo a integral também é positiva. Quando  $\theta$  está próximo de  $\theta^0$  este termo é próximo de um, o que faz a integral ser positiva.

Esta estrutura pode representar a estrutura ARMAX (Auto-regressivo com média mó-

vel e entrada exógena) se

$$B(q) = \begin{bmatrix} q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_b} \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C(q) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \hline q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_c} \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \text{ e } F(q) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \hline q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_f} \end{bmatrix}.$$

#### 4.4.3 Estrutura OE generalizada

Nesta estrutura os vetores  $C(q) = D(q) = 0$  e o que significa que  $v(t)$  é ruído branco. Neste caso

$$(\theta - \theta^0)^T \nabla \bar{V}(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) |(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \Re \left\{ \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta} \right\} d\omega. \quad (4.6)$$

Observe que a positividade da integral depende do termo

$$\Re \left\{ \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta} \right\} \quad (4.7)$$

já que se ele for positivo a integral também é positiva. Note que quando  $\theta$  está próximo de  $\theta^0$  este termo é próximo de um, o que faz a integral ser positiva.

Este tipo de estrutura pode representar a estrutura OE (Output-Error) se os vetores  $F(q)$  e  $B(q)$  são parametrizados da seguinte forma

$$B(q) = \begin{bmatrix} q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_b} \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \text{ e } F(q) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \hline q^{-1} \\ q^{-2} \\ q^{-3} \\ \vdots \\ q^{-n_f} \end{bmatrix}.$$

Na sequência vamos apresentar alguns resultados importantes sobre a convergência no caso em que o sistema tem a estrutura tipo OE.

**Corolário 1** *O conjunto em que a função custo é semi-convexa não depende dos zeros do processo se for utilizada uma estrutura do tipo OE.*

*Prova:* Note que a condição (4.7) não depende do termo  $B(q)$  o qual corresponde aos zeros do processo. ■

Um outro resultado bastante interessante que pode ser obtido dos resultados apresentados trata dos processos de primeira ordem.

**Corolário 2** *Considere uma estrutura OE de primeira ordem. Se o modelo e o sistema são estáveis então a função custo é semi-convexa em todo o conjunto  $\Gamma$ .*

*Prova:* Vamos definir a função

$$T(e^{j\omega}, \theta) = \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta}.$$

Usando as condições do teorema  $T(e^{j\omega}, \theta)$  tem um polo, um zero, é estável e de fase mínima. Então é fácil mostrar que  $-\frac{\pi}{2}rad < \angle T(e^{j\omega}, \theta) < \frac{\pi}{2}rad$  o que significa que  $\Re\{T(e^{j\omega}, \theta)\} > 0 \forall \omega$ . Como o integrando é positivo, a integral (4.6) também é positiva. ■

O resultado mostra que para todo processo de primeira ordem que é estável não existem problemas de convergência. De outro ponto de vista,  $\Re\{T(e^{j\omega}, \theta)\} > 0$  somente se

$$|\angle(T(e^{j\omega}, \theta))| = |\angle(1 + F^T(e^{j\omega})\theta^0) - \angle(1 + F^T(e^{j\omega})\theta)| < \frac{\pi}{2}rad. \quad (4.8)$$

Esta condição somente pode ser satisfeita se a ordem do processo for dois ou mais. Além disso, a condição (4.8) é violada apenas quando  $\theta$  está distante de  $\theta^0$ .

## 4.5 Parametrização independente

Vamos assumir agora que o modelo possui uma estrutura fixa onde a função de transferência  $G(q, \theta)$  é função de  $\theta$  mas que a função de transferência do ruído é independente do vetor de parâmetros

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + F^T(q)\theta}, \quad H(q, \theta) = \hat{H}(q)$$

onde os vetores  $B(q), F(q) \in \mathbb{R}^n$  são formados por funções de transferência conhecidas e  $\theta \in \mathbb{R}^n$  é o vetor de parâmetros que desejamos encontrar.

Este tipo de modelo pode ser utilizado quando não se deseja identificar o modelo do ruído. Neste caso é utilizada apenas uma estimativa do modelo do ruído  $\hat{H}(q)$  na identificação dos parâmetros do modelo, mas o modelo do ruído em si não será identificado.

Nesta estrutura tanto o numerador como o denominador são funções afins nos parâmetros  $\theta$ , o que simplifica o cálculo das derivadas parciais, pois

$$\begin{aligned} \nabla G(q, \theta) &= \frac{B(q)}{1 + F^T(q)\theta} - \frac{B^T(q)\theta F(q)}{(1 + F^T(q)\theta)^2}, \\ \nabla H^{-1}(q, \theta) &= 0. \end{aligned}$$

Utilizando esta estrutura de modelo podemos obter condições para que a função custo seja semi-convexa em um conjunto.

**Teorema 4.3** A função  $V(\theta)$  é semi-convexa no conjunto  $\mathcal{C}$  se

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \left| \hat{H}^{-1}(q) (G_0 - G(\theta)) \right|^2 \Re \left\{ \frac{(1 + F^T \theta^0)}{(1 + F^T \theta)} \right\} d\omega > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^0 \quad (4.9)$$

□

*Prova:*

Para garantir que a função  $V(\theta)$  seja semi-convexa no conjunto  $\mathcal{C}$  devemos garantir que

$$(\theta - \theta^0)^T \nabla \bar{V}(\theta) > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^0.$$

Necessitamos portanto calcular  $\nabla \bar{V}(\theta)$ ,

$$\begin{aligned} \nabla \bar{V}(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u \nabla \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right)^* \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right) d\omega \\ &\quad + \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right)^* \nabla \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right) d\omega \end{aligned}$$

que pode ser rearranjado como

$$\nabla \bar{V}(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} 2\Re \left\{ \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right)^* \nabla \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right) \right\} \Phi_u d\omega.$$

e simplificado como

$$\nabla \bar{V}(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \Re \left\{ \left( \hat{H}^{-1}(q) (G_0 - G(\theta)) \right)^* \left( -\hat{H}^{-1}(q) \nabla G(\theta) \right) \right\} d\omega.$$

Vamos agora calcular

$$\begin{aligned} (\theta - \theta^0)^T \nabla \bar{V}(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Re \left\{ \left[ \hat{H}^{-1}(q) (G_0 - G(\theta)) \right]^* \right. \\ &\quad \left. \left[ \hat{H}^{-1}(q) \left( -\frac{B^T(\theta - \theta^0)}{1 + F^T \theta} + \frac{B^T \theta F^T (\theta - \theta^0)}{(1 + F^T \theta)^2} \right) \right] \right\} \Phi_u d\omega \end{aligned}$$

a qual pode ser reescrita como

$$\begin{aligned} (\theta - \theta^0)^T \nabla \bar{V}(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Re \left\{ \left[ \hat{H}^{-1}(q) (G_0 - G(\theta)) \right]^* \right. \\ &\quad \left. \left[ \hat{H}^{-1}(q) \left( \frac{-(1 + F^T \theta) B^T (\theta - \theta^0) + B^T \theta F^T (\theta - \theta^0)}{(1 + F^T \theta)^2} \right) \right] \right\} \Phi_u d\omega. \end{aligned}$$

Utilizando a hipótese  $G(\theta^0) = G_0 = G(\theta^0)$  pode-se escrever

$$\begin{aligned} (\theta - \theta^0)^T \nabla \bar{V}(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Re \left\{ \left[ \hat{H}^{-1}(q) (G_0 - G(\theta)) \right]^* \right. \\ &\quad \left. \left[ \hat{H}^{-1}(q) \left( \frac{(1 + F^T \theta^0)}{(1 + F^T \theta)} \right) \left( G_0 - \frac{B^T \theta}{1 + F^T \theta} \right) \right] \right\} \Phi_u d\omega \end{aligned}$$

que pode ser simplificada para

$$(\theta - \theta^0)^T \nabla \bar{V}(\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u \left| \hat{H}^{-1}(q) (G_0 - G(\theta)) \right|^2 \Re \left\{ \frac{(1 + F^T \theta^0)}{(1 + F^T \theta)} \right\} d\omega.$$

Para que a função  $\bar{V}(\theta)$  seja semi-convexa no conjunto  $\mathcal{C}$  é necessário que a equação acima seja positiva para todo  $\theta$  dentro do conjunto com exceção do mínimo global  $\theta^0$ . Note que esta é a condição do enunciado e portanto está demonstrado o teorema. ■

Este teorema mostra que é possível convergir para o modelo  $G(q, \theta^0)$  mesmo quando o modelo do ruído é desconhecido. No teorema foi utilizado um modelo para o ruído  $\hat{H}(q)$  o qual não é necessariamente  $H_0(q)$ . Mesmo neste caso é possível encontrar um algoritmo que converge para  $\theta^0$  e portanto obter uma estimativa adequada para  $G(q, \theta)$ .

## 4.6 Considerações finais

Este capítulo apresentou condições inéditas para que a função  $\bar{V}(\theta)$  seja semi-convexa em um conjunto  $\mathcal{C}$ , considerando diversas estruturas para o modelo. Todas as condições obtidas estão na forma de uma função integral, a qual deve ser positiva para todo ponto dentro de um conjunto, para garantir que função  $V(\theta)$  seja semi-convexa neste conjunto. Estas funções integrais dependem de vários fatores: das funções de transferência do sistema ( $G_0(q)$  e  $H_0(q)$ ), da estrutura do modelo ( $G(q, \theta)$  e  $H(q, \theta)$ ), da variância do ruído  $\lambda_0$  e do espectro do sinal de entrada  $\Phi_u(\omega)$ .

Nos próximos capítulos vamos analisar mais profundamente como o espectro do sinal de entrada  $\Phi_u(\omega)$  influencia o conjunto em que a função  $\bar{V}(\theta)$  é semi-convexa. Vamos mostrar que é sempre possível escolher um espectro para o sinal de entrada o qual garante que a função custo seja semi-convexa em um dado conjunto. Além disso, vamos mostrar como escolher o espectro do sinal de entrada para ampliar o tamanho do conjunto em que a função  $\bar{V}(\theta)$  é semi-convexa e como melhorar a convergência para o mínimo global da função custo.

Nos procedimentos para a identificação de sistemas, a escolha do sinal que será utilizado em um experimento é chamada de *projeto de experimentos*. No próximo capítulo vamos apresentar o projeto de experimentos no formato de um problema de otimização. Após, no capítulo seguinte, vamos apresentar restrições para o projeto de experimentos as quais garantem que a função  $\bar{V}(\theta)$  seja semi-convexa em um certo conjunto pré-escolhido. Desta forma, podemos projetar um sinal de entrada para o experimento que facilitará a convergência dos algoritmos para o mínimo global da função custo.

## 5 PROJETO DE EXPERIMENTOS

O projeto do experimento consiste em fazer diversas escolhas sobre o experimento que será realizado para coletar os dados que serão utilizados na identificação. De uma certa forma, as variáveis associadas ao projeto de experimento são mais cruciais que as outras escolhas que devem ser feitas em todo o processo de identificação. Enquanto que a estrutura de modelo e os métodos de identificação podem ser escolhidos e testados diversas vezes através do uso do computador, os dados obtidos durante o experimento só podem ser trocados se realizado um novo experimento, o que constitui o principal custo e o fator limitante na aplicação de métodos de identificação.

A primeira escolha consiste em definir quais os sinais serão as saídas e as entradas do processo, ou seja, onde posicionar os sensores e quais sinais devem ser manipulados de forma a excitar o sistema durante o experimento. Outra decisão importante a ser tomada é quando medir. Geralmente os sinais são amostrados usando um tempo de amostragem  $T$  constante, valor que deve ser escolhido pelo usuário.

Os sinais de entrada determinam o ponto de operação e quais as partes e os modos do sistema que são excitados durante o experimento. A liberdade que o usuário tem para escolher os sinais de entrada pode variar consideravelmente dependendo do processo a ser identificado. Em processos industriais, por exemplo, pode não ser permitido manipular o sistema quando este está em modo de operação. Em sistemas econômicos e ecológicos pode ser impossível afetar o sistema com o objetivo de identificação. Já em aplicações didáticas ou durante fases de desenvolvimento de um equipamento novo, a escolha do sinal de entrada só é restringida por limitações físicas. Todas estas restrições devem ser consideradas ao projetar o sinal de entrada de um experimento.

Será visto neste capítulo que muitas das restrições dos sinais de entrada são feitas em relação ao espectro do sinal, ou seja, na sua representação frequencial. Logo, é comum fazer o projeto do sinal de entrada no domínio da frequência. Depois que o espectro é projetado é necessário escolher a maneira que este espectro será realizado no domínio do tempo. É comum realizar os sinais como soma de senóides ou como ruído branco filtrado.

Por fim, ainda é necessário escolher o tamanho do experimento  $N$ , ou seja, o número de medidas de sinais de entrada e saída que serão coletados.

### 5.1 Projeto do sinal de entrada

Uma forma de escolher o melhor sinal para um experimento é descrever o que se deseja na forma de um problema de otimização. Podemos escrever o seguinte problema de otimização conceitual para fazer a escolha do sinal de entrada (HJALMARSSON,

2005):

$$\begin{aligned} & \min_{\Phi_u(\omega)} \text{objetivo} & (5.1) \\ & \text{sujeito a} \text{ restrições de qualidade,} \\ & \text{restrições no sinal e} \\ & \text{restrições de convergência.} \end{aligned}$$

Muitas vezes o *objetivo* está relacionado com a qualidade da estimativa do modelo que será obtido com os dados gerados pelo projeto de experimentos. Contudo, outros objetivos podem também ser utilizados e é comum minimizar a energia do sinal de entrada. Diversas restrições podem ser impostas ao problema, das quais podemos citar as restrições nos sinais, na qualidade da estimativa e na convergência do método de identificação. As restrições de sinal incluem restrições na energia do sinal, na amplitude do sinal no domínio do tempo e na amplitude do espectro em certas faixas de frequência. As restrições de qualidade tratam da qualidade do modelo que será obtido na etapa de identificação. Já as restrições de convergência serão propostas nesta tese com o objetivo de facilitar a convergência dos métodos de identificação para o mínimo global.

Embora pareça natural o problema de otimização apresentado, ele é intratável, por diversas razões:

- o problema possui dimensão infinita;
- as restrições são tipicamente não-convexas;
- as condições dependem dos parâmetros reais do sistema que são desconhecidos.

A fim de solucionar os dois primeiros itens, algumas adaptações são realizadas em (JANSSON, 2004), de forma que o projeto de experimentos possa ser solucionado. Para isso, o espectro do sinal de entrada é parametrizado de forma linear e de dimensão finita com o intuito de eliminar o primeiro dos problemas. As restrições podem muitas vezes ser reescritas de forma convexa, o que elimina o segundo dos problemas.

Considerando o terceiro problema, diversas condições de qualidade e de convergência dependem dos parâmetros reais do sistema que não são conhecidos. Se estes parâmetros fossem conhecidos, não seria necessário identificar um modelo para o sistema. Na prática, assume-se que está disponível uma estimativa para o modelo que não é suficientemente precisa. Esta estimativa é utilizada no projeto de experimentos para fazer a escolha do espectro do sinal de entrada. Este sinal de entrada pode então ser utilizado em procedimento de identificação para obter uma estimativa mais precisa para os parâmetros do modelo. Portanto, estimativas dos parâmetros do modelo são utilizadas nas condições que dependem dos parâmetros reais.

Genericamente, o espectro do sinal de entrada pode ser descrito por

$$\Phi_u(\omega) = \sum_{k=1}^{\infty} c_k \mathbf{B}_k(e^{j\omega}) \quad (5.2)$$

onde as funções de base  $\mathbf{B}_k(e^{j\omega})$  são escolhidas pelo usuário e  $c_k \in \mathbb{R}$ . Note que existem infinitos parâmetros  $c_k$  nesta parametrização, o que torna a dimensão desta parametrização infinita. Para tornar a dimensão finita é proposta a seguinte aproximação para o espectro do sinal de entrada:

$$\Phi_u(\omega) = \sum_{k=1}^M c_k \mathbf{B}_k(e^{j\omega}) \quad (5.3)$$

onde agora o somatório possui apenas  $M$  termos. Esta aproximação é utilizada para simplificar o problema de otimização e tornar factível a sua solução.

Note contudo que não é qualquer combinação de valores  $c_k$  que pode ser utilizada. O espectro de um sinal possui a seguinte propriedade

$$\Phi_u(\omega) \geq 0, \quad \forall \omega. \quad (5.4)$$

Logo é necessário impor restrições nos valores de  $c_k$  para garantir que o espectro obtido possa ser realizado. Nas próximas seções vamos apresentar algumas escolhas usuais para a base  $\mathbf{B}_k(e^{j\omega})$  e algumas restrições aos valores  $c_k$  para garantir a positividade do espectro. Mas antes disso vamos apresentar algumas restrições usuais utilizadas nos espectros.

Uma das restrições mais usuais está relacionada com a potência média do sinal (BOMBOIS et al., 2006). A restrição da potência média do sinal de entrada pode ser expressa por

$$\bar{E}(u(t))^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega < \gamma \quad (5.5)$$

enquanto que a restrição da potência média do sinal de saída pode ser dada por

$$\bar{E}(y(t))^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_y(\omega) d\omega < \gamma. \quad (5.6)$$

Podemos incluir restrições na amplitude do espectro, as quais podem ser genericamente expressas por

$$\alpha(\omega) \leq \Phi(\omega) \leq \beta(\omega) \quad \forall \omega. \quad (5.7)$$

Outro tipo de restrição muito utilizada depende da matriz  $P_\theta$ , a qual representa a variância da estimativa que será obtida da identificação. Algumas das restrições mais usuais são (MEHRA, 1974)

- L-otimalidade

$$\begin{aligned} Tr(WP_\theta) &< \gamma \\ W &= VV^* \\ &\Downarrow \\ Tr(V^*P_\theta V) &< \gamma \\ &\Downarrow \\ Tr(Z) &< \gamma \\ V^*P_\theta V &< Z \\ &\Downarrow \\ Tr(Z) &< \gamma \\ \begin{bmatrix} Z & V \\ V^* & P_\theta^{-1} \end{bmatrix} &> 0 \end{aligned}$$

onde foi utilizado o complemento de Schur (BOYD et al., 1994).

- A-otimalidade -  $Tr(P_\theta) < \gamma$

Basta substituir  $V$  pela matriz identidade na condição de L-otimalidade.

- E-otimalidade -  $\lambda_{max}(P_\theta) < \gamma$

$$\begin{aligned} \lambda_{max}(P_\theta) &< \gamma \\ \Downarrow \\ \gamma I - P_\theta &> 0 \\ \Downarrow \\ \begin{bmatrix} \gamma I & I \\ I & P_\theta^{-1} \end{bmatrix} &> 0 \end{aligned}$$

- D-otimalidade -  $\det(P_\theta) < \gamma$

$$\begin{aligned} \det(P_\theta) &< \gamma \\ \Downarrow \\ \log(\det(P_\theta)) &< \log(\gamma) \\ \Downarrow \\ -\log(\det(P_\theta^{-1})) &< \log(\gamma) \end{aligned}$$

Para todas as restrições apresentadas na matriz de covariância  $P_\theta$  existe uma condição convexa equivalente, a qual depende da inversa  $P_\theta^{-1}$ . Como a inversa da matriz de covariância (2.16) é afim ao espectro do sinal de entrada  $\Phi_u(\omega)$  e o espectro é afim às variáveis de decisão  $c_k$ , então as restrições na matriz de covariância podem ser implementadas de forma convexa no problema de otimização do projeto de experimentos. As restrições de convergência, as quais são propostas nesta tese, serão apresentadas no capítulo seguinte.

Vamos apresentar a seguir duas bases  $\mathbf{B}_k(e^{j\omega})$  diferentes para a parametrização do espectro. Serão apresentadas realizações para estes espectros e condições convexas para as diversas restrições apresentadas nesta seção.

## 5.2 Espectro discreto

Uma das parametrizações mais utilizadas para o espectro do sinal de entrada consiste em escolher

$$\mathbf{B}_k(e^{j\omega}) = \delta(\omega - \omega_k) + \delta(\omega + \omega_k)$$

de maneira que

$$\Phi_u(\omega) = \sum_{k=1}^M c_k (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)). \quad (5.8)$$

Nestas expressões  $\delta(\cdot)$  representa a função *delta de dirac* e portanto o espectro  $\Phi_u(\omega)$  é discreto.

Podemos realizar este espectro no domínio do tempo da seguinte maneira

$$u(t) = \sum_{k=1}^M \sqrt{\frac{2c_k}{\pi}} \sin(\omega_k t + \phi_k). \quad (5.9)$$

onde o parâmetro  $\phi_k \in \mathbb{R}$  pode ser livremente escolhido.

### 5.2.1 Positividade do espectro

Para que o espectro (5.8) seja realizável é necessário garantir que o mesmo seja não negativo para todas frequências. Isto pode ser garantido utilizando as seguintes restrições:

$$c_k \geq 0 \quad k = 1, \dots, M.$$

Note que estas restrições são convexas e podem ser incluídas no projeto de experimentos.

### 5.2.2 Restrição em potência média

A restrição da potência média do sinal de entrada é dada por

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k=1}^M c_k (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)) d\omega < \gamma$$

que pode ser escrita como

$$\frac{1}{2\pi} \sum_{k=1}^M c_k \left( \int_{-\pi}^{\pi} \delta(\omega - \omega_k) + \delta(\omega + \omega_k) d\omega \right) < \gamma.$$

A restrição na potência média do sinal de entrada pode ser simplificada como

$$\frac{1}{\pi} \sum_{k=1}^M c_k < \gamma. \quad (5.10)$$

pois

$$\int_{-\pi}^{\pi} \delta(\omega - \omega_k) + \delta(\omega + \omega_k) d\omega = 2.$$

A restrição (5.10) pode ser incluída no projeto de experimentos já que é convexa.

Já a restrição na potência média do sinal de saída é dada por

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_y(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} (|G_0(e^{j\omega})|^2 \Phi_u(\omega) + \Phi_e(\omega)) d\omega < \gamma$$

que pode ser escrita como

$$\sum_{k=1}^M c_k \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |G_0(e^{j\omega})|^2 (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)) d\omega \right) + \lambda_0 < \gamma.$$

Esta condição pode ser simplificada como

$$\lambda_0 + \sum_{k=1}^M c_k \mathbf{B}_k^y < \gamma \quad (5.11)$$

onde

$$\mathbf{B}_k^y = \int_{-\pi}^{\pi} |G_0(e^{j\omega})|^2 (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)) d\omega = 2 \sum_{k=1}^M |G_0(e^{j\omega_k})|^2.$$

Note que a restrição (5.11) também é convexa e pode ser incluída no projeto de experimentos. Contudo, esta restrição depende de  $G_0(e^{j\omega})$  o qual é desconhecido. Para utilizar esta condição é necessário obter uma estimativa para  $G_0(e^{j\omega})$ .

### 5.2.3 Restrição na amplitude do espectro

As restrições na amplitude do espectro podem ser implementadas como

$$\alpha_k < c_k < \beta_k \quad k = 1, \dots, M.$$

visto que as condições são convexas.

### 5.2.4 Restrição na matriz de covariância

Diversas restrições de qualidade são expressas em função da matriz de covariância  $P_\theta$ . Foi mostrado anteriormente que diversas restrições possuem condições equivalentes que são convexas com relação à inversa de  $P_\theta$ . A equação (2.16) mostra que  $P_\theta^{-1}$  é uma função afim no espectro no sinal de entrada. Utilizando (5.8) podemos simplificar a expressão (2.16) como

$$P_\theta^{-1} = R_0 + \sum_{k=1}^M c_k \mathbf{B}_k^P \quad (5.12)$$

onde

$$\begin{aligned} \mathbf{B}_k^P &= \frac{1}{2\pi\lambda_0} \int_{-\pi}^{\pi} \mathcal{F}_u(e^{j\omega}, \theta^0) (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)) \mathcal{F}_u^*(e^{j\omega}, \theta^0) d\omega \\ \mathbf{B}_k^P &= \frac{1}{2\pi\lambda_0} \sum_{k=1}^M \mathcal{F}_u(e^{j\omega_k}, \theta^0) \mathcal{F}_u^*(e^{j\omega_k}, \theta^0) + \mathcal{F}_u(e^{-j\omega_k}, \theta^0) \mathcal{F}_u^*(e^{-j\omega_k}, \theta^0) \\ \mathbf{B}_k^P &= \frac{1}{\pi\lambda_0} \sum_{k=1}^M \Re \{ \mathcal{F}_u(e^{j\omega_k}, \theta^0) \mathcal{F}_u^*(e^{j\omega_k}, \theta^0) \}. \end{aligned}$$

A representação (5.12) para a inversa da matriz de covariância  $P_\theta^{-1}$  é afim nas variáveis  $c_k$  e portanto pode ser utilizada no projeto de experimentos gerando uma restrição convexa. Contudo, esta restrição depende de  $\theta^0$ , o qual é desconhecido. Para utilizar esta condição é necessário obter uma estimativa para  $\theta^0$ .

## 5.3 Espectro contínuo

Outra parametrização muito utilizada para o espectro do sinal de entrada consiste em escolher

$$\mathbf{B}_k(e^{j\omega}) = e^{j(k-1)\omega} + e^{-j(k-1)\omega} = 2\cos((k-1)\omega)$$

de maneira que

$$\Phi_u(\omega) = 2 \sum_{k=1}^M c_k \cos((k-1)\omega). \quad (5.13)$$

Note que neste caso o espectro  $\Phi_u(\omega)$  é contínuo.

O espectro (5.13) pode ser modelado por um filtro do tipo FIR e portanto ele pode ser realizado utilizando um filtro FIR de ordem  $M$  e ruído branco. O sinal de entrada pode ser obtido a partir de

$$u(t) = w(q)e(t)$$

onde  $e(t)$  é ruído branco com variância 1 e  $w(q)$  é a função de transferência de um filtro do tipo FIR. Como o ruído branco possui mesma amplitude em todas as frequências, basta

escolher um filtro tal que  $|W(e^{j\omega})|^2 = \Phi_u(\omega)$ . Isto pode ser feito utilizando uma fatoração espectral.

Considere que o filtro é

$$W(e^{j\omega}) = \sum_{k=1}^M a_k e^{j(k-1)\omega}.$$

Para encontrar os parâmetros  $a_k$  basta resolver a seguinte equação

$$\begin{aligned} |W(e^{j\omega})|^2 &= W(e^{j\omega})W(e^{-j\omega}) = \Phi_u(\omega) \\ \left( \sum_{k=1}^M a_k e^{j(k-1)\omega} \right) \left( \sum_{k=1}^M a_k e^{-j(k-1)\omega} \right) &= \sum_{k=1}^M c_k (e^{j(k-1)\omega} + e^{-j(k-1)\omega}). \end{aligned}$$

Como existem várias soluções para os parâmetros  $a_k$  que respeitam a equação acima, é usual escolher o conjunto de parâmetros que gera um filtro com fase mínima, ou seja, todos os zeros do filtro estão dentro do círculo unitário (ÅSTRÖM, 1970; LJUNG, 1987).

### 5.3.1 Positividade do espectro

Para que o espectro (5.13) seja realizável é necessário garantir que o mesmo seja não negativo para todas frequências. Uma condição convexa que garante a positividade do espectro é apresentada no seguinte Lema.

**Lema 5.1** *Existe  $Q = Q^T$  tal que*

$$\begin{bmatrix} Q - A^T Q A & -A^T Q B \\ -B^T Q A & -B^T Q B \end{bmatrix} + \begin{bmatrix} 0 & C^T \\ C & D + D^T \end{bmatrix} \geq 0 \quad (5.14)$$

onde

$$\begin{aligned} A &= \begin{bmatrix} 0_{1 \times M-2} & 0 \\ I_{M-2} & 0_{M-2 \times 1} \end{bmatrix} \\ B &= [1 \ 0 \ \dots \ 0]^T \\ C &= [c_2 \ c_3 \ \dots \ c_M] \\ D &= c_1 \end{aligned}$$

se e somente se

$$\Phi_u(\omega) = 2 \sum_{k=1}^M c_k \cos((k-1)\omega) \geq 0 \quad \forall \omega$$

□

*Prova:* Basta utilizar o Lema KYP (A.1) onde  $v = 1$ ,  $u = [e^{j\omega} \ e^{j2\omega} \ \dots \ e^{jM\omega}] e$   
 $M = \begin{bmatrix} 0 & C^T \\ C & D + D^T \end{bmatrix}$ . ■

Na condição (5.14), as variáveis livres  $c_k$  estão presentes apenas nos termos  $C$  e  $D$ , o que torna a condição (5.14) uma LMI (Linear Matrix Inequality) (BOYD et al., 1994). Para simplificar a notação vamos escrever a condição (5.14) como

$$K(Q, A, B, C, D) \geq 0.$$

### 5.3.2 Restrição em potência média

A restrição da potência média do sinal de entrada é dada por

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega = \frac{1}{\pi} \int_{-\pi}^{\pi} \left( \sum_{k=1}^M c_k \cos((k-1)\omega) \right) d\omega < \gamma$$

que pode ser escrita como

$$\frac{1}{\pi} \sum_{k=1}^M c_k \left( \int_{-\pi}^{\pi} \cos((k-1)\omega) d\omega \right) < \gamma.$$

Note que

$$\int_{-\pi}^{\pi} \cos(0\omega) d\omega = 2\pi \quad \text{e} \quad \int_{-\pi}^{\pi} \cos(k\omega) d\omega = 0 \quad \forall k > 1$$

e portanto a restrição na potência média do sinal de entrada pode ser simplificada como

$$2c_1 < \gamma. \quad (5.15)$$

A restrição (5.15) é convexa e pode ser incluída no projeto de experimentos.

Já a restrição na potência média do sinal de saída é dada por

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_y(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G_0(e^{j\omega})|^2 \Phi_u(\omega) + \Phi_e(\omega) d\omega < \gamma$$

que pode ser escrita como

$$\sum_{k=1}^M c_k \left( \frac{1}{\pi} \int_{-\pi}^{\pi} |G_0(e^{j\omega})|^2 \cos((k-1)\omega) \right) + \lambda_0 < \gamma.$$

Esta condição pode ser simplificada como

$$\lambda_0 + \sum_{k=1}^M c_k \mathbf{B}_k^y < \gamma \quad (5.16)$$

onde

$$\mathbf{B}_k^y = \frac{1}{\pi} \int_{-\pi}^{\pi} |G_0(e^{j\omega})|^2 \cos((k-1)\omega) d\omega.$$

A restrição (5.16) também é convexa e pode ser incluída no projeto de experimentos. Contudo, esta restrição depende de  $G_0(e^{j\omega})$ , a qual é desconhecida. Para utilizar esta condição é necessário obter uma estimativa para  $G_0(e^{j\omega})$ .

### 5.3.3 Restrição na amplitude do espectro

Se as restrições na amplitude do espectro podem ser expressas por

$$\alpha(\omega) \leq \Phi(\omega) \leq \beta(\omega) \quad \forall \omega \quad (5.17)$$

onde  $\alpha(\omega)$  e  $\beta(\omega)$  são funções racionais, então podemos implementar as restrições de maneira similar ao que foi feito para garantir a positividade do espectro.

Precisamos adicionar duas condições extras:

$$\Phi(\omega) - \alpha(\omega) \geq 0 \quad \forall \omega, \quad (5.18)$$

$$\beta(\omega) - \Phi(\omega) \geq 0 \quad \forall \omega, \quad (5.19)$$

as quais podem ser representadas de maneira convexa utilizando LMIs como foi feito no Lema 5.1.

### 5.3.4 Restrição na matriz de covariância

Diversas restrições de qualidade são expressas em função da matriz de covariância  $P_\theta$ . Foi mostrado anteriormente que diversas restrições possuem condições equivalentes que são convexas com relação à inversa de  $P_\theta$ . A equação (2.16) mostra que  $P_\theta^{-1}$  é uma função afim no espectro no sinal de entrada. Utilizando (5.13) podemos simplificar a expressão (2.16) como

$$P_\theta^{-1} = R_0 + \sum_{k=1}^M c_k \mathbf{B}_k^P \quad (5.20)$$

onde

$$\mathbf{B}_k^P = \frac{1}{\pi \lambda_0} \int_{-\pi}^{\pi} \mathcal{F}_u(e^{j\omega}, \theta^0) \cos((k-1)\omega) \mathcal{F}_u^*(e^{j\omega}, \theta^0) d\omega$$

A representação (5.20) para a inversa da matriz de covariância  $P_\theta^{-1}$  é afim nas variáveis  $c_k$  e portanto pode ser utilizada no projeto de experimentos gerando uma restrição convexa. Contudo, esta restrição depende de  $\theta^0$  o qual é desconhecido. Para utilizar esta condição é necessário obter uma estimativa para  $\theta^0$ .

## 5.4 Exemplos

Neste capítulo serão apresentados exemplos que demonstram como utilizar os lemas e teoremas descritos nas seções anteriores.

### 5.4.1 Sistema tipo FIR

Seja o sistema verdadeiro

$$\mathcal{S} : y(t) = u(t-1) + 0,8u(t-2) + e_0(t) \quad (5.21)$$

onde  $e_0(t)$  é ruído branco com variância 1.

Desejamos encontrar os parâmetros da seguinte estrutura

$$y(t) = \theta_1 u(t-1) + \theta_2 u(t-2) + e_0(t). \quad (5.22)$$

Gostaríamos de encontrar o espectro com menor energia cujo sinal possa ser usado para estimar os parâmetros  $\theta_1$  e  $\theta_2$  com precisão melhor que 0,01 em um intervalo de confiança de 95%. Além disso, gostaríamos de utilizar um experimento de duração  $N = 5000$  amostras.

Inicialmente vamos parametrizar o espectro do sinal utilizando parametrização espectral de dimensão finita com base  $\mathbf{B}_k = e^{j\omega k}$ .

$$\Phi_u(\omega) = \sum_{k=1}^M c_k [e^{j(k-1)\omega} + e^{-j(k-1)\omega}]$$

Utilizando este espectro temos que

$$\mathbf{B}_k^P = \frac{1}{\pi \lambda_0} \int_{-\pi}^{\pi} \mathcal{F}_u(e^{j\omega}, \theta^0) \cos((k-1)\omega) \mathcal{F}_u^*(e^{j\omega}, \theta^0) d\omega$$

Note que  $\mathbf{B}_k^P = \mathbf{0}$  para todo  $k > 2$ , portanto vamos escolher  $M = 2$ .

Logo,

$$P_{\theta}^{-1} = c_1 \mathbf{B}_1^P + c_2 \mathbf{B}_2^P$$

pois  $R_0 = 0$ .

Para garantir a positividade do espectro é utilizado o Lema 5.1 com as seguintes matrizes:

$$A = 0 \quad (5.23)$$

$$B = 1 \quad (5.24)$$

$$C = c_2 \quad (5.25)$$

$$D = c_1. \quad (5.26)$$

A parametrização da restrição em energia torna-se então

$$2c_1 < \gamma$$

onde  $\gamma$  é o valor que será minimizado.

Para garantir esta restrição de qualidade podemos utilizar a seguinte restrição

$$\lambda_{max}(P_{\theta}) < \frac{(0,01)^2 N}{\chi_{0,95}^2} = \frac{0,0001 * 5000}{5,99} = \frac{50}{599}$$

a qual pode ser escrita de maneira convexa como

$$\begin{bmatrix} \frac{50}{599}I & I \\ I & P_{\theta}^{-1} \end{bmatrix} > 0.$$

Portanto o problema de otimização que deve ser resolvido é o seguinte

$$\begin{aligned} & \min_{c_0, c_1, \gamma, Q} \gamma \\ & \text{sujeito a} \\ & 2c_1 < \gamma \\ & \begin{bmatrix} \frac{50}{599}I & I \\ I & c_1 \mathbf{B}_1^P + c_2 \mathbf{B}_2^P \end{bmatrix} > 0, \\ & K(Q, A, B, C, D) \geq 0. \end{aligned}$$

Utilizando o solver `lmi lab` do MATLAB obtém-se:

$$\begin{aligned} c_1 &= 5,99 \\ c_2 &= 0 \\ Q &= 5,99. \end{aligned}$$

Como apenas  $c_1$  é diferente de zero o sinal ótimo é ruído branco com variância 5,99.

A Figura 5.1 apresenta uma simulação por Monte-Carlo com 1000 rodadas. A figura mostra que as simulações respeitam as restrições do problema.

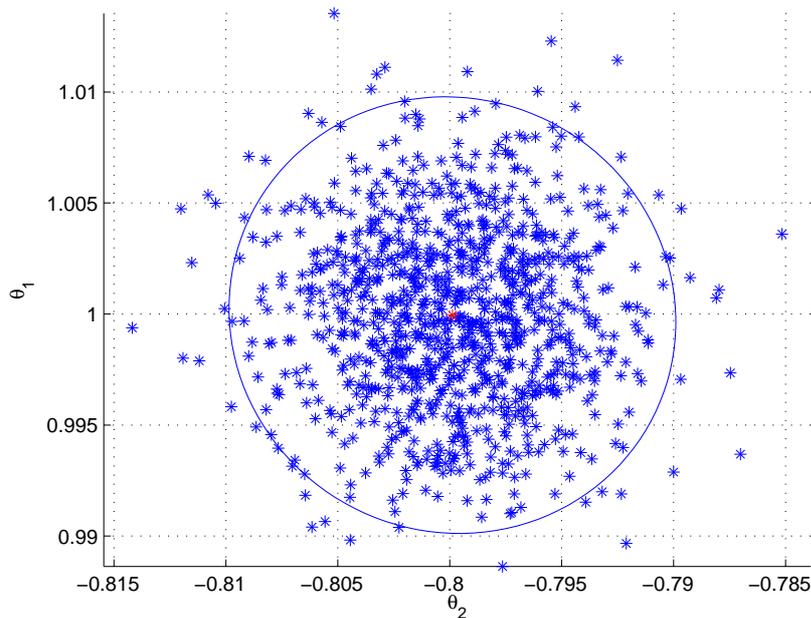


Figura 5.1: Simulação Monte-Carlo e Intervalo de Confiança de 95%.

#### 5.4.2 Sistema tipo ARX

Seja o sistema descrito por

$$\mathcal{S} : y(t) = 0,5u(t-1) + 0,9y(t-1) + e_0(t) \quad (5.27)$$

onde  $e_0(t)$  é ruído branco com variância 1.

Desejamos encontrar os parâmetros da seguinte estrutura

$$y(t) = \theta_1 u(t-1) - \theta_2 y(t-1) + e_0(t). \quad (5.28)$$

Gostaríamos de encontrar o espectro com menor energia cujo sinal possa ser usado para estimar os parâmetros  $\theta_1$  e  $\theta_2$  com precisão melhor que 0,01 em um intervalo de confiança de 95%. Além disso, gostaríamos de utilizar um experimento de duração  $N = 5000$  amostras.

Inicialmente vamos parametrizar o espectro do sinal utilizando parametrização espectral de dimensão finita com base  $\mathbf{B}_k = e^{j\omega k} + e^{-j\omega k}$ .

$$\Phi_u(\omega) = \sum_{k=1}^{20} c_k [e^{j\omega k} + e^{-j\omega k}]$$

Para garantir a positividade do espectro é utilizado o Lema 5.1 com as seguintes matrizes:

$$A = \begin{bmatrix} 0_{1 \times 18} & 0 \\ I_{18} & 0_{18 \times 1} \end{bmatrix} \quad (5.29)$$

$$B = [1 \ 0 \ \cdots \ 0]^T \quad (5.30)$$

$$C = [c_2 \ c_3 \ \cdots \ c_{20}] \quad (5.31)$$

$$D = c_1 \quad (5.32)$$

A parametrização da restrição em energia torna-se então

$$2c_1 < \gamma$$

onde  $\gamma$  é o valor que será minimizado.

Para garantir esta restrição de qualidade podemos utilizar a seguinte restrição

$$\lambda_{max}(P_\theta) < \frac{N(0, 01)^2}{\chi_{0,95}^2} = \frac{0,0001 * 5000}{5,99} = \frac{50}{599}$$

a qual pode ser escrita de maneira convexa como

$$\begin{bmatrix} \frac{50}{599}I & I \\ I & P_\theta^{-1} \end{bmatrix} > 0$$

onde

$$P_\theta^{-1} = R_0 + \sum_{k=1}^{20} c_k \mathbf{B}_k^P.$$

Portanto o problema de otimização que deve ser resolvido é o seguinte

$$\begin{aligned} & \min_{c_1, \dots, c_{20}, \gamma, Q} \gamma \\ & \text{sujeito a} \\ & 2c_1 < \gamma \\ & \begin{bmatrix} \frac{50}{599}I & I \\ I & R_0 + \sum_{k=1}^{20} c_k \mathbf{B}_k^P \end{bmatrix} > 0 \\ & K(Q, A, B, C, D) > 0 \end{aligned}$$

Utilizando o solver `lmi lab` do MATLAB obtém-se que

$$\begin{aligned} [c_1 \cdots c_{20}] = & [6,0018 - 0,1040 - 0,0886 - 0,0754 - 0,0638 - 0,0539 - 0,0452 \\ & -0,0378 - 0,0314 - 0,0259 - 0,0212 - 0,0172 - 0,0137 - 0,0108 \\ & -0,0083 - 0,0063 - 0,0045 - 0,0030 - 0,0018 - 0,0008]. \end{aligned}$$

A Figura 5.2 apresenta uma simulação por Monte-Carlo com 1000 rodadas. A figura mostra que as simulações respeitam as restrições do problema.

A Figura 5.3 apresenta o espectro obtido através da otimização e a Figura 5.4 apresenta uma realização do sinal  $u(t)$ .

## 5.5 Considerações finais

O projeto de experimentos consiste na escolha de diversos elementos do experimento relacionados com a identificação do modelo. O elemento mais importantes é o sinal que será aplicado na entrada do processo. A escolha do sinal pode ser feita resolvendo um problema de otimização. Diversos critérios de qualidade foram apresentados para duas parametrizações do espectro do sinal de entrada. O projeto de experimentos será utilizado para construir a primeira ferramenta para melhorar a convergência dos métodos de identificação por erro de predição. Esta ferramenta nada mais é que restrições para o projeto de experimentos, os quais vão ampliar a região em que a função custo do método de identificação é semi-convexa, e portanto melhorar a convergência dos métodos de identificação.

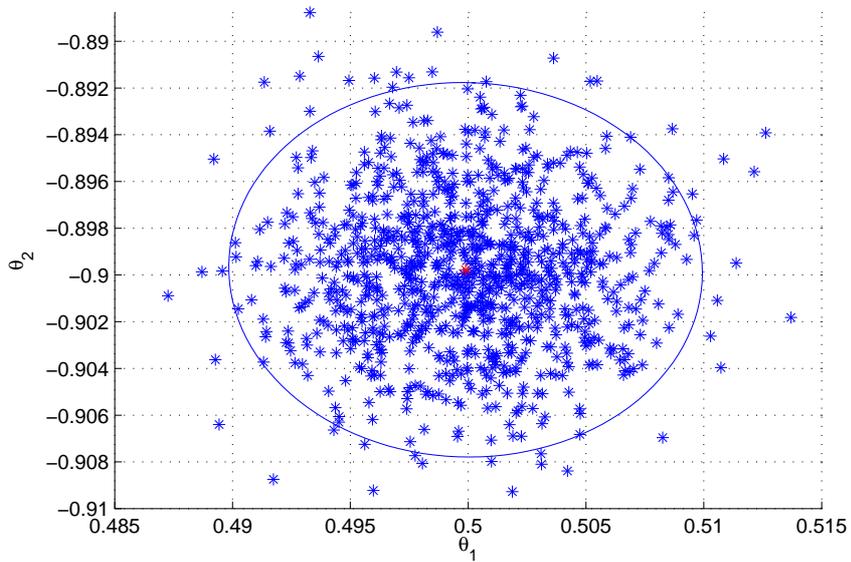


Figura 5.2: Simulação Monte-Carlo e Intervalo de Confiança de 95%.

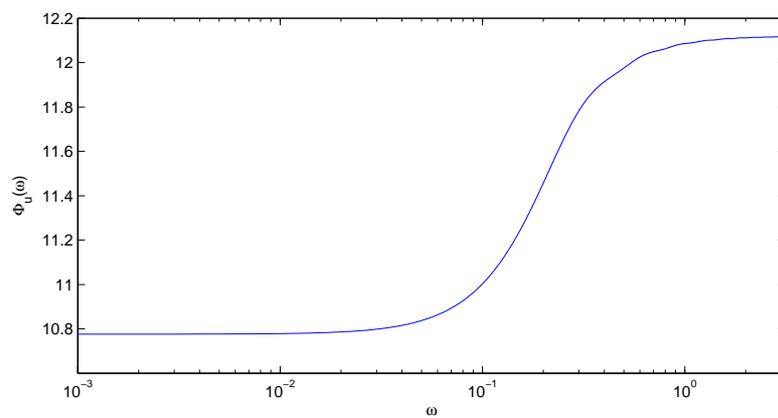


Figura 5.3: Espectro do sinal de entrada

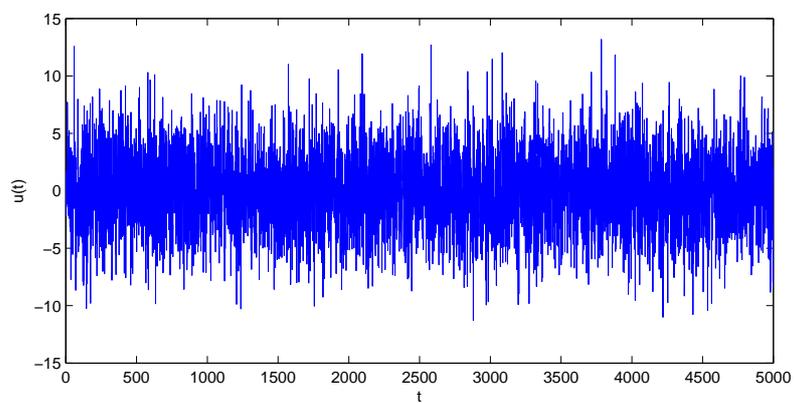


Figura 5.4: Realização do sinal de entrada  $u(t)$

## 6 RESTRIÇÃO DE CONVERGÊNCIA

Nos Capítulos 3 e 4 foram apresentadas condições para que uma função seja semi-convexa em um conjunto. Foi mostrado que a convergência para o mínimo global da função custo é mais fácil se a condição inicial dos algoritmos pertencer ao conjunto em que a função é semi-convexa.

As condições obtidas do Capítulo 4 dependem de vários fatores, dentre os quais foi dada atenção especial ao espectro do sinal de entrada. O espectro  $\Phi_u(\omega)$  possui grande influência sobre o formato da função custo e sobre o domínio que ela é semi-convexa. Escolhendo apropriadamente este espectro é possível garantir que a função custo seja semi-convexa em um certo conjunto  $\mathcal{C}$ . Seguindo esta ideia, vamos neste capítulo apresentar maneiras de projetar o espectro do sinal de entrada para garantir que a função custo seja semi-convexa em um conjunto previamente escolhido.

No Capítulo 5 foi apresentado o projeto de experimentos como uma ferramenta para a escolha do espectro do sinal de entrada. O projeto de experimentos foi descrito como um problema de otimização convexo no qual pode-se incluir diversos tipos de restrições. Foram utilizadas restrições tanto no espectro dos sinais de entrada e saída como na covariância da estimativa que será obtida na etapa de identificação.

Neste capítulo vamos apresentar mais uma contribuição inédita desta tese: um novo tipo de restrição para o projeto de experimentos, o qual é uma das ferramentas para melhoria da convergência dos métodos de identificação por erro de predição.

### 6.1 Restrições infinitas

No Capítulo 3 foi mostrado que uma função  $\bar{V}(\theta)$  é semi-convexa em um conjunto  $\mathcal{C}$  se

$$\nabla \bar{V}^T(\theta)(\theta - \theta^0) > 0 \quad \forall \theta \in \mathcal{C}, \theta \neq \theta^0. \quad (6.1)$$

Esta condição possui algumas propriedades que facilitam a convergência dos algoritmos para o mínimo global da função custo.

Quando o método de identificação pela minimização do erro de predição é utilizado, deseja-se encontrar o mínimo global da função custo  $\bar{V}(\theta)$ , conforme apresentado no capítulo 2.

No Capítulo 4 foi mostrado que  $\nabla \bar{V}^T(\theta)(\theta - \theta^0) > 0$  é equivalente a

$$\begin{aligned} & \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \\ & \quad \cdot \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} + \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta} - \frac{1 + D^T(e^{j\omega})\theta^0}{1 + D^T(e^{j\omega})\theta} \right\} + d\omega \\ & \frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 |H_0(e^{j\omega})(H^{-1}(e^{j\omega}, \theta) - H_0^{-1}(e^{j\omega}))|^2 \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} \right\} d\omega > 0. \end{aligned} \quad (6.2)$$

Utilizando a condição acima é possível garantir que um certo conjunto  $\mathcal{C}$  seja um domínio de atração para o mínimo global  $\theta^0$ . Para tanto, basta garantir que a condição seja satisfeita para todo  $\theta$  que pertence ao conjunto. Seguindo esta ideia, um usuário pode especificar um conjunto  $\mathcal{C}$  no qual ele gostaria que a função custo fosse semi-convexa. Na sequência, este usuário pode verificar a condição (6.2) para todos os pontos do conjunto e então concluir se a função custo é semi-convexa neste conjunto.

Observe que a condição (6.2) depende explicitamente do espectro do sinal de entrada  $\Phi_u(\omega)$ . Desta forma, pode-se escolher tal espectro para garantir que a condição seja satisfeita para todos os pontos de  $\mathcal{C}$ .

No capítulo 5 foi apresentado o projeto de experimentos como uma ferramenta para a escolha do espectro do sinal de entrada  $\Phi_u(\omega)$ . Podemos portanto utilizar o projeto de experimentos para fazer a escolha do espectro do sinal de entrada de forma sistemática. O projeto de experimentos foi apresentado como um problema de otimização no qual pode-se incluir diversas restrições. Vamos neste capítulo apresentar algumas novas restrições que estão relacionadas ao domínio em que a função custo é semi-convexa.

Relembre que no projeto de experimentos o espectro do sinal de entrada foi parametrizado da seguinte forma:

$$\Phi_u(\omega) = \sum_{k=1}^M c_k \mathbf{B}_k(e^{j\omega}).$$

Utilizando esta expressão para o espectro, e fixando  $\theta = \theta^1$  podemos escrever a condição (6.2) como

$$\phi(\theta^1) + \sum_{k=1}^M c_k \psi_k(\theta^1) > 0 \quad (6.3)$$

onde

$$\begin{aligned} \psi_k(\theta^1) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \mathbf{B}_k(e^{j\omega}) |H^{-1}(\theta^1)(G_0 - G(\theta^1))|^2 \\ & \quad \cdot \Re \left\{ \frac{1 + C^T\theta^0}{1 + C^T\theta^1} + \frac{1 + F^T\theta^0}{1 + F^T\theta^1} - \frac{1 + D^T\theta^0}{1 + D^T\theta^1} \right\} d\omega \end{aligned}$$

e

$$\phi(\theta^1) = \lambda_0 \frac{1}{\pi} \int_{-\pi}^{\pi} |H_0(H^{-1}(\theta^1) - H_0^{-1})|^2 \Re \left\{ \frac{1 + C^T\theta^0}{1 + C^T\theta^1} \right\} d\omega.$$

A condição (6.3) é convexa nas variáveis  $c_k$  e portanto pode ser incluída no projeto de experimentos. Contudo, observe que a condição diz respeito a apenas um valor fixo  $\theta^1$ , e necessitamos garantir a condição para todo  $\theta$  dentro do conjunto  $\mathcal{C}$ . Como existem infinitos pontos dentro do conjunto  $\mathcal{C}$ , a princípio precisamos incluir infinitas restrições no projeto de experimentos para garantir que a função  $\bar{V}(\theta)$  seja semi-convexa no conjunto

$\mathcal{C}$ . Como esta abordagem não é factível, vamos procurar por aproximações para estas condições.

Vamos obter duas aproximações para verificar a condição (6.3)  $\forall \theta \in \mathcal{C}$ . A primeira abordagem é baseada no *Scenario Approach* (CAMPI; GARATTI; PRANDINI, 2009) a qual apresenta restrições aleatórias que são necessárias para garantir que a condição seja satisfeita com uma probabilidade pré-definida. A segunda abordagem é baseada em *S-procedure* a qual apresenta condições suficientes.

## 6.2 Scenario Approach

Para que a função  $\bar{V}(\theta)$  seja semi-convexa em um conjunto  $\mathcal{C}$  escolhido pelo usuário, é preciso garantir que a condição (6.3) seja válida para todo  $\theta \in \mathcal{C}$ . Como existem infinitos pontos dentro do conjunto  $\mathcal{C}$  precisamos de infinitas restrições para alcançar o objetivo. A Figura 6.1 apresenta um exemplo de conjunto  $\mathcal{C}$  que pode ser escolhido pelo usuário. Para que a função  $V(\theta)$  seja semi-convexa neste conjunto precisamos garantir que (6.3) seja válida para toda a área hachurada.

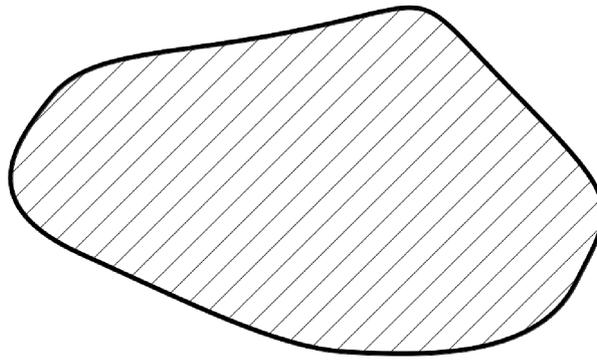


Figura 6.1: Exemplo de conjunto  $\mathcal{C}$ .

Como não é possível incluir infinitas restrições no projeto de experimentos, vamos propor que apenas um conjunto finito de restrições seja imposto ao problema de otimização. Em vez de utilizar infinitos pontos dentro do conjunto  $\mathcal{C}$ , vamos escolher aleatoriamente um grande número  $Q$  de pontos dentro do conjunto. Então vamos apenas impor restrições ao projeto de experimentos que estão relacionadas com estes pontos escolhidos aleatoriamente. A Figura 6.2 apresenta o conjunto  $\mathcal{C}$  e alguns dos pontos escolhidos aleatoriamente.

Para incluir as  $Q$  restrições no projeto de experimentos vamos adicionar ao problema de otimização as seguintes inequações

$$\phi(\theta^i) + \sum_{k=1}^M c_k \psi_k(\theta^i) > 0 \quad i = 1, \dots, Q \quad (6.4)$$

onde os vetores  $\theta^i \in \mathcal{C}$  são os  $Q$  pontos escolhidos aleatoriamente. As  $Q$  condições em (6.4) são convexas e podem ser incluídas no projeto de experimentos.

Como não foram utilizados todos os pontos do conjunto, **não** vamos garantir (6.3) para **todos** os pontos do conjunto; as restrições se aplicam apenas aos  $Q$  pontos escolhidos. Embora seja possível implementar (6.4) no projeto de experimentos, esta condição

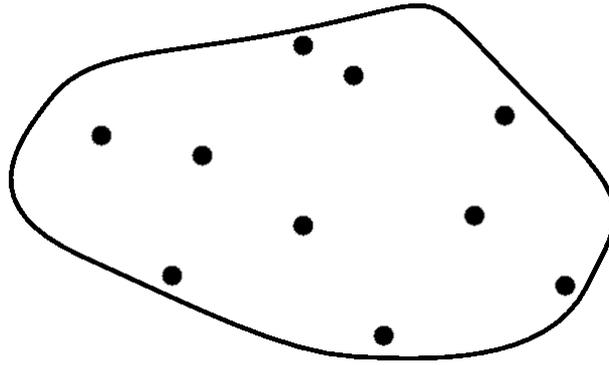


Figura 6.2: Exemplo de alguns pontos do conjunto  $\mathcal{C}$ .

não garante que (6.3) será satisfeito para todo o conjunto  $\mathcal{C}$ . Contudo, é razoável pensar que quanto maior o número de pontos  $Q$  escolhidos dentro do conjunto  $\mathcal{C}$ , melhor será a aproximação utilizada. Se fosse possível utilizar infinitos pontos então estaríamos cobrindo o conjunto todo. Precisamos então de uma métrica que avalie a qualidade da aproximação com relação ao número de pontos  $Q$ .

O trabalho (CAMPI; GARATTI; PRANDINI, 2009) apresenta uma métrica para a utilização de restrições escolhidas aleatoriamente em problemas de otimização convexos. Esta técnica foi batizada pelos autores de *Scenario Approach*. Como os pontos são escolhidos de maneira aleatória, a abordagem é classificada como estocástica e todos os resultados que utilizam esta abordagem passam a apenas ser válidos com um certo nível de confiança.

O trabalho estabelece que dentro do conjunto  $\mathcal{C}$  deve existir um outro conjunto o qual é uma fração  $\epsilon$  de  $\mathcal{C}$ . Isto significa que o tamanho deste novo conjunto é  $\epsilon$  vezes o tamanho do conjunto  $\mathcal{C}$ . A Figura 6.3 apresenta um exemplo destes conjuntos. Além disso, este novo conjunto é desconhecido, apenas sabe-se o tamanho que ele possui.

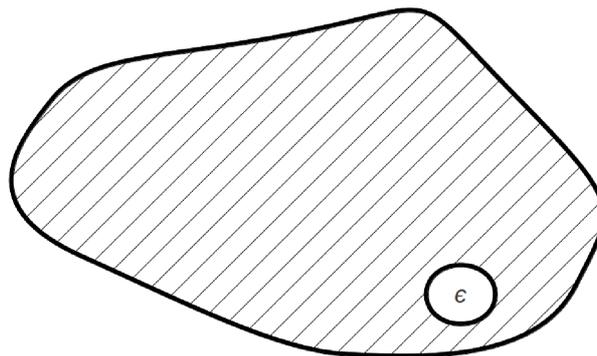


Figura 6.3: Exemplo dos conjuntos.

O trabalho (CAMPI; GARATTI; PRANDINI, 2009) estabelece a probabilidade com que a condição (6.3) seja válida para todo o conjunto  $\mathcal{C}$ , com exceção da fração  $\epsilon$ . Esta probabilidade é função tanto do número de pontos  $Q$  como da fração  $\epsilon$  que representa o conjunto dos pontos que a condição não precisa ser verificada. O principal resultado do trabalho é apresentado no seguinte lema.

**Lema 6.1** (CAMPI; GARATTI; PRANDINI, 2009) *Selecione um parâmetro de violação*

$\epsilon \in (0, 1)$  e um **parâmetro de confiança**  $\beta \in (0, 1)$ . Se (6.4) é satisfeita e

$$\sum_{i=0}^{n-1} \binom{Q}{i} \epsilon^i (1 - \epsilon)^{Q-i} < \beta$$

então, com probabilidade não menor que  $1 - \beta$ , a condição (6.3) é satisfeita para  $\forall \theta \in \mathcal{C}$  se desconsiderada uma fração  $\epsilon$  do conjunto.  $\square$

Considere que do conjunto  $\mathcal{C}$  seja retirada uma pequena fração correspondente a  $\epsilon$ . Se a condição do lema é satisfeita, então a probabilidade da condição ser verificada para todo este conjunto reduzido é maior ou igual a  $1 - \beta$ .

Este resultado é interessante pois ajuda na escolha da quantidade de pontos  $Q$ . Pode-se por exemplo escolher  $\epsilon = 0.01$  de maneira que o conjunto reduzido represente 99% do conjunto  $\mathcal{C}$ . Se for escolhido  $\beta = 0.01$ , o que representa uma probabilidade de 99%, podemos escolher a quantidade de pontos  $Q$ , de maneira que a condição do lema seja satisfeita. Desta maneira, vamos garantir, com 99% de probabilidade que a condição (6.3) seja satisfeita para ao menos 99% do conjunto  $\mathcal{C}$ .

Observando a condição do Lema 6.1 podemos notar que se tendemos o valor de  $\beta$  para zero então o número de pontos  $Q$  tende ao infinito. Por outro lado, se tendemos o valor  $\epsilon$  a zero, o número de pontos  $Q$  também tende ao infinito. O resultado expõe que não é possível garantir que a condição seja satisfeita para todo o conjunto  $\mathcal{C}$ , pois segundo a condição do lema seriam necessários infinitos pontos. Contudo, em muitas aplicações é aceitável a utilização de um intervalo de confiança razoavelmente grande ( $\beta$  pequeno) e de um conjunto razoavelmente grande ( $\epsilon$  pequeno).

No próximo exemplo vamos mostrar como utilizar esta abordagem ao projeto de experimentos e como calcular o número mínimo de pontos para garantir as condições com um nível de confiança pré-estabelecido.

### 6.2.1 Exemplo

Considere o sistema

$$G_0(q) = \frac{-0,2073q^{-1} - 0,5083q^{-2} + 0,5564q^{-3} + 0,1592q^{-4}}{1 - 3,125q^{-1} + 3,977q^{-2} - 2,382q^{-3} + 0,5677q^{-4}}$$

$$H_0(q) = 1, \quad \lambda_0 = 20.$$

o qual representa um sistema de potência que consiste de dois geradores síncronos conectados à rede, cujos dados foram retirados de (ANDERSON; FOUAD, 1977). A função de transferência  $G_0(q)$  descreve a relação entre a tensão aplicada no campo de um dos geradores com a velocidade angular do mesmo gerador, em um determinado ponto de operação. A Figura 6.4 apresenta o diagrama de Bode de  $G_0(q)$ .

Vamos identificar um modelo para este sistema utilizando uma estrutura Output-Error:

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + F^T(q)\theta}, \quad H(q, \theta) = 1$$

onde

$$B(q) = [q^{-1} \ q^{-2} \ q^{-3} \ q^{-4} \ 0 \ 0 \ 0 \ 0]^T,$$

$$F(z) = [0 \ 0 \ 0 \ 0 \ q^{-1} \ q^{-2} \ q^{-3} \ q^{-4}]^T.$$

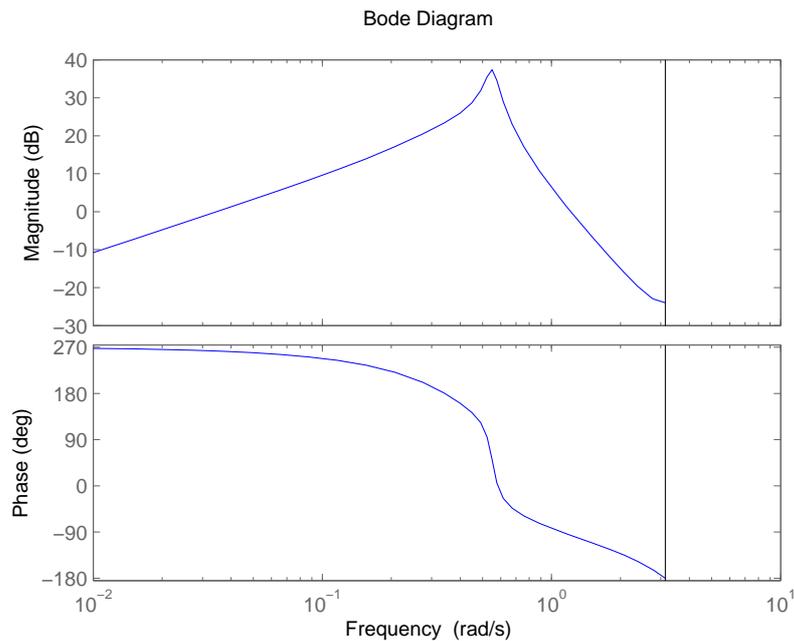


Figura 6.4: Diagrama de Bode de  $G_0(z)$ .

O sistema pertence ao conjunto de modelos definido pela estrutura, ou seja,  $G_0(z) = G(q, \theta^0)$  onde  $\theta^0$  é dado por

$$\theta^0 = [-0,2073 \quad -0,5083 \quad 0,5564 \quad 0,1592 \quad -3,125 \quad 3,977 \quad -2,382 \quad 0,5677]^T.$$

Antes de rodar o experimento é necessário escolher o sinal de entrada que será aplicado ao sistema. O sinal será escolhido utilizando o projeto de experimentos apresentado no Capítulo 5. O número de amostras utilizado será  $N = 1.000$  e o espectro será discreto com a seguinte parametrização:

$$\Phi(\omega) = \sum_{k=1}^{40} c_k (\delta(\omega - \omega_k) + \delta(\omega + \omega_k))$$

onde

$$\omega_k = 10^{(2\frac{(k-1)}{40}-2)}\pi.$$

As frequências foram escolhidas de maneira a cobrir uma ampla faixa do espectro (observe a Figura 6.4).

Vamos utilizar diferentes restrições no projeto de experimento. O primeiro espectro será denotado  $\Phi_1(\omega)$  e será obtido pela minimização do maior autovalor da matriz de covariância, tal que a potência média do sinal de entrada seja menor ou igual a 1. O

problema de otimização é o seguinte:

$$\begin{aligned} & \min_{c_1, \dots, c_{40}, \gamma} \gamma \\ & \text{sujeito a} \\ & c_k \geq 0 \quad i = k, \dots, 40 \\ & \frac{1}{\pi} \sum_{k=1}^{40} c_k \leq 1 \\ & \begin{bmatrix} \gamma I & I \\ I & \sum_{k=1}^{40} c_k \mathbf{B}_k^P \end{bmatrix} > 0 \end{aligned}$$

onde  $\mathbf{B}_k^P$  são as matrizes definidas anteriormente.

Este problema de otimização foi resolvido utilizando o solver `lmi lab` do Matlab. A solução do problema gerou o espectro apresentado na Figura 6.5 em forma de quadrados. O menor autovalor de  $P_\theta^{-1}$  é 0.0044.

A matriz  $P_\theta^{-1}$  calculada com o primeiro espectro foi computada como

$$\begin{bmatrix} 55, 2193 & 48, 6381 & 30, 7532 & 6, 5475 & -575, 9925 & -1334, 0086 & -1670, 023 & -1512, 8205 \\ 48, 6381 & 55, 2193 & 48, 6381 & 30, 7532 & 404, 0486 & -575, 9925 & -1334, 0086 & -1670, 023 \\ 30, 7532 & 48, 6381 & 55, 2193 & 48, 6381 & 1329, 957 & 404, 0486 & -575, 9925 & -1334, 0086 \\ 6, 5475 & 30, 7532 & 48, 6381 & 55, 2193 & 1927, 206 & 1329, 957 & 404, 0486 & -575, 9925 \\ -575, 9925 & 404, 0486 & 1329, 957 & 1927, 206 & 146190, 1915 & 125456, 9388 & 69214, 0205 & -6396, 2663 \\ -1334, 0086 & -575, 9925 & 404, 0486 & 1329, 957 & 125456, 9388 & 146190, 1915 & 125456, 9388 & 69214, 0205 \\ -1670, 023 & -1334, 0086 & -575, 9925 & 404, 0486 & 69214, 0205 & 125456, 9388 & 146190, 1915 & 125456, 9388 \\ -1512, 8205 & -1670, 023 & -1334, 0086 & -575, 9925 & -6396, 2663 & 69214, 0205 & 125456, 9388 & 146190, 1915 \end{bmatrix}$$

O segundo espectro será denotado  $\Phi_2(\omega)$  e também será obtido pela minimização do maior autovalor da matriz de covariância, tal que a potência média do sinal de entrada seja menor ou igual a 1. Além disso, será incluída a restrição (6.3) para melhorar a convergência do método de identificação.

O conjunto  $\mathcal{C}$  é definido como

$$\mathcal{C} = \{ \theta : (\theta - \theta^0)^T 10^3 P_\theta^{-1} (\theta - \theta^0) < \chi_8^2(95\%) \}$$

de tal maneira que ele representa o intervalo de confiança de 95% da estimativa que seria obtida com 1000 amostras de uma realização de  $\Phi_1(\omega)$ .

O número de pontos utilizados dentro do conjunto  $\mathcal{C}$  para compor as restrições de convergência é  $Q = 1000$ . Esta quantidade de pontos garante com 99.92% de probabilidade que a restrição (6.4) é satisfeita para pelo menos 98% do conjunto  $\mathcal{C}$ .

O problema de otimização é o seguinte:

$$\begin{aligned} & \min_{c_1, \dots, c_{40}, \gamma} \gamma \\ & \text{sujeito a} \\ & c_k \geq 0 \quad i = k, \dots, 40 \\ & \frac{1}{\pi} \sum_{k=1}^{40} c_k \leq 1 \\ & \begin{bmatrix} \gamma I & I \\ I & \sum_{k=1}^{40} c_k \mathbf{B}_k^P \end{bmatrix} > 0 \\ & \phi(\theta^i) + \sum_{k=1}^{40} c_k \psi_k(\theta^i) > 0 \quad i = 1, \dots, 1000 \end{aligned}$$

onde  $B_k^P$ ,  $\phi(\cdot)$  e  $\psi_k(\cdot)$  foram definidos anteriormente e  $\theta^i$  é escolhido aleatoriamente do conjunto  $\mathcal{C}$ .

O espectro relacionado com a solução deste problema é plotado na Figura 6.5, o qual é representado pelos círculos. A matriz  $P_\theta^{-1}$  calculada com o segundo espectro foi computada como

$$\begin{bmatrix} 56,3632 & 51,5002 & 37,9659 & 18,6354 & 676,8681 & 310,9146 & -70,9465 & -398,0481 \\ 51,5002 & 56,3632 & 51,5002 & 37,9659 & 949,5046 & 676,8681 & 310,9146 & -70,9465 \\ 37,9659 & 51,5002 & 56,3632 & 51,5002 & 1060,9036 & 949,5046 & 676,8681 & 310,9146 \\ 18,6354 & 37,9659 & 51,5002 & 56,3632 & 969,5702 & 1060,9036 & 949,5046 & 676,8681 \\ 676,8681 & 949,5046 & 1060,9036 & 969,5702 & 47157,5433 & 41540,7255 & 26112,9214 & 4763,5478 \\ 310,9146 & 676,8681 & 949,5046 & 1060,9036 & 41540,7255 & 47157,5433 & 41540,7255 & 26112,9214 \\ -70,9465 & 310,9146 & 676,8681 & 949,5046 & 26112,9214 & 41540,7255 & 47157,5433 & 41540,7255 \\ -398,0481 & -70,9465 & 310,9146 & 676,8681 & 4763,5478 & 26112,9214 & 41540,7255 & 47157,5433 \end{bmatrix}$$

A solução do problema está relacionada com uma matriz de covariância cujo menor autovalor de  $P_\theta^{-1}$  é 0.0037. O menor autovalor é menor que aquele obtido com o primeiro espectro. A solução deste problema de otimização resultou em um espectro que está relacionado com uma maior matriz de covariância. A inclusão da restrição de convergência forçou o problema de otimização a encontrar outra solução, cuja matriz de covariância possui maiores autovalores.

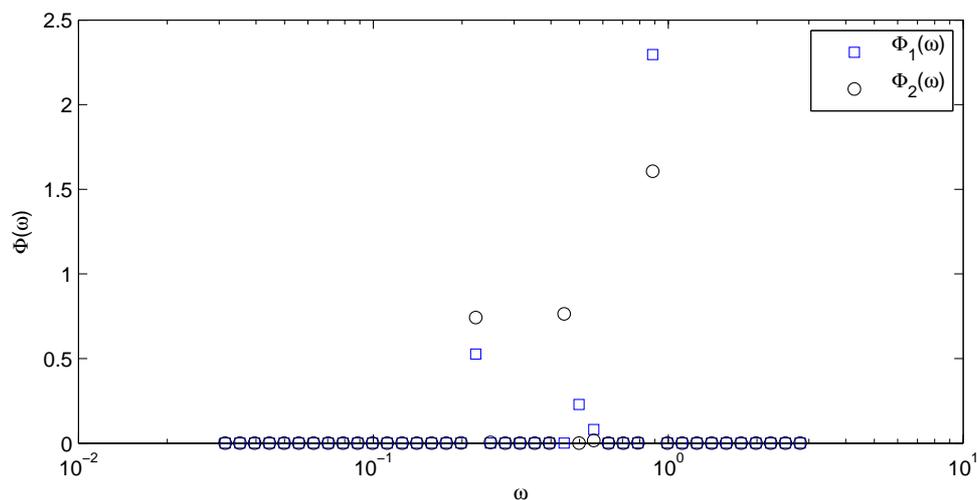


Figura 6.5: Espectros.

O vetor de parâmetros  $\theta$  foi estimado utilizando o toolbox de identificação de sistemas UNIT (University of Newcastle Identification Toolbox). Para cada um dos dois espectros foram realizadas 100 rodadas de Monte-Carlo, gerando portando 100 estimativas diferentes para o vetor de parâmetros. Em cada rodada foi utilizada uma realização diferente para o ruído adicionado ao sinal de saída.

As Figuras 6.6 e 6.7 apresentam as estimativas obtidas para o primeiro espectro. Na Figura 6.6 o eixo das ordenadas apresenta o valor de cada um dos oito elementos do vetor de parâmetros  $\theta$ , enquanto que a abcissa corresponde a cada uma das rodadas de Monte-Carlo. As rodadas de Monte-Carlo foram reordenadas em termos do menor valor de  $\theta$ .

Pode-se ver que após o 19º ponto do gráfico os valores estão bastante próximos dos parâmetros corretos  $\theta^0$ , indicando que 82 rodadas convergiram para o mínimo global da função custo. Contudo, para 18 das rodadas os valores estão agrupados em torno de um outro ponto no espaço, distante de  $\theta^0$ , indicando que o algoritmo não convergiu para o mínimo global.

A Figura 6.7 apresenta o *boxplot* dos parâmetros onde o eixo das ordenadas indica o elemento do vetor de parâmetros e o eixo das abcissas apresenta o valor obtido em cada rodada. Pode-se notar que o algoritmo não convergiu para o mínimo global em todas as rodadas.

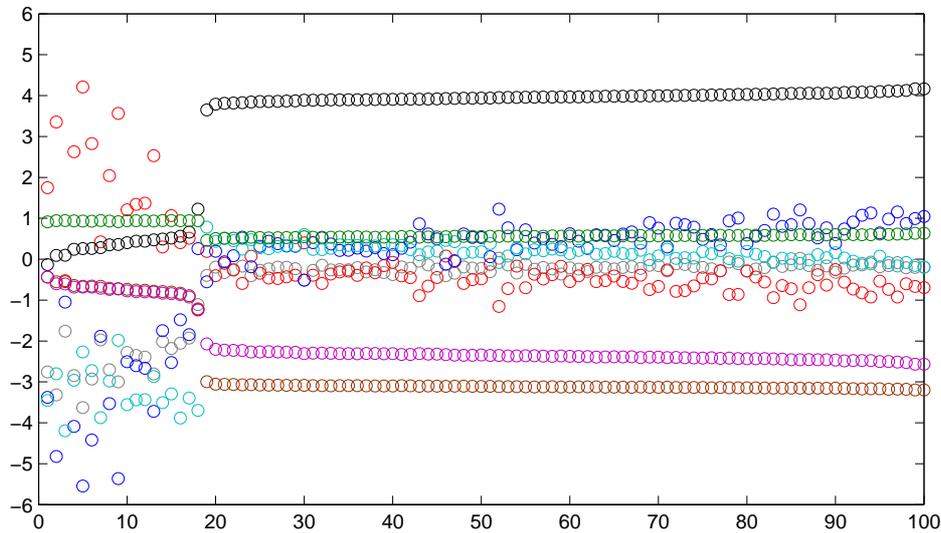


Figura 6.6: Rodadas de Monte-Carlo para  $\Phi_1(\omega)$ .

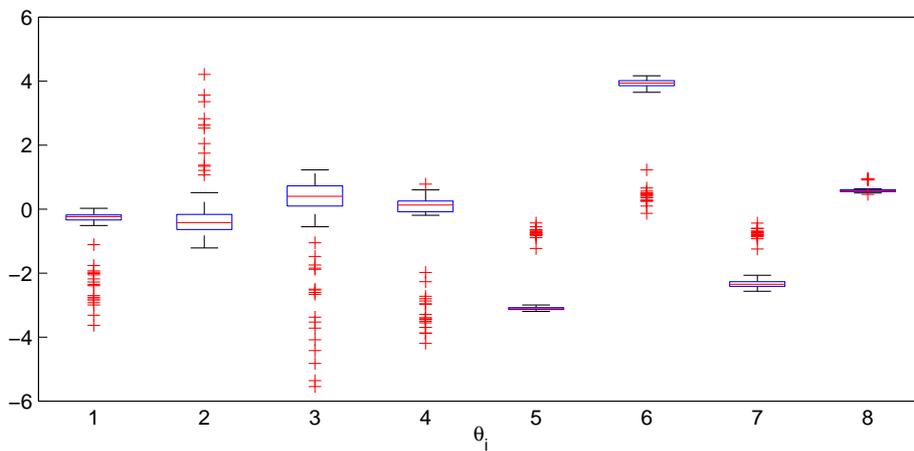


Figura 6.7: Boxplot das rodadas de Monte-Carlo para  $\Phi_1(\omega)$ .

As Figuras 6.8 e 6.9 apresentam os mesmos gráficos para o segundo espectro. Pode-se ver que o mínimo global foi obtido em todas as rodadas de Monte-Carlo.

Vamos explorar um pouco mais os resultados obtidos calculando a média dos parâmetros para as rodadas de Monte-Carlo. Vamos definir

$$\theta_m = \frac{1}{100} \sum_{i=1}^{100} \hat{\theta}_N^i$$

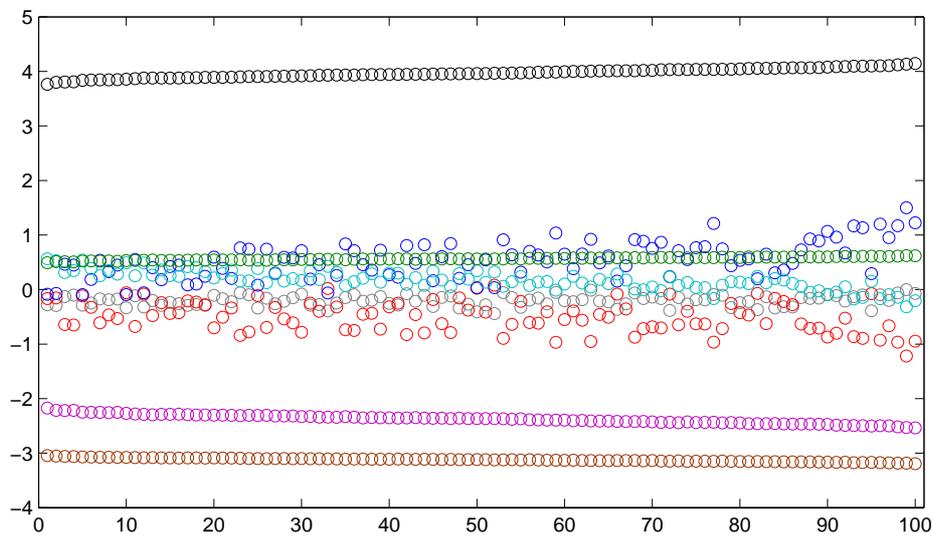


Figura 6.8: Rodadas de Monte-Carlo para  $\Phi_2(\omega)$ .

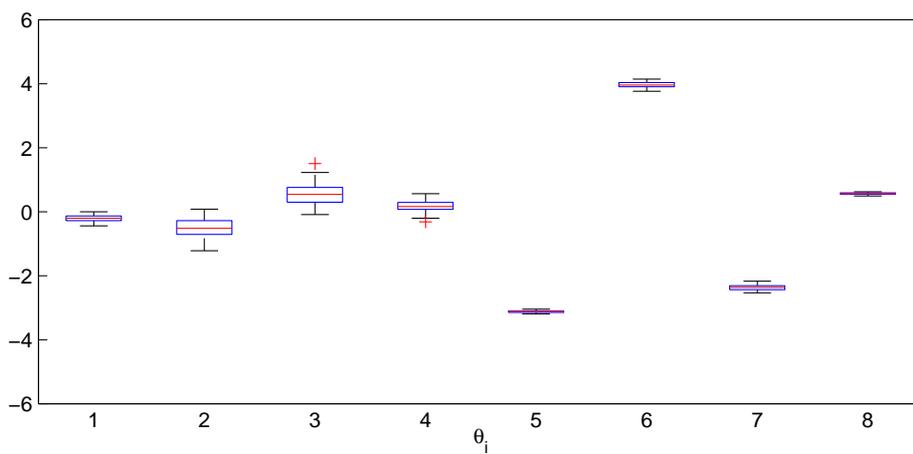


Figura 6.9: Boxplot das rodadas de Monte-Carlo para  $\Phi_2(\omega)$ .

onde cada  $\hat{\theta}_N^i$  representa a estimativa obtida na  $i$ -ésima rodada de Monte-Carlo.

$$\theta_m(\Phi_1) = [-0,61811 \quad -0,1173 \quad -0,1005 \quad -0,43944 \\ -2,691 \quad 3,323 \quad -2,083 \quad 0,63166]^T$$

$$\theta_m(\Phi_2) = [-0,20804 \quad -0,51126 \quad 0,56363 \quad 0,1546 \\ -3,1264 \quad 3,9807 \quad -2,3859 \quad 0,56916]^T$$

Comparando estes valores com os parâmetros corretos de  $\theta^0$  podemos ver que em média o método proposto se aproxima mais dos parâmetros corretos que o método usual.

Vamos também calcular a matriz de covariância amostral obtida com as 100 rodadas de Monte-Carlo de cada método. Utilizando o segundo espectro obtivemos uma estimativa para a matriz de covariância cujo menor autovalor de  $P_\theta^{-1}$  foi 0,00412469, o qual é bastante próximo do valor teórico 0,0037.

Por outro lado, quando o espectro  $\Phi_1(\omega)$  foi utilizado, o menor autovalor de  $P_\theta^{-1}$  foi calculado como 0,00015350, o qual é muito menor que o valor teórico 0,0044. Como esperado, as 18 rodadas de Monte-Carlo que não convergiram para o mínimo global tornam a matriz de covariância muito maior do que era esperado (em torno de 30 vezes).

### 6.3 S-procedure

Na seção anterior apresentamos uma condição necessária para garantir (6.1). A condição apresentada é necessária mas não é suficiente, o que revela que a condição apresentada não garante que (6.3) é válida para todo o conjunto  $\mathcal{C}$ . A condição proposta é estocástica e o resultado do Lema 6.1 descreve o intervalo de confiança da solução que será obtida. A condição é convexa e pode ser utilizada no projeto de experimentos.

Nesta seção vamos apresentar condições suficientes para garantir (6.3). Estas condições vão garantir que (6.3) seja válida para todo o conjunto  $\mathcal{C}$ , mas em contrapartida serão mais restritivas que as anteriormente apresentadas.

A ideia principal aqui é abrir o somatório (6.3) em seus  $M + 1$  termos, e garantir que cada um dos termos não seja negativo para todo  $\theta$  que pertence ao conjunto  $\mathcal{C}$ . Se todos os termos da soma não forem negativos, então a condição (6.3) será válida para todo o conjunto  $\mathcal{C}$ .

O primeiro dos termos é dado por

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 |H_0(e^{j\omega})(H^{-1}(e^{j\omega}, \theta) - H_0^{-1}(e^{j\omega}))|^2 \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} \right\} d\omega. \quad (6.5)$$

Esta equação não depende do espectro do sinal de entrada e portanto o projeto de experimentos não pode ser utilizado para garantir a positividade deste termo. Para utilizar a abordagem proposta nesta seção é necessário garantir por algum meio que a equação acima seja positiva para todo  $\theta$  que pertence ao conjunto  $\mathcal{C}$ . Para algumas estruturas de modelo esta equação é sempre positiva, por exemplo quando o modelo é do tipo Output-Error, o que significa que  $C(e^{j\omega}) = 0$  na expressão. Caso seja utilizada outra estrutura para o modelo, é necessário garantir que (6.5) seja positiva utilizando alguma outra ferramenta.

Cada um dos outros termos é definido como

$$c_k \frac{1}{\pi} \int_{-\pi}^{\pi} \mathbf{B}_k(e^{j\omega}) |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \cdot \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} + \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta} - \frac{1 + D^T(e^{j\omega})\theta^0}{1 + D^T(e^{j\omega})\theta} \right\} d\omega. \quad (6.6)$$

Como estes termos dependem dos parâmetros  $c_k$ , os quais definem o espectro do sinal de entrada  $\Phi_u(\omega)$ , podemos utilizar o projeto de experimentos para garantir que cada um dos termos não seja negativo.

No integrando da expressão acima, apenas o seguinte termo pode ser negativo:

$$R(\omega) \triangleq \Re \left\{ \frac{1 + C^T(e^{j\omega})\theta^0}{1 + C^T(e^{j\omega})\theta} + \frac{1 + F^T(e^{j\omega})\theta^0}{1 + F^T(e^{j\omega})\theta} - \frac{1 + D^T(e^{j\omega})\theta^0}{1 + D^T(e^{j\omega})\theta} \right\}.$$

Quando este termo é positivo, então o integrando é positivo e portanto a integral também é positiva. Portanto a positividade do termo (6.6) depende da positividade de  $R(\omega)$ .

Vamos agora considerar apenas o caso em que o espectro é discreto e parametrizado com  $\mathbf{B}_k(e^{j\omega}) = \delta(\omega - \omega_k) + \delta(\omega + \omega_k)$ . Neste caso a expressão (6.6) pode ser escrita como

$$\frac{2}{\pi} |H^{-1}(e^{j\omega_k}, \theta)(G_0(e^{j\omega_k}) - G(e^{j\omega_k}, \theta))|^2 \cdot c_k \Re \left\{ \frac{1 + C^T(e^{j\omega_k})\theta^0}{1 + C^T(e^{j\omega_k})\theta} + \frac{1 + F^T(e^{j\omega_k})\theta^0}{1 + F^T(e^{j\omega_k})\theta} - \frac{1 + D^T(e^{j\omega_k})\theta^0}{1 + D^T(e^{j\omega_k})\theta} \right\} \quad (6.7)$$

pois devido à parametrização do espectro, podemos resolver a integral (6.6) analiticamente. Note que (6.7) não será negativo se

$$c_k R(\omega_k) \geq 0 \quad k = 1, \dots, M. \quad (6.8)$$

Podemos garantir que (6.1) seja verificada fazendo uma escolha apropriada dos valores  $c_k$ . Para tanto, basta utilizar valores positivos para  $c_k$  quando  $R(\omega_k) \geq 0$  para todo  $\theta$  pertencente ao conjunto  $\mathcal{C}$  e fazer  $c_k = 0$  quando  $R(\omega_k) < 0$  para algum  $\theta$  pertencente a  $\mathcal{C}$ .

É importante ressaltar que o procedimento apresentado é suficiente para garantir que (6.1) seja verificado, mas não é necessário. Estamos garantindo que todos os termos do somatório (6.3) sejam positivos e concluindo que a soma será positiva. Entretanto, mesmo se alguns dos termos do somatório fossem negativos, a soma poderia ainda assim ser positiva, portanto o procedimento é conservador.

Precisamos agora de um método para verificar se  $R(\omega_k)$  é não negativo para todo  $\theta$  que pertence ao conjunto  $\mathcal{C}$ . Observe que  $R(\omega) \geq 0$  é equivalente a

$$\begin{aligned} \Xi(\theta) \triangleq & |1 + F^T\theta|^2 |1 + D^T\theta|^2 \Re \{ (1 + C^T\theta)(1 + C^*\theta^0) \} \\ & + |1 + C^T\theta|^2 |1 + D^T\theta|^2 \Re \{ (1 + F^T\theta)(1 + F^*\theta^0) \} \\ & - |1 + F^T\theta|^2 |1 + C^T\theta|^2 \Re \{ (1 + D^T\theta)(1 + D^*\theta^0) \} \geq 0. \end{aligned} \quad (6.9)$$

Para obter este resultado basta somar o termo com seu complexo conjugado e então multiplicar o resultado pelo mínimo múltiplo comum do denominador. A vantagem de utilizar

a condição (6.9) está no fato que tal condição é polinomial em relação a  $\theta$ , enquanto que  $R(\omega)$  é racional.

Vamos agora analisar se  $\Xi(\theta) \geq 0$  para todo  $\theta$  pertencente ao conjunto  $\mathcal{C}$ . Para tanto vamos especificar que o conjunto  $\mathcal{C}$  é descrito por

$$\mathcal{C} = \{\theta : \Pi(\theta) \geq 0\}. \quad (6.10)$$

Gostaríamos de garantir que  $\Xi(\theta) \geq 0$  para todo  $\theta$  pertencente a  $\mathcal{C}$  e portanto vamos utilizar *S-procedure* para obter a seguinte condição

$$\Xi(\theta) \geq \kappa \Pi(\theta) \quad \forall \theta \in \mathbb{R}^n. \quad (6.11)$$

Note que se existe  $\kappa > 0, \kappa \in \mathbb{R}$  que respeita a condição acima então  $\Xi(\theta) \geq 0 \forall \theta \in \mathcal{C}$  pois  $\Pi(\theta) \geq 0 \forall \theta \in \mathcal{C}$ . Logo, podemos utilizar a condição (6.11) para garantir que  $R(\omega)$  seja positiva para todo  $\theta \in \mathcal{C}$ . Esta técnica é chamada de *S-procedure* (BOYD et al., 1994).

É importante ressaltar que a abordagem por *S-procedure* torna o problema ainda mais conservador. A condição (6.11) é necessária e suficiente apenas em alguns casos específicos; em muitos casos esta condição é apenas suficiente. Um caso conhecido e muito utilizado em que o *S-procedure* é necessário e suficiente é quando tanto  $\Xi(\theta)$  como  $\Pi(\theta)$  são funções quadráticas em  $\theta$ .

O conservadorismo da *S-procedure* pode ser reduzido se for trocado o escalar  $\kappa$  por um polinômio  $p(\theta)$  tal que  $p(\theta) > 0 \forall \theta$ . Neste caso obtemos a seguinte condição

$$\Xi(\theta) \geq p(\theta)\Pi(\theta) \quad \forall \theta \in \mathbb{R}^n. \quad (6.12)$$

Para garantir que  $p(\theta) > 0$  e que (6.12) seja verificada é possível utilizar a abordagem por soma de quadrados (*sum of squares*) (PUTINAR, 1993). Embora a abordagem por soma de quadrados seja convexa, seu custo computacional é bastante elevado, o que restringe sua aplicação para apenas problemas de pequena ordem.

### 6.3.1 Estrutura Output-Error generalizada

Quando a estrutura do modelo é do tipo Output-Error generalizada, tanto  $C(e^{j\omega}) = 0$  como  $D(e^{j\omega}) = 0$ , o que simplifica bastante as condições obtidas anteriormente.

A primeira simplificação fica evidente em (6.5), pois a expressão é sempre positiva para esta estrutura. Porém, esta não é a única simplificação oriunda desta parametrização. Note que a expressão para  $\Xi(\theta)$  fica mais simples

$$\Xi(\theta) = (1 + F^T(e^{j\omega_i})\theta^0)(1 + F^*(e^{j\omega_i})\theta) + (1 + F^*(e^{j\omega_i})\theta^0)(1 + F^T(e^{j\omega_i})\theta)$$

o qual é afim na variável  $\theta$ .

Neste caso a condição  $\Xi(\theta) \geq 0$  é equivalente a

$$\begin{bmatrix} \theta \\ 1 \end{bmatrix}^T \begin{bmatrix} 0 & \Re\{F^T + F^T F^* \theta^0\}^T \\ \Re\{F^T + F^T F^* \theta^0\} & 2\Re\{1 + F^T \theta^0\} \end{bmatrix} \begin{bmatrix} \theta \\ 1 \end{bmatrix} \geq 0.$$

onde foi omitida a dependência na variável  $\omega$ .

Se o conjunto  $\mathcal{C}$  é elipsoidal, então podemos simplificar ainda mais as condições. Vamos definir o seguinte conjunto

$$\mathcal{C} = \{\theta | (\theta - \theta^0)^T P (\theta - \theta^0) \leq 1\}.$$

Este conjunto pode ser descrito na forma  $\mathcal{C} = \{\theta : F(\theta) \geq 0\}$  onde

$$\Pi(\theta) = \begin{bmatrix} \theta \\ 1 \end{bmatrix}^T \begin{bmatrix} -P & P\theta^0 \\ \theta^{0T}P & 1 - \theta^{0T}P\theta^0 \end{bmatrix} \begin{bmatrix} \theta \\ 1 \end{bmatrix} \geq 0.$$

Utilizando as expressões acima a condição do *S-procedure* (6.11) se torna

$$\begin{bmatrix} 0 & \Re\{F^T + F^T F^* \theta^0\}^T \\ \Re\{F^T + F^T F^* \theta^0\} & 2\Re\{1 + F^T \theta^0\} \end{bmatrix} + \kappa \begin{bmatrix} P & -P\theta^0 \\ -\theta^{0T}P & \theta^{0T}P\theta^0 - 1 \end{bmatrix} \geq 0. \quad (6.13)$$

Como tanto  $\Xi(\theta)$  como  $\Pi(\theta)$  são quadráticos na variável  $\theta$  então o *S-procedure* não impõe conservadorismo extra na solução do problema.

### 6.3.2 Utilização da abordagem

Para utilizar a abordagem proposta nesta seção é necessário realizar os seguintes passos:

1. Escolher um conjunto  $\mathcal{C}$  que possa ser descrito pela equação (6.10).
2. Escolher o número  $M$  de componentes do espectro e suas frequências, de tal maneira que o espectro possa ser descrito por

$$\Phi_u(\omega) = \sum_{k=1}^M c_k (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)).$$

3. Verificar se (6.5) é positivo.
4. Para cada  $k = 1, \dots, M$  verificar se existe  $\kappa$  tal que a condição (6.11) seja satisfeita. Se a condição não é satisfeita deve-se forçar  $c_k = 0$ .
5. Utilizar o projeto de experimentos para calcular os valores  $c_k$  para todas as componentes que não foram previamente forçadas igual a zero.

No final deste procedimento será obtido um espectro tal que a condição (6.3) é satisfeita para todo o conjunto  $\mathcal{C}$ . O exemplo a seguir utiliza este procedimento e mostra que ele pode ser utilizado para melhorar a convergência dos métodos de identificação.

### 6.3.3 Exemplo

Considere o seguinte processo

$$y(t) = \frac{0,0030q^{-3}}{1 - 2,5500q^{-1} + 2,1650q^{-2} - 0,6120q^{-3}}u(t) + e(t)$$

onde  $e(t)$  é ruído branco com variância  $\lambda_0 = 0.01$ . Este sistema pode ser representado por

$$y(t) = \frac{0,0030q^{-3}}{(1 - 0,9q^{-1})(1 - 0,85q^{-1})(1 - 0,8q^{-1})}u(t) + e(t).$$

A resposta ao degrau deste sistema é apresentada na Figura 6.10. O Diagrama de Bode do sistema pode ser visto na Figura 6.11.

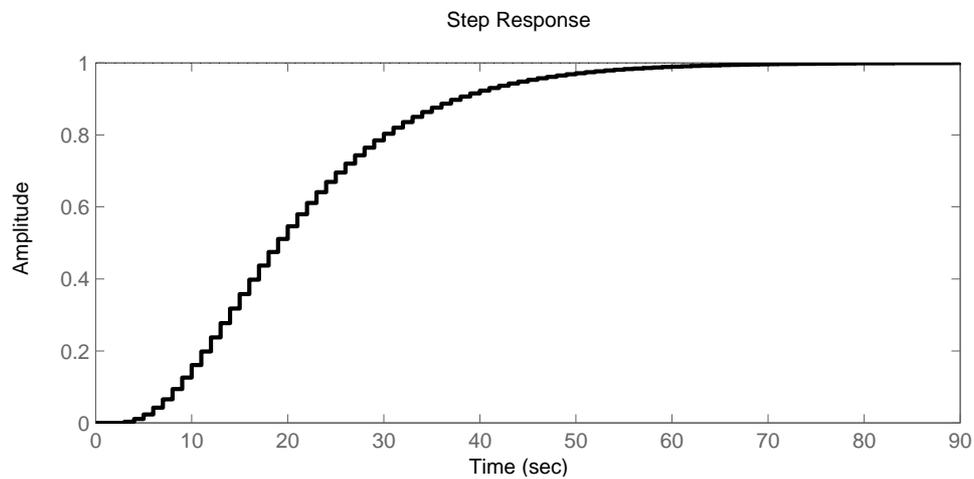


Figura 6.10: Resposta ao degrau do sistema.

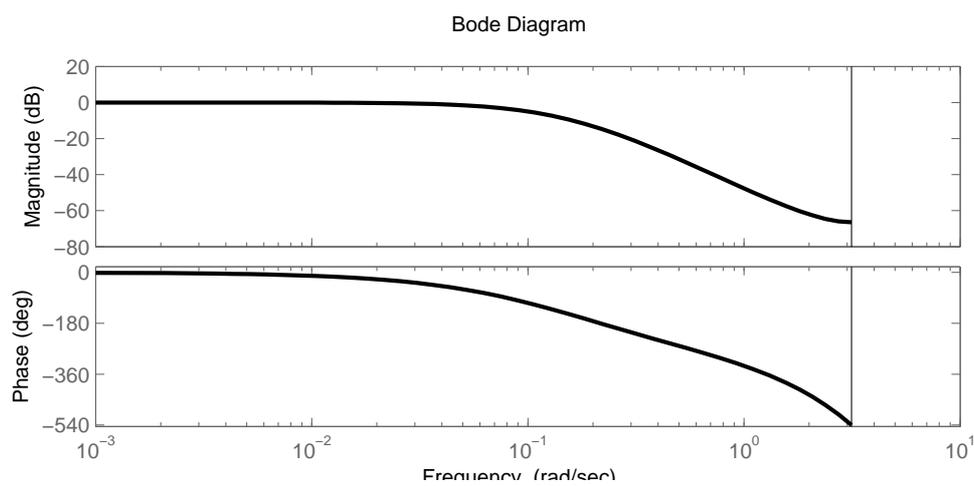


Figura 6.11: Diagrama de Bode do sistema.

Vamos considerar que o sistema é desconhecido e que gostaríamos de identificar os parâmetros do seguinte modelo a partir de dados coletados do sistema.

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + F^T(q)\theta},$$

$$B^T(q) = [ q^{-3} \quad 0 \quad 0 \quad 0 ],$$

$$F^T(q) = [ 0 \quad q^{-1} \quad q^{-2} \quad q^{-3} ].$$

Note que a estrutura do modelo pode representar o sistema real quando

$$\theta = \theta^0 = [ 0,0030 \quad -2,5500 \quad 2,1650 \quad -0,6120 ]^T.$$

Vamos projetar o sinal de entrada que será utilizado para identificar o processo. Isto será feito por

$$\begin{array}{ll} \text{minimizar} & \text{energia do sinal de entrada} \\ & \Phi_u(\omega) \\ \text{sujeito a} & \text{restrições de qualidade e de convergência} \end{array}$$

Vamos considerar que o espectro do sinal de entrada é parametrizado da seguinte maneira

$$\Phi_u(\omega) = \sum_{k=1}^{100} c_k (\delta(\omega - \omega_k) + \delta(\omega + \omega_k))$$

onde  $\omega_k = \pi 10^{\frac{3(k-1)}{100}-3}$ ,  $k = 1, \dots, 100$ . Este sinal cobre o espectro de  $0,001\pi \text{ rad/s}$  até  $0,93\pi \text{ rad/s}$ . A restrição de qualidade será a variância da estimativa do modelo. Gostaríamos que a estimativa tivesse precisão melhor que 0,001 em um intervalo de confiança de 95%.

Dois projetos de experimento serão realizados. No primeiro as restrições de convergência não serão aplicadas. No segundo vamos utilizar as restrições de convergência, conforme foi descrito anteriormente.

### 6.3.4 Projeto de experimento sem restrições de convergência

Projetamos o espectro de entrada para minimizar a energia do sinal de entrada e garantir que a precisão das estimativas seja melhor que 0,001 em um intervalo de confiança de 95%. O espectro do sinal foi obtido utilizando o projeto de experimentos. Uma realização deste espectro é

$$u_1(t) = 71,01\cos(0,0433t) + 137\cos(0,1982t).$$

a qual pode ser vista na Figura 6.12.

O sinal de entrada  $u_1(t)$  foi aplicado ao sistema e a saída foi coletada. Os sinais de entrada e saída foram utilizados para estimar os parâmetros do modelo. A condição inicial utilizada foi

$$\theta^1 = [ 0,0196 \quad -2,5461 \quad 2,1701 \quad -0,6044 ]^T.$$

O algoritmo do gradiente rodou 10000 iterações e obteve a seguinte estimativa para os parâmetros do modelo

$$\bar{\theta}^1 = [ -0,0004 \quad -2,5447 \quad 2,1703 \quad -0,6033 ]^T.$$

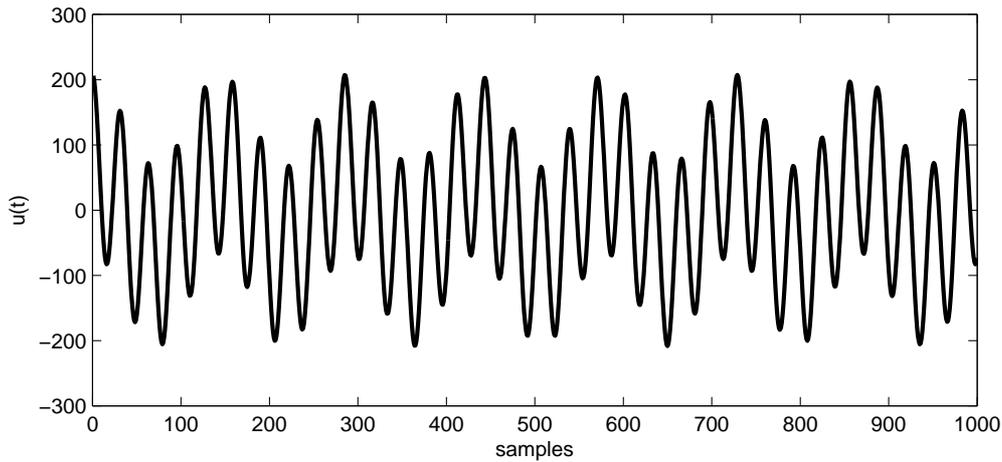


Figura 6.12: Sinal de entrada projetado sem as restrições de convergência.

Estes parâmetros representam o modelo

$$G(q, \bar{\theta}^1) = \frac{-0,00042015q^{-3}}{(1 - 0,6063q^{-1})(1 - 1,939q^{-1} + 0,9952q^{-2})}.$$

Este modelo é bastante diferente do sistema real. A estimativa do custo foi

$$V(\bar{\theta}^1) = \sum_{t=1}^{1000} (\varepsilon(t, \bar{\theta}^1))^2 = 2289.$$

### 6.3.5 Projeto de experimento com restrições de convergência

Para utilizar o procedimento proposto é necessário escolher o conjunto  $\mathcal{C}$ . Foi escolhido o seguinte conjunto

$$\mathcal{C} = \{\theta | (\theta - \theta^0)^T P (\theta - \theta^0) \leq 1\}$$

onde

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 847885 & -198919 & -296429 \\ 0 & -198919 & 739874 & -387637 \\ 0 & -296429 & -387637 & 422343 \end{bmatrix}.$$

Este conjunto foi escolhido de maneira a incluir a condição inicial  $\theta^1$ . O conjunto representa a região em que gostaríamos a função  $V(\theta)$  fosse semi-convexa. Note que somente os parâmetros relacionados aos polos do sistemas estão restringidos neste conjunto. Isto foi feito porque somente os polos afetam a convergência do algoritmo. Este conjunto está representado na Figura 6.13.

O procedimento proposto foi utilizado, o que resultou em um espectro calculado utilizando o projeto de experimentos. Uma realização do sinal ótimo é

$$u_2(t) = 47,01\cos(0,0703t) + 163\cos(0,2275t).$$

o qual pode ser visto na Figura 6.14. O sinal de entrada  $u_2(t)$  foi aplicado ao processo e dados de entrada e saída foram coletados. Os sinais foram utilizados para estimar os parâmetros do modelo. A condição inicial foi a mesma utilizada no caso anterior. O

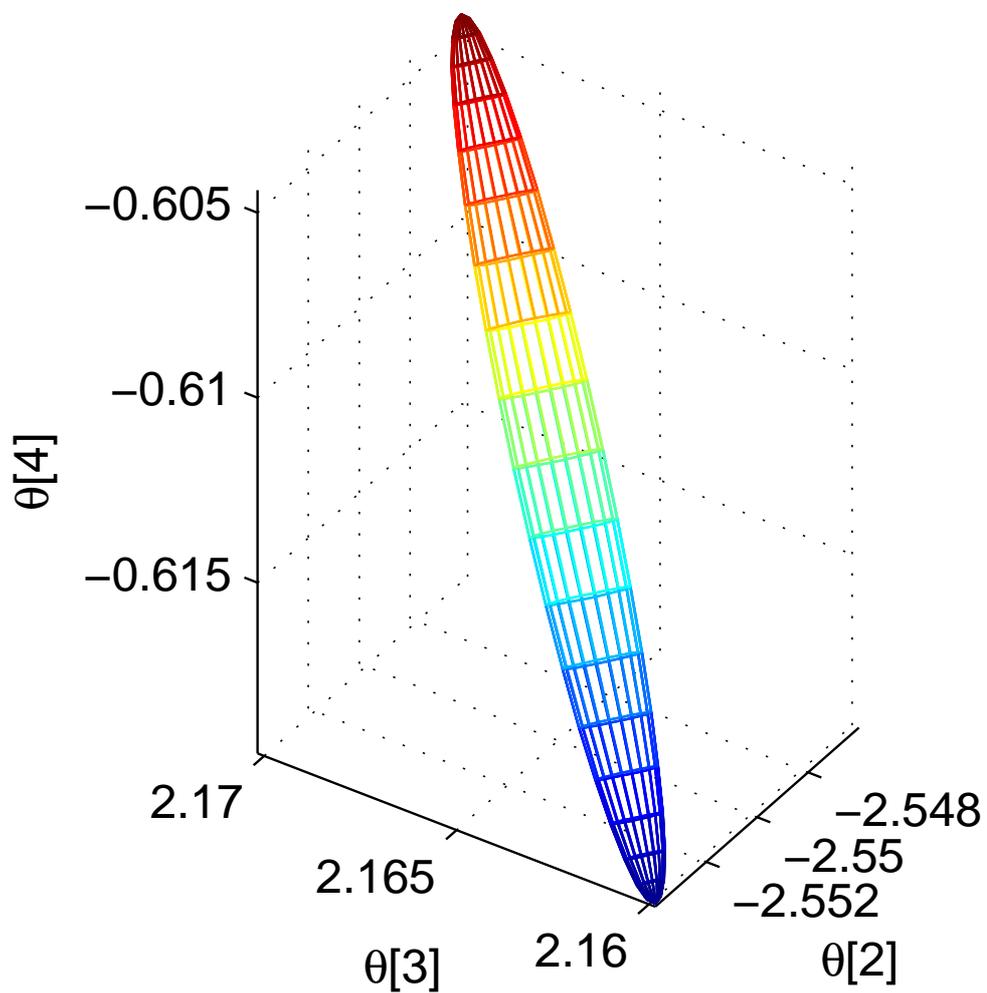


Figura 6.13: Elipsoide de restrições  $\mathcal{C}$  projetado no subespaço  $\theta = [0 \ x \ x]^T$ .

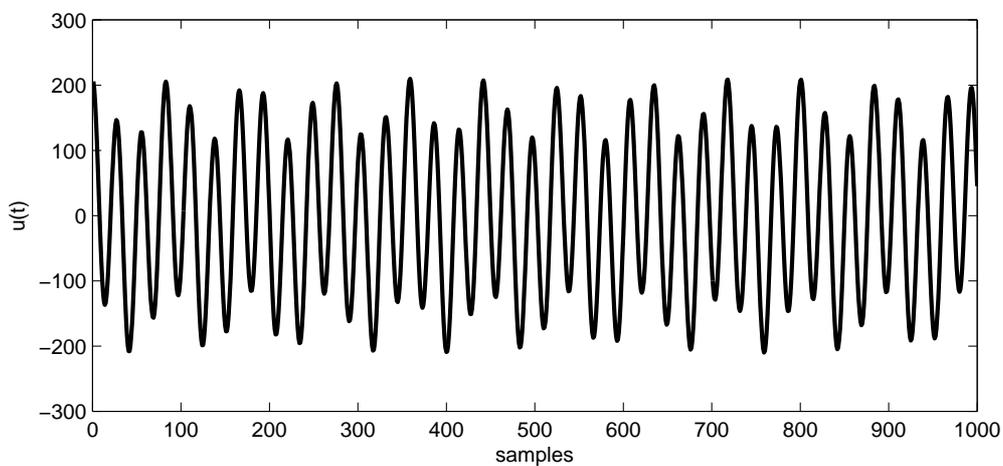


Figura 6.14: Sinal de entrada calculado com as restrições de convergência.

algoritmo rodou 10000 iterações e ele obteve a seguinte estimativa para os parâmetros do modelo

$$\bar{\theta}^2 = [ 0,0029 \quad -2,5496 \quad 2,1644 \quad -0,6116 ] .$$

Estes parâmetros representam o modelo

$$G(q, \bar{\theta}^2) = \frac{0,0029q^{-3}}{1 - 2,5496q^{-1} + 2,1644q^{-2} - 0,6116q^{-3}} .$$

A restrição de qualidade foi respeitada e todos os parâmetros foram estimados com precisão melhor que 0.001.

Desta vez o algoritmo convergiu para o mínimo global e o modelo obtido é uma boa aproximação do processo. O valor da função custo obtida foi

$$V(\bar{\theta}^2) = \sum_{i=1}^{1000} (\varepsilon(t, \bar{\theta}^2))^2 = 0,2188 .$$

A resposta ao degrau deste modelo é mostrada na Figura 6.15.

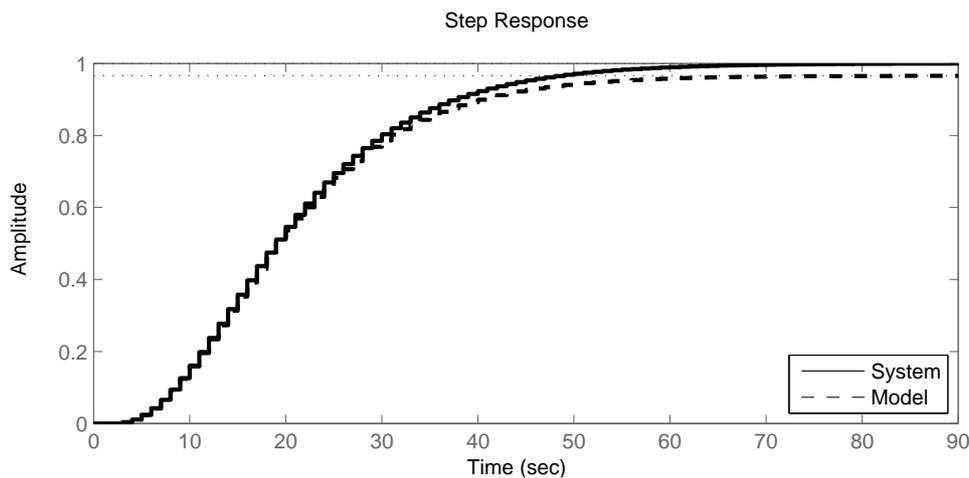


Figura 6.15: Resposta ao degrau do sistema real e do modelo.

## 6.4 Considerações finais

Neste capítulo foram apresentadas restrições para o projeto de experimentos que podem ser utilizadas para melhorar a convergência dos métodos de identificação por erro de predição. As contribuições inéditas deste capítulo foram em parte publicadas em (ECKHARD; BAZANELLA, 2011b). As restrições garantem que a função  $\bar{V}(\theta)$  seja semi-convexa em um conjunto  $\mathcal{C}$  pré-escolhido. Foi apresentada uma abordagem estocástica baseada no *Scenario Approach*, a qual é necessária mas não é suficiente. Uma análise desta abordagem foi feita e foram apresentados níveis de confiança para o procedimento proposto. Foi também apresentada uma condição suficiente baseada em *S-procedure*. Esta restrição é conservadora por dois motivos: ela exige que todos os termos de um somatório (6.3) sejam positivos e utiliza o *S-procedure*, o qual é conservador em muitos casos. Foram apresentados exemplos para as duas abordagens que mostraram que é possível utilizar as condições propostas em conjunto com o projeto de experimentos.

## 7 FILTRAGEM ITERATIVA

Neste capítulo será apresentada mais uma contribuição inédita desta tese, a qual é a segunda ferramenta proposta melhorar a convergência dos métodos de identificação por erro de predição. Nos capítulos anteriores foi mostrado que é possível modificar a função custo para aumentar a região em que a função é semi-convexa. O tamanho desta região depende de vários fatores, dos quais foi ressaltado no capítulo anterior o espectro do sinal de entrada. Foi proposto um método para a escolha do espectro do sinal de entrada de maneira a aumentar o tamanho dessa região.

O projeto de experimentos é bastante eficaz quando é possível realizar um experimento em que se pode escolher o sinal de entrada. Contudo, em muitas aplicações não existe liberdade para a escolha do experimento e portanto o projeto de experimentos não pode ser utilizado. Um exemplo muito claro desta limitação está nos casos em que existem dados históricos sobre um determinado processo, os quais serão utilizados para a identificação do modelo. Como os dados já foram coletados, não é possível mais modificar os sinais. Neste caso, é necessário utilizar outras ferramentas para aumentar o tamanho da região que a função  $\bar{V}(\theta)$  é semi-convexa e facilitar a convergência para o mínimo global da função custo.

A primeira ferramenta proposta nesta tese foi a utilização de restrições ao projeto do espectro do sinal de entrada. Modificando o espectro dos sinais estamos mudando a função custo utilizada. Entretanto, o projeto do sinal de entrada não é a única maneira de modificar os sinais utilizados. Se os dados coletados do processo forem filtrados, obteremos outros sinais cujos espectros são diferentes. A segunda ferramenta proposta nesta tese para melhorar a convergência dos algoritmos é baseada na filtragem dos dados coletados, com o objetivo de modificar o espectro dos sinais utilizados e aumentar o tamanho da região que a função  $\bar{V}(\theta)$  é semi-convexa.

Vamos considerar novamente que o processo do qual desejamos obter um modelo é descrito por

$$y(t) = G_0(q)u(t) + H_0(q)e_0(t).$$

Deste processo é possível coletar um conjunto de dados

$$\mathcal{Z}^N = \{u(1), y(1), \dots, u(N), y(N)\}.$$

Vamos considerar agora que tanto os dados de entrada como os dados de saída são filtrados pelo mesmo filtro  $L(q)$ , resultando nos sinais:

$$y_L(t) = L(q)y(t) \quad u_L(t) = L(q)u(t) \quad (7.1)$$

os quais são as versões filtradas dos sinais  $y(t)$  e  $u(t)$ . Estes dados são coletados e formam o conjunto

$$\mathcal{Z}_L^N = \{u_L(1), y_L(1), \dots, u_L(N), y_L(N)\}.$$

O filtro  $L(q)$  é linear, invariante no tempo, BIBO-estável e mônico. Na sequência vamos apresentar maneiras para projetar o filtro  $L(q)$  para melhorar a identificação do modelo. Note por enquanto que o espectro dos novos sinais depende do filtro utilizado.

$$\Phi_{y_L}(\omega) = |L(e^{j\omega})|^2 \Phi_y(\omega) \quad \Phi_{u_L}(\omega) = |L(e^{j\omega})|^2 \Phi_u(\omega). \quad (7.2)$$

Utilizando os sinais  $u_L(t)$  e  $y_L(t)$  a função custo torna-se

$$\begin{aligned} \bar{V}_L(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_u(\omega) |L(e^{j\omega})H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 d\omega \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_0 |L(e^{j\omega})H^{-1}(e^{j\omega}, \theta)H_0(e^{j\omega})|^2 d\omega. \end{aligned} \quad (7.3)$$

Vamos assumir novamente que o modelo possui uma estrutura fixa determinada por

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + F^T(q)\theta}, \quad H(q, \theta) = \frac{1 + C^T(q)\theta}{1 + D^T(q)\theta}$$

onde os vetores  $B(q), C(q), D(q), F(q) \in \mathbb{R}^n$  são formados por funções de transferência conhecidas e  $\theta \in \mathbb{R}^n$  é o vetor de parâmetros que se deseja estimar.

Neste caso o gradiente da função custo é dado por

$$\begin{aligned} \nabla \bar{V}_L(\theta) &= \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u |L|^2 \Re \left\{ [H^{-1}(\theta)(G_0 - G(\theta))]^* \right. \\ &\quad \cdot \left[ \left( \frac{D}{1 + C^T\theta} - \frac{(1 + D^T\theta)C}{(1 + C^T\theta)^2} \right) (G_0 - G(\theta)) \right. \\ &\quad \left. \left. - H^{-1}(\theta) \left( \frac{B}{1 + F^T\theta} - \frac{B^T\theta F}{(1 + F^T\theta)^2} \right) \right] \right. \\ &\quad \left. + \lambda_0 [H_0 H^{-1}(\theta)]^* H_0 \left( \frac{D}{1 + C^T\theta} - \frac{(1 + D^T\theta)C}{(1 + C^T\theta)^2} \right) \right\} d\omega. \end{aligned} \quad (7.4)$$

Usando esta expressão para o gradiente da função custo podemos obter uma condição equivalente para  $(\theta - \theta_0)^T \nabla \bar{V}_L(\theta) > 0$  dada por

$$\begin{aligned} \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u |L|^2 |H^{-1}(\theta)(G_0 - G(\theta))|^2 \Re \left\{ \frac{1 + C^T\theta_0}{1 + C^T\theta} + \frac{1 + F^T\theta_0}{1 + F^T\theta} - \frac{1 + D^T\theta_0}{1 + D^T\theta} \right\} \\ + \lambda_0 \Re \left\{ (LH_0H^{-1} - 1)^* (LH_0H^{-1}(\theta) - L) \frac{1 + C^T\theta_0}{1 + C^T\theta} \right\} d\omega > 0. \end{aligned} \quad (7.5)$$

Vamos utilizar a condição (7.5) como referência no projeto do filtro  $L(q)$  para expandir a região que a função custo é semi-convexa. Como a condição (7.5) é bastante complicada vamos inicialmente trabalhar com uma versão mais simples, válida no caso em que o modelo possui uma estrutura Output-Error, para depois generalizar para outras estruturas.

## 7.1 Estrutura Output-Error generalizada

Quando o modelo possui uma estrutura Output-Error generalizada, a condição (7.5) pode ser simplificada como

$$\int_{-\pi}^{\pi} \Phi_u(\omega) |L(e^{j\omega})|^2 |H^{-1}(e^{j\omega}, \theta)(G_0(e^{j\omega}) - G(e^{j\omega}, \theta))|^2 \Re \left\{ \frac{1 + F^T(e^{j\omega})\theta_0}{1 + F^T(e^{j\omega})\theta} \right\} d\omega > 0. \quad (7.6)$$

Note que na condição (7.6) aparece uma integral cujo integrando é composto de vários termos dos quais o único que pode ser negativo é

$$\Re \{K_F(e^{j\omega}, \theta)\} \triangleq \Re \left\{ \frac{1 + F^T(e^{j\omega})\theta_0}{1 + F^T(e^{j\omega})\theta} \right\}.$$

Logo, a positividade da integral depende da função  $K_F(e^{j\omega})$ . Esta função pode ser reescrita como

$$K_F(e^{j\omega}, \theta) = \frac{(1 + p_0^1 e^{-j\omega})(1 + p_0^2 e^{-j\omega}) \cdots (1 + p_0^n e^{-j\omega})}{(1 + p^1 e^{-j\omega})(1 + p^2 e^{-j\omega}) \cdots (1 + p^n e^{-j\omega})}.$$

onde  $p^i$  são os  $n$  zeros do polinômio  $(1 + F^T(q)\theta)$  e  $p_0^i$  são os  $n$  zeros do polinômio  $(1 + F^T(q)\theta_0)$ .

A resposta desta função na frequência  $\omega = 0$  é

$$K_F(e^{j0}, \theta) = \frac{(1 + p_0^1)(1 + p_0^2) \cdots (1 + p_0^n)}{(1 + p^1)(1 + p^2) \cdots (1 + p^n)}$$

enquanto que a resposta na frequência  $\omega = \pi$  é

$$K_F(e^{j\pi}, \theta) = \frac{(1 - p_0^1)(1 - p_0^2) \cdots (1 - p_0^n)}{(1 - p^1)(1 - p^2) \cdots (1 - p^n)}.$$

Portanto se todos os polos  $p^i$  e zeros  $p_0^i$  da função de transferência  $K(e^{j\omega}, \theta)$  estiverem dentro do círculo unitário, tanto  $K_F(e^{j0}, \theta)$  como  $K_F(e^{j\pi}, \theta)$  são reais e positivos. Note que o numerador de  $K_F(e^{j\omega}, \theta)$  é o denominador de  $G_0(q)$  e que o denominador de  $K_F(e^{j\omega}, \theta)$  é o denominador de  $G(q, \theta)$ . Como o processo é estável e o modelo é estável para todo  $\theta$  que pertence ao conjunto  $\Gamma$ , então  $K_F(e^{j0}, \theta)$  e  $K_F(e^{j\pi}, \theta)$  são positivos  $\forall \theta \in \Gamma$ . Como a função  $K_F(e^{j\omega}, \theta)$  é contínua nós podemos chegar no seguinte resultado.

**Lema 7.1** *Para todo  $\theta \in \Gamma$ , existem  $\omega_l, \omega_h \in \mathbb{R}$  tais que*

$$\begin{aligned} \Re \{K_F(e^{j\omega}, \theta)\} &> 0 \quad 0 \leq \omega \leq \omega_l, \\ \Re \{K_F(e^{j\omega}, \theta)\} &> 0 \quad \omega_h \leq \omega \leq \pi. \end{aligned} \quad (7.7)$$

De acordo com este resultado, a parte real da função de transferência  $K_F(e^{j\omega}, \theta)$  é positiva para frequências suficientemente baixas e para frequências suficientemente altas. Observe que esta propriedade não depende dos parâmetros do processo  $G_0(q)$ ,  $H_0(q)$  e  $\lambda_0$ . Contudo, o valor exato de  $\omega_l$  e  $\omega_h$  depende das funções do processo. Esta propriedade será muito útil para a escolha do filtro usado para modificar a região em que a função  $V(\theta)$  é semi-convexa.

Observe que o integrando de (7.6) é sempre positivo para as frequências baixas e altas, mas não se sabe se o integrando é positivo nas frequências intermediárias. Se fosse possível enfatizar as frequências baixas e altas, então seria possível garantir a positividade da integral. De fato, esta é a abordagem deste capítulo, e o filtro  $L(q)$  é utilizado para enfatizar algumas faixas de frequência e para negligenciar outras.

A principal ideia consiste em escolher um filtro  $L(q)$  que tenha baixo ganho nas frequências intermediárias, de maneira que a integral (7.6) seja positiva para todo  $\theta \in \Gamma$ . Quando os dados são filtrados é removida uma parte importante da informação contida. As frequências removidas são as intermediárias, as quais tipicamente contém uma parte

significativa da informação sobre o comportamento do processo. Por consequência, a qualidade da estimativa obtida com os dados filtrados é menor do que se fossem utilizados os dados originais.

Para evitar esta perda de informação é proposto um algoritmo iterativo. Este algoritmo utiliza inicialmente dados que possuem informações apenas sobre as baixas frequências do processo. Utilizando estes dados, vamos obter um domínio de atração candidato que é relativamente grande, mas a variância da estimativa será alta. A estimativa obtida nesta etapa será utilizada como condição inicial da próxima etapa. Na próxima etapa, vamos utilizar um filtro que remove menos informação do conjunto de dados. Nesta caso, vamos obter um domínio de atração menor que o obtido anteriormente, mas a variância da estimativa será menor. Espera-se que a estimativa obtida na etapa anterior esteja dentro do domínio de atração da etapa atual, e que então o algoritmo de otimização convirja para o mínimo global da função custo, utilizando a estimativa anterior como condição inicial. O procedimento pode ser repetido diversas vezes, até que finalmente sejam utilizados os dados originais, os quais geram a melhor estimativa possível.

O algoritmo pode ser descrito pelos seguintes passos:

1. Colete um conjunto de dados de entrada e saída ( $y(t)$  e  $u(t)$ ) do processo e faça  $k = 0$ ;
2. Escolha uma condição inicial  $\theta^1$ ;
3. Escolha um filtro passa-baixa  $L_k(q)$  com banda passante bastante baixa;
4. Obtenha os dados filtrados  $u_L(t) = L_k(q)u(t)$  e  $y_L(t) = L_k(q)y(t)$ ;
5. Realize a identificação do modelo com os dados  $u_L(t)$  e  $y_L(t)$  e com a condição inicial  $\theta^k$ . Faça  $\theta^{k+1}$  igual a solução do procedimento de identificação;
6. Escolha um filtro passa-baixas  $L_{k+1}(q)$  com banda passante maior que  $L_k(q)$ ;
7. Faça  $k = k + 1$  e vá para o passo 4.

Após algumas iterações o filtro  $L_k(q)$  pode ser escolhido como  $L_k(q) = 1$ , o que significa que os dados não são mais filtrados e que o procedimento de identificação utilizará os dados originais.

A principal ideia deste algoritmo é que se  $L_k(q)$  é similar a  $L_{k+1}(q)$  então o mínimo global das respectivas funções custos também é similar, e portanto o mínimo global  $\theta^k$  é uma boa condição inicial para o próximo procedimento de identificação que usará os dados filtrados por  $L_{k+1}(q)$ . Se os filtros  $L_k(q)$  são suficientemente próximos então este algoritmo iterativo converge para o mínimo global da função custo  $V^N(\theta)$ .

## 7.2 Estrutura ARMAX generalizada

Quando o modelo possui uma estrutura ARMAX generalizada, a condição (7.5) pode ser simplificada como

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u |L|^2 |H^{-1}(\theta)(G_0 - G(\theta))|^2 \Re \left\{ \frac{1 + C^T \theta_0}{1 + C^T \theta} \right\} + \lambda_0 \Re \left\{ (LH_0 H^{-1} - 1)^* (LH_0 H^{-1}(\theta) - L) \frac{1 + C^T \theta_0}{1 + C^T \theta} \right\} d\omega > 0. \quad (7.8)$$

Note que na condição (7.8) aparece o termo

$$K_C(e^{j\omega}, \theta) \triangleq \frac{1 + C^T(e^{j\omega})\theta_0}{1 + C^T(e^{j\omega})\theta}$$

o qual é similar ao termo  $K_F(e^{j\omega}, \theta)$  apresentado anteriormente. Utilizando os mesmos argumentos apresentados na seção anterior pode-se mostrar que

$$\Re \{K_C(e^{j\omega}, \theta)\} > 0$$

para frequências baixas e para frequências altas.

Vamos considerar agora que o filtro  $L(q)$  é um passa-baixas tal que

$$L(e^{j\omega}) = \begin{cases} 1 & \text{se } -\omega_a < \omega < \omega_a \\ 0 & \text{se } -\pi \leq \omega \leq -\omega_a \text{ ou } \omega_a \leq \omega \leq \pi \end{cases}$$

Neste caso a condição (7.8) pode ser escrita como

$$\frac{1}{\pi} \int_{-\omega_a}^{\omega_a} \left( \Phi_u |H^{-1}(\theta)(G_0 - G(\theta))|^2 + \lambda_0 |H_0 H^{-1} - 1|^2 \right) \Re \left\{ \frac{1 + C^T \theta_0}{1 + C^T \theta} \right\} d\omega > 0. \quad (7.9)$$

Se a frequência do filtro  $\omega_a$  for escolhida suficientemente pequena então a condição (7.9) será respeitada pois neste caso apenas as frequências em que  $K_C(e^{j\omega}, \theta) > 0$  serão utilizadas.

Neste caso, para expandir a região que a função custo é semi-convexa pode-se utilizar o mesmo algoritmo proposto para a estrutura Output-Error. Em cada iteração será escolhido um valor  $\omega_a$  um pouco maior que na iteração anterior, de maneira que a banda passante do filtro será incrementada iteração por iteração.

### 7.3 Estrutura Box-Jenkins generalizada

A estrutura Box-Jenkins generalizada é similar à estrutura geral sendo que a única especificidade é que as funções de transferência  $G(q, \theta)$  e  $H(q, \theta)$  são parametrizadas independentemente. Utilizando a notação utilizada nos outros capítulos podemos escrever as funções de transferência como

$$G(q, \eta) = \frac{B^T(q)\eta}{1 + F^T(q)\eta}, \quad H(q, \xi) = \frac{1 + C^T(q)\xi}{1 + D^T(q)\xi}.$$

onde

$$\theta = \begin{bmatrix} \eta \\ \xi \end{bmatrix} \text{ e } \theta_0 = \begin{bmatrix} \eta_0 \\ \xi_0 \end{bmatrix}.$$

Neste caso a condição (7.5) pode ser simplificada como

$$\begin{aligned} & \frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u |L|^2 |H^{-1}(\xi)(G_0 - G(\eta))|^2 \Re \left\{ \frac{1 + C^T \xi_0}{1 + C^T \xi} + \frac{1 + F^T \eta_0}{1 + F^T \eta} - \frac{1 + D^T \xi_0}{1 + D^T \xi} \right\} \\ & + \lambda_0 \Re \left\{ (LH_0 H^{-1}(\xi) - 1)^* (LH_0 H^{-1}(\xi) - L) \frac{1 + C^T \xi_0}{1 + C^T \xi} \right\} d\omega > 0. \quad (7.10) \end{aligned}$$

A relação do filtro  $L(q)$  na condição acima é bastante complicada e portanto vamos dividir o procedimento de identificação em duas etapas. Na primeira etapa vamos identificar o vetor de parâmetros  $\eta$  enquanto que na etapa seguinte será identificado o vetor parâmetros  $\xi$ .

Para identificar o parâmetro  $\eta$ , a variável  $\xi$  será fixada em um valor aleatório  $\xi = \xi_a$ . Como o modelo do ruído  $H(\xi_a)$  é fixo, a condição (7.10) pode ser simplificada como

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \Phi_u |L|^2 |H^{-1}(\xi_a)(G_0 - G(\eta))|^2 \Re \left\{ \frac{1 + F^T \eta_0}{1 + F^T \eta} \right\} d\omega > 0. \quad (7.11)$$

Observe que o segundo termo da condição some, pois o modelo do ruído foi fixado e portanto seu gradiente é nulo.

Note que a condição acima é similar à condição obtida para a estrutura Output-Error. Logo, podemos utilizar o algoritmo proposto anteriormente para obter a estimativa de  $\hat{\eta}$ . Ao final do procedimento espera-se que o algoritmo tenha convergido para o mínimo global do critério e portanto  $G(q, \hat{\eta}) \approx G_0(q)$ .

Agora que já foi obtida a estimativa para  $\eta$  podemos partir para a segunda etapa do procedimento, que consiste em identificar o modelo do ruído. Como já obtivemos a estimativa correta ( $\eta \approx \eta_0$ ), a condição (7.10) será simplificada como

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \lambda_0 \Re \left\{ (LH_0H^{-1}(\xi) - 1)^* (LH_0H^{-1}(\xi) - L) \frac{1 + C^T \xi_0}{1 + C^T \xi} \right\} d\omega > 0.$$

Observe que a condição acima é similar à condição obtida para a estrutura ARMAX. Neste caso podemos novamente utilizar o procedimento proposto, o qual consiste da filtragem iterativa, para obter uma estimativa para o parâmetro  $\xi$ .

Novamente vamos considerar que o filtro  $L(q)$  é um passa-baixas tal que

$$L(e^{j\omega}) = \begin{cases} 1 & \text{se } -\omega_a < \omega < \omega_a \\ 0 & \text{se } -\pi \leq \omega \leq -\omega_a \text{ ou } \omega_a \leq \omega \leq \pi \end{cases}$$

Neste caso a condição acima pode então ser escrita como

$$\frac{1}{\pi} \int_{-\omega_a}^{\omega_a} \lambda_0 |H_0H^{-1}(\xi) - 1|^2 \Re \left\{ \frac{1 + C^T \xi_0}{1 + C^T \xi} \right\} d\omega > 0.$$

A cada iteração será escolhido um valor  $\omega_a$  um pouco maior que na iteração anterior, de maneira que a banda passante do filtro seja incrementada iteração por iteração e a estimativa correta do mínimo global do critério seja obtida.

## 7.4 Exemplo

Considere que o processo é descrito por

$$G_0(q) = \left( -0, 2073q^{-1} + 0, 1815q^{-2} + 1, 352q^{-3} - 3, 356q^{-4} \right. \\ \left. + 3, 061q^{-5} - 1, 045q^{-6} - 0, 07957q^{-7} + 0, 0944q^{-8} \right) / \\ \left( 1 - 6, 549q^{-1} + 19, 2q^{-2} - 32, 83q^{-3} + 35, 77q^{-4} - 25, 39q^{-5} \right. \\ \left. + 11, 45q^{-6} - 3q^{-7} + 0, 3491q^{-8} \right),$$

$$H_0(q) = 1, \quad \lambda_0 = 0, 01.$$

Este modelo representa um sistema de potência que consiste de dois geradores síncronos conectados à rede (ANDERSON; FOUAD, 1977). A função de transferência  $G_0(q)$  descreve a relação entre a tensão no enrolamento de campo do primeiro gerador com a velocidade angular no mesmo gerador, considerando um dado ponto de operação.

Vamos identificar um modelo que tem uma estrutura Output-Error

$$G(q, \theta) = \frac{B^T(q)\theta}{1 + F^T(q)\theta} \quad H(q, \theta) = 1$$

onde

$$B(q) = [q^{-1} \ q^{-2} \ q^{-3} \ q^{-4} \ q^{-5} \ q^{-6} \ q^{-7} \ q^{-8} \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T,$$

$$F(q) = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ q^{-1} \ q^{-2} \ q^{-3} \ q^{-4} \ q^{-5} \ q^{-6} \ q^{-7} \ q^{-8}]^T.$$

O sistema pertence à classe de modelos onde

$$\theta^0 = [-0,2073 \ 0,1815 \ 1,352 \ -3,356 \ 3,061 \ -1,045 \ -0,07957 \ 0,0944 \\ -6,549 \ 19,2 \ -32,83 \ 35,77 \ -25,39 \ 11,45 \ -30,3491]^T$$

Com o objetivo de evitar a convergência para mínimos locais da função custo será utilizada a *filtragem iterativa* proposta neste trabalho. A cada iteração será utilizado um filtro FIR passa-baixas de ordem 20 com diferente banda passante. Serão utilizadas 6 iterações do procedimento, nas quais a frequência de corte será respectivamente  $0,4\pi$ ,  $0,5\pi$ ,  $0,6\pi$ ,  $0,7\pi$ ,  $0,8\pi$  e  $0,9\pi$ .

Para comparação, o modelo também será estimado utilizando algoritmos comerciais disponíveis com o software MATLAB. Serão usadas as plataformas `ident` e `unit`, além do método proposto neste trabalho.

O mesmo sinal de entrada  $u(t)$  será utilizado nos três casos, para gerar os dados que serão utilizados para identificar os modelos. O sinal é ruído branco gaussiano com variância unitária ( $\lambda_0 = 1$ ) e tamanho  $N = 1.000$  amostras. Com este sinal de entrada, a variância do sinal de saída é aproximadamente  $\lambda_y = 100$ .

Para cada um dos três algoritmos foram feitas 100 rodadas de Monte-Carlo, gerando portanto 100 estimativas para os parâmetros do modelo. Os resultados obtidos com a plataforma `unit` são mostrados na Figura 7.1, onde o eixo das ordenadas apresenta o valor de cada um dos 16 elementos do vetor de parâmetros  $\theta$  e o eixo das abcissas apresenta cada uma das 100 rodadas de Monte-Carlo. Os valores foram reordenados em termos do menor  $\theta$ . Pode-se ver que os 45 primeiros pontos do gráfico apresentam valores próximos dos parâmetros do processo  $\theta^0$ , indicando que nestas 45 rodadas o algoritmo convergiu para o mínimo global da função custo. Contudo nas outras 55 rodadas os valores estão próximos de um outro ponto no espaço de parâmetros, distante de  $\theta^0$ , indicando que o algoritmo convergiu para um outro ponto da função custo.

Quando a plataforma `ident` foi utilizada, apenas 17 das 100 rodadas convergiram para o mínimo global da função custo, como pode ser visto na Figura 7.2.

Na Figura 7.3 é apresentado o mesmo gráfico para o caso em que o algoritmo proposto, que consiste da filtragem iterativa, é utilizado. Observe que o mínimo global da função custo foi obtido em todas as rodadas.

Para ilustrar melhor os resultados, vamos calcular o valor médio do vetor de parâmetros,

$$\theta_m = \frac{1}{100} \sum_{i=1}^{100} \hat{\theta}_N^i$$



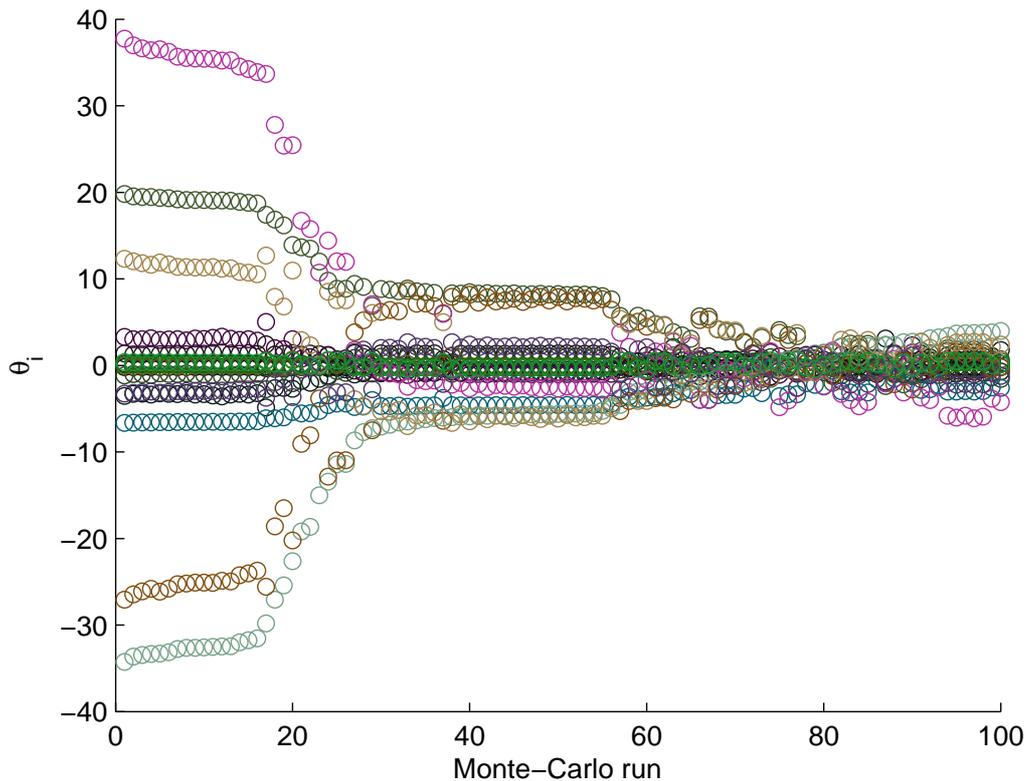


Figura 7.2: Rodadas de Monte-Carlo com a plataforma ident.

## 7.5 Considerações finais

Neste capítulo foi apresentada uma ferramenta para identificação de sistemas que melhora a convergência dos algoritmos de otimização para o mínimo global da função custo. A ferramenta foi criada observando propriedades da função custo relacionadas com o espectro dos sinais utilizados pelos métodos de identificação. Foi mostrado que para diversas estruturas, se os espectros dos sinais são concentrados nas baixas e altas frequências, então a função custo possui um formato que facilita a convergência para o mínimo global. Utilizando esta propriedade, foi proposto um algoritmo iterativo que filtra os dados disponíveis para que os espectros sejam concentrados nas faixas de frequência desejada. Como os filtros retiram informações importantes sobre a dinâmica do processo, o algoritmo a cada iteração utiliza um filtro com maior banda passante. A cada iteração, menos informação é retirada dos sinais, e como consequência é ampliada a quantidade de informação contida no conjunto de dados. O algoritmo proposto foi comparado com plataformas de identificação comerciais e se mostrou consideravelmente melhor nos exemplos que foram testados.

Este algoritmo é a segunda ferramenta proposta nesta tese e também é uma contribuição inédita. Alguns dos resultados deste capítulo foram publicados em (ECKHARD et al., 2012).

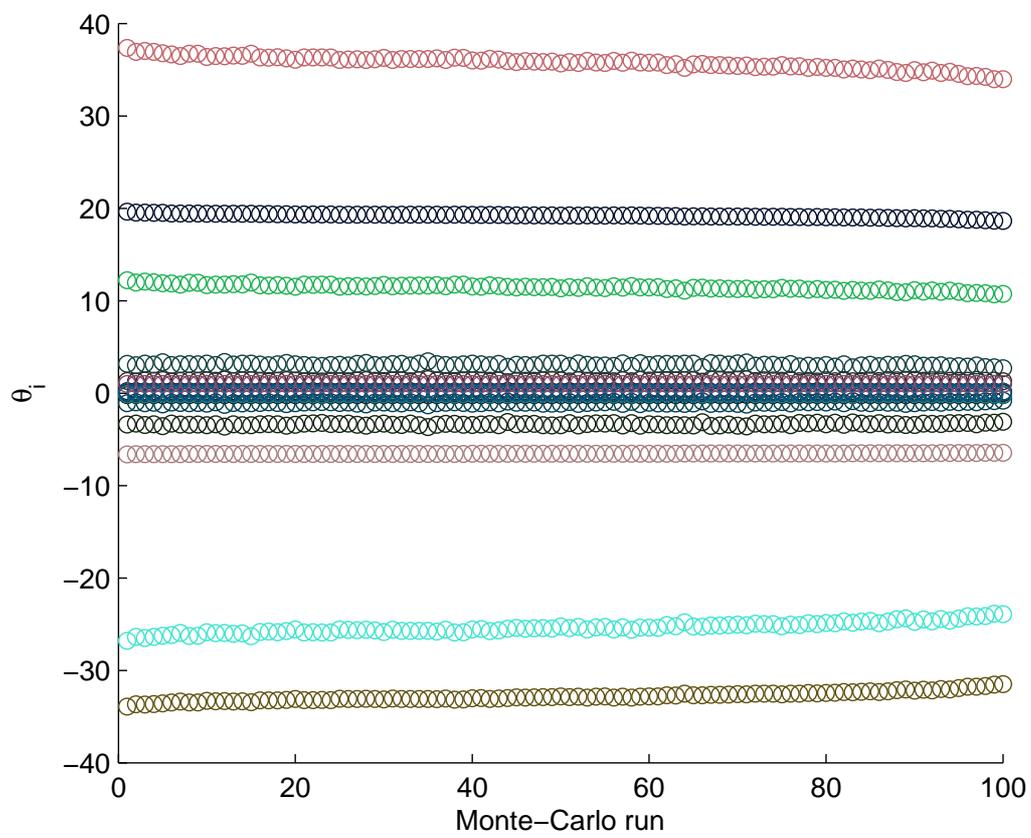


Figura 7.3: Rodadas de Monte-Carlo com o algoritmo proposto.

## 8 CONCLUSÃO

Este trabalho apresentou técnicas para melhorar a convergência dos algoritmos de otimização utilizados com o método de identificação pela minimização do erro de predição. Este método de identificação consiste na busca dos parâmetros, de um modelo com estrutura fixa, que melhor descrevem a dinâmica do processo. O método é descrito como um problema de otimização, o qual procura pelos parâmetros que minimizam o erro médio quadrático entre a saída do processo e a predição do modelo. O critério de otimização não é convexo e para procurar o mínimo global do critério são empregados algoritmos iterativos.

A convergência dos algoritmos iterativos, para o mínimo global do critério de otimização, foi estudada utilizando a teoria de sistemas não-lineares. Considerou-se que os algoritmos são sistemas não-lineares em tempo discreto e foi realizada uma análise de estabilidade dos pontos de equilíbrio destes algoritmos utilizando a teoria de Lyapunov. Uma contribuição inédita desta tese foi a análise da robustez destes algoritmos à incertezas.

Foi realizada uma análise do critério de otimização do método de identificação pela minimização do erro de predição. Esta análise resultou em condições para que a função custo seja semi-convexa em um conjunto. Estas condições são contribuições inéditas desta tese. Observou-se nestas condições que o espectro do sinal de entrada utilizado no experimento exerce grande influência na função custo. Foi então proposta a manipulação do espectro do sinal de entrada com o objetivo de ampliar o tamanho da região em que a função custo é semi-convexa, e melhorar portanto a convergência para o mínimo global do critério de otimização. A manipulação do espectro do sinal de entrada, com o objetivo de modificar a função custo e melhorar a convergência dos algoritmos de otimização, é outra contribuição inédita desta tese. Foram então propostas duas ferramentas para manipular o espectro dos sinais de entrada: a primeira é baseada no projeto de experimentos enquanto que a segunda é baseada na filtragem dos dados. Ambas as ferramentas são contribuições desta tese.

A escolha de espectros dos sinais de entrada é usualmente realizada durante a fase de projeto do experimento. Esta escolha pode ser realizada utilizando um problema de otimização, onde as características desejadas para o espectro são impostas através de restrições ao problema. Foram apresentadas diversas restrições de qualidade que podem ser descritas de forma convexa, o que facilita a resolução do problema de otimização. Foram então propostas restrições de convergência para o projeto de experimentos. Estas restrições garantem que a função custo seja semi-convexa em um conjunto escolhido pelo usuário. Duas abordagens foram utilizadas para obter as restrições de convergência. A primeira abordagem foi baseada no *Scenario Approach*, o qual resultou em condições necessárias para que a função custo seja semi-convexa em um conjunto. Foram apresentados também

os intervalos de confiança desta abordagem estocástica. A segunda abordagem foi baseada em *S-procedure* e resultou em condições suficientes para garantir que a função custo seja semi-convexa em um conjunto.

Em alguns processos não é possível realizar um experimento para coletar os dados. Nestes casos, não se pode utilizar as técnicas baseadas no projeto de experimento para melhorar a convergência dos algoritmos de otimização. Foi então proposta neste trabalho uma outra técnica para melhorar a convergência dos algoritmos de otimização. Esta técnica é baseada na filtragem dos dados, com o objetivo de modificar o espectro dos sinais utilizados pelos algoritmos de otimização. Observou-se que quando os dados são filtrados é removida uma parte importante da informação que estava contida. Para minimizar a perda de informação, foi proposta uma filtragem iterativa dos dados, que em cada iteração remove uma quantidade menor de informação dos dados originais. Desta maneira pode-se utilizar a filtragem para melhorar a convergência dos algoritmos de otimização sem a desvantagem da perda de informação. Os métodos propostos foram comparados com plataformas comerciais de identificação de sistemas, e os resultados obtidos mostraram o melhor desempenho dos métodos propostos.

Neste trabalho foi considerado que todos os experimentos realizados para coletar dados são realizados em malha-aberta. Num futuro não distante pretende-se expandir os resultados obtidos nesta tese para o caso em que os dados são coletados de um experimento em malha-fechada. Outra hipótese utilizada neste tese é que a estrutura do modelo consegue representar o sistema do qual se deseja obter o modelo. Pretende-se obter resultados sobre a convergência dos algoritmos para os casos em que esta hipótese não é respeitada.

## APÊNDICE A

### A.1 Teorema KYP

Seja  $A \in \mathbb{C}^{n \times n}$ ,  $B \in \mathbb{C}^{n \times m}$  e  $M \in \mathbb{H}^{n+m}$  então as seguintes afirmativas são equivalentes (YIPING, 2007).

1. Existe  $P \in \mathbb{H}^n$  tal que

$$\begin{bmatrix} P - A^*PA & -A^*PB \\ B^*PA & B^*PB \end{bmatrix} + M > 0$$

2. Para todo  $\omega \in \mathbb{R}$ ,  $u \in \mathbb{C}^n$ ,  $v \in \mathbb{C}^m$ , se

- 

$$\begin{bmatrix} u \\ v \end{bmatrix} \neq 0$$

- 

$$(e^{j\omega}I - A)u = Bv$$

então

$$\begin{bmatrix} u^* & v^* \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} > 0.$$

Se  $(A, B)$  é controlável então a equivalência se mantém entre as duas afirmativas se todas as ocorrências de  $>$  são trocadas por  $\geq$ .

### A.2 Teorema de Parseval

Seja  $s(t)$  um sinal quasi-estacionário  $\Phi_s(\omega)$  seu espectro. Então

$$\bar{E}s^2(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_s(\omega) d\omega.$$

## REFERÊNCIAS

- ANDERSON, P.; FOUAD, A. **Power System Control and Stability**. New York: IEEE Press, 1977.
- ARMIJO, L. Minimization of functions having continuous partial derivatives. **Pacific Journal of Mathematics**, Berkeley, v.16, n.1, p.1–3, 1966.
- ÅSTRÖM, K. J. **Introduction to Stochastic Control Theory**. New York: Academic Press, 1970.
- ÅSTRÖM, K. J.; SÖDERSTRÖM, T. Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model. **IEEE Transactions on Automatic Control**, Brookline, v.19, n.6, p.769–773, 1974.
- BARENTHIN, M. et al. Identification for control of multivariable systems: controller validation and experiment design via lmis. **Automatica**, Elmsford, v.44, n.12, p.3070–3078, 2008.
- BAZANELLA, A. S.; CAMPESTRINI, L.; ECKHARD, D. **Data-driven Controller Design: the  $H_2$  approach**. Netherlands: Springer, 2012.
- BAZANELLA, A. S. et al. Iterative minimization of control performance criteria. **Automatica**, Elmsford, v.44, n.10, p.2549–2559, 2008.
- BOMBOIS, X. et al. Least costly identification experiment for control. **Automatica**, Elmsford, v.42, n.10, p.1651–1662, 2006.
- BOYD, S. et al. **Linear matrix inequalities in system & control theory**. Philadelphia: Society for Industrial and Applied Mathematic, 1994.
- CAMPESTRINI, L. et al. Identificação não-linear de um biorreator através da minimização do erro de predição. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 19., 2012, Campina Grande. **Anais...** São Paulo: SBA, 2012. p.3066–3072.
- CAMPI, M. C.; GARATTI, S.; PRANDINI, M. The scenario approach for systems and control design. **Annual Reviews in Control**, Oxford, v.33, n.2, p.149 – 157, 2009.
- ECKHARD, D.; BAZANELLA, A. Data-based controller tuning: improving the convergence rate. In: IEEE CONFERENCE ON DECISION AND CONTROL, 49., 2010, Atlanta. **Proceedings...** New York: IEEE, 2010. p.4801–4806.

- ECKHARD, D.; BAZANELLA, A. S. Optimizing the Convergence of Data-Based Controller Tuning. In: EUROPEAN CONTROL CONFERENCE, 2009, Budapest. **Proceedings...** Budapest: EUCA, 2009. p.910–915.
- ECKHARD, D.; BAZANELLA, A. S. Optimizing the convergence of data-based controller tuning. **Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering**, Bath, v.226, n.4, p.563–574, 2011.
- ECKHARD, D.; BAZANELLA, A. S. On the global convergence of identification of output error models. In: IFAC WORLD CONGRESS, 18., 2011, Milan. **Proceedings...** Laxemburg: IFAC, 2011. p.9058–9063.
- ECKHARD, D.; BAZANELLA, A. S. Robust convergence of the steepest descent method for data-based control. **International Journal of Systems Science**, Hants, v.43, n.10, p.1969–1975, 2012.
- ECKHARD, D. et al. On the convergence of the Prediction Error Method to its global minimum. In: SYMPOSIUM ON SYSTEM IDENTIFICATION, 2012, Brussels. **Proceedings...** Laxemburg: IFAC, 2012. p.698–703.
- GERENCSEER, L.; HJALMARSSON, H. Adaptive input design in system identification. In: IEEE CONFERENCE ON DECISION AND CONTROL AND 2005 EUROPEAN CONTROL CONFERENCE, 44., 2005, Seville. **Proceedings...** New York: IEEE, 2005. p.4988–4993.
- GEVERS, M.; BOMBOIS, X. Input design : from open-loop to control-oriented design. In: IFAC SYMPOSIUM ON SYSTEM IDENTIFICATION, 14., 2006, Newcastle. **Proceedings...** Laxemburg: IFAC, 2006. p.1329–1334.
- GOODWIN, G. C.; AGÜERO, J. C.; SKELTON, R. Conditions for Local Convergence of Maximum Likelihood Estimation for ARMAX Models. In: IFAC SYMPOSIUM ON SYSTEM IDENTIFICATION, 13., 2003, Rotterdam. **Proceedings...** Laxemburg: IFAC, 2003. p.797–802.
- HJALMARSSON, H. From experiment design to closed-loop control. **Automatica**, Elmsford, v.41, n.3, p.393–438, 2005.
- HJALMARSSON, H.; JANSSON, H. Closed loop experiment design for linear time invariant dynamical systems via LMIs. **Automatica**, Elmsford, v.44, n.3, p.623–636, 2008.
- HJALMARSSON, H.; MARTENSSON, J. Optimal Input Design for Identification of Non-linear Systems: learning from the linear case. In: AMERICAN CONTROL CONFERENCE, 2007, New York. **Proceedings...** New York: IEEE, 2007. p.1572–1576.
- JANSSON, H. **Experiment Design with Applications in Identification for Control**. Stockholm: KTH, 2004. 207p. Tese (Doutorado) – Royal Institute of Technology, Stockholm, 2004.
- JANSSON, H.; HJALMARSSON, H. Input design via LMIs admitting frequency-wise model specifications in confidence regions. **IEEE Transactions on Automatic Control**, New York, v.50, n.10, p.1534–1549, 2005.

- KHALIL, H. K. **Nonlinear systems**. Upper Saddle River: Prentice Hall, 1996.
- LEE, W. S. et al. On some key issues in the Windsurfer approach to adaptive robust control. **Automatica**, Elmsford, v.31, n.11, p.1619–1636, 1995.
- LINDQVIST, K.; HJALMARSSON, H. Identification for control: adaptive input design using convex optimization. In: IEEE CONFERENCE DECISION AND CONTROL, 40., 2001, Orlando. **Proceedings...** New York: IEEE, 2001. v.5, p.4326–4331.
- LJUNG, L. **System identification: theory for the user**. Englewood Cliffs: Prentice Hall, 1987.
- LJUNG, L. **System Identification Toolbox, User's Guide**. 4.ed. Natick: The Mathworks, 2000.
- LJUNG, L. Prediction error estimation methods. **Circuits, Systems, and Signal Processing**, Cambridge, v.21, n.1, p.11–21, 2002.
- LJUNG, L. Perspectives on system identification. **Annual Reviews in Control**, Oxford, v.34, n.1, p.1–12, 2010.
- MEHRA, R. Optimal input signals for parameter estimation in dynamic systems—Survey and new results. **Automatic Control, IEEE Transactions on**, Brookline, USA, v.19, n.6, p.753–768, 1974.
- POLAK, E. An historical survey of computational methods in optimal control. **SIAM review**, Philadelphia, v.15, n.2, p.553–584, 1973.
- PUTINAR, M. Positive polynomials on compact semi-algebraic sets. **Indiana University Mathematics Journal**, Bloomington, v.42, n.3, p.969–984, 1993.
- RIEDLE, B.; PRALY, L.; KOKOTOVIC, P. Examination of the SPR condition in output error parameter estimation. **Automatica**, Elmsford, v.22, n.4, p.495–498, 1986.
- ROJAS, C.; HJALMARSSON, H.; HILDEBRAND, R. MIMO experiment design based on asymptotic model order theory. In: IEEE CONFERENCE ON DECISION AND CONTROL AND 28TH CHINESE CONTROL CONFERENCE, 48., 2009, Shanghai. **Proceedings...** New York: IEEE, 2009. p.488–493.
- SÖDERSTRÖM, T. On the uniqueness of Maximum Likelihood identification. **Automatica**, Elmsford, v.11, n.2, p.193–197, 1975.
- SÖDERSTRÖM, T.; STOICA, P. Some properties of the output error method. **Automatica**, Elmsford, v.18, n.1, p.93–99, 1982.
- SÖDERSTRÖM, T.; STOICA, P. **System Identification**. Hemel Hempstead, Hertfordshire: Prentice-Hall International, 1989.
- SPIEGEL, M. R. **Manual de Fórmulas, Métodos e Tabelas de Matemática**. 2.ed. São Paulo: Mc Graw Hill, 1992.
- YIPING, C. A Proof of the Discrete-Time KYP Lemma Using Semidefinite Programming Duality. In: CHINESE CONTROL CONFERENCE, 2007, Beijing. **Proceedings...** New York: IEEE, 2007. p.156–160.

ZOU, Y.; HEATH, W. P. Conditions for Attaining Global Minimum in Maximum Likelihood System Identification. In: IFAC SYMPOSIUM ON SYSTEM IDENTIFICATION, 15., 2009, Saint-Malo. **Proceedings...** Laxemburg: IFAC, 2009. p.1110–1115.

ZOU, Y.; HEATH, W. P. Global convergence conditions in maximum likelihood estimation. **International Journal of Control**, Southampton, v.85, n.5, p.475–490, 2012.