RODRIGO ALVITE ROMANO

IDENTIFICAÇÃO DE PROCESSOS NÃO-LINEARES E
QUANTIFICAÇÃO DE ATRITO EM VÁLVULAS DE CONTROLE

São Paulo

2010

RODRIGO ALVITE ROMANO

IDENTIFICAÇÃO DE PROCESSOS NÃO-LINEARES E
QUANTIFICAÇÃO DE ATRITO EM VÁLVULAS DE CONTROLE

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

Área de concentração:
Engenharia de Sistemas
Orientador:
Prof. Dr. Claudio Garcia

São Paulo
2010

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

São Paulo, 06 de Janeiro de 2010.

Assinatura do Autor _____

Assinatura do Orientador _____

FICHA CATALOGRÁFICA

Romano, Rodrigo Alvite

Identificação de processos não-lineares e quantificação de atrito em válvulas de controle / R.A. Romano. – ed.rev. – São Paulo, 2010.

130 p.

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

1. Identificação de sistemas 2. Válvulas de controle pneumático 3. Atrito 4. Otimização não linear I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle II. t.

Este trabalho é dedicado, com todo o meu apreço e carinho, aos meus pais Roberto e Nancy, aos meus irmãos Renato e Iara e à minha querida esposa Vanessa.

Agradecimentos

Ao professor Claudio Garcia, não apenas pela orientação e sábias palavras, mas principalmente pela amistosa convivência nos últimos anos.

Aos mestres: Álvaro Puga Paz, José Carlos de Souza Jr., José Jaime da Cruz e, em especial, Vanderlei Cunha Parro. Cada um a sua maneira, vocês são exemplos para mim.

À Faculdade de Engenharia Industrial (FEI) por proporcionar uma formação acadêmica à altura dos desafios da minha vida profissional.

Aos colegas de jornada Alain Segundo Potts, Átila Madureira Bueno, Daniel Uehara, Flávio Batista, Paulo Henrique da Rocha e Ricardo Bressan Pinheiro pelos momentos de reflexão e descontração durante esta jornada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

A todos os meus familiares por todo apoio, amor e união, mesmo nos momentos mais difíceis. À minha esposa Vanessa por sempre estar ao meu lado. Por fim, gostaria de ressaltar a minha gratidão aos meus pais Roberto e Nancy, pois não sei como eu chegaria até aqui sem o exemplo de fibra e perseverança que vocês sempre demonstraram.

Resumo

O atrito em válvulas e a sintonia inadequada de controladores são duas das maiores causas de degradação no desempenho das malhas de controle que incluem tais dispositivos. Assim como modelos de atrito são necessários para diagnosticar o mau funcionamento das válvulas ou para compensar os efeitos indesejáveis causados pelo atrito, modelos de processos são de fundamental importância para o projeto de controladores. Este trabalho estende métodos existentes para estimar parâmetros de modelos de atrito e processo, de modo que uma estrutura não-linear é adotada para representar o processo. O procedimento é baseado em dados de operação em malha fechada. Os algoritmos de estimação desenvolvidos são testados com dados simulados e gerados por uma plataforma híbrida (composta por uma válvula real e por uma planta simulada de neutralização de pH), a partir da qual avaliam-se as influências de perturbações, da magnitude do sinal de teste e da sintonia do controlador nos modelos estimados. Os resultados demonstram que o nível de atrito é corretamente quantificado, assim como “bons” modelos para o processo são estimados em diversas situações. Além disso, a extensão proposta apresenta vantagens significativas em relação a outros métodos, como: (1) maior exatidão na quantificação do nível de atrito, principalmente para processos em que as não-lineares sejam mais severas e (2) estimativas razoáveis do comportamento estático não-linear.

Palavras-chave: identificação de sistemas; válvulas de controle pneumático; atrito; otimização não-linear.

Abstract

The friction in control valves and inadequate controller tuning are two of the major sources of performance degradation in control loops that include such devices. As friction models are needed to diagnose abnormal valve operation or to compensate such undesirable effects, process models play an essential role in controller design. This work extends existing methods that jointly identify the friction and process model parameters, so that a nonlinear structure is adopted to represent the process model. The procedure is based on data from closed-loop experiments. The developed estimation algorithms are tested with data from simulations and generated by a hybrid setup (composed of a real valve and a simulated pH neutralization process), in which the influences of the process disturbances, of the excitation signal magnitude and of the controller tuning on estimated models are investigated. The results demonstrate that the friction is accurately quantified, as well as “good” process models are estimated in several situations. In addition, the proposed extension presents significant advantages in relation to other methods, such as: (1) greater accuracy for friction quantification, especially for highly nonlinear processes and (2) reasonable estimates of the nonlinear steady state characteristics.

Keywords: system identification; pneumatic control valves; friction; nonlinear optimization.

Lista de Figuras

2.1	Válvula de controle com atuador pneumático	10
2.2	Comparação do comportamento da força de atrito reproduzida pelos modelos clássico e de Karnopp	15
2.3	Entrada e saída reproduzida pelo modelo de histerese e a correspon- dente relação $x(k) = f(F_{ext}(k))$	18
2.4	Comportamento da posição da haste em função do tempo e relação entrada-saída reproduzida pelo modelo de atrito de Kano	19
2.5	Fluxograma do modelo empírico de Kano	21
3.1	Diagrama de blocos de um modelo de Hammerstein	33
3.2	Estrutura do modelo de Wiener	33
3.3	Modelos de Hammerstein-Wiener e de Lur'e	34
3.4	Esquema para identificação em malha fechada	38
3.5	Fluxograma do procedimento para a estimação de parâmetros do mo- delo de Wiener	49
4.1	Malha de controle diante da presença de atrito	52
4.2	Diagrama de blocos do modelo da malha de controle	53
4.3	Densidade espectral de potência do GBN para diferentes $\bar{T}_{sw}$	56
4.4	Fluxograma do algoritmo para geração do sinal GBN	57
4.5	Lugar geométrico dos valores candidatos ao par (S, J)	59
4.6	Condição inicial para otimização na busca do par (S, J)	63

5.1	Diagrama de blocos da malha simulada	74
5.2	Gráfico da função estática não-linear da planta simulada	74
5.3	Gráficos da simulação com perturbações de baixa intensidade	76
5.4	Dados da simulação para a estimação de parâmetros referentes ao cenário de perturbação elevada	76
5.5	Curvas de nível do comportamento do erro de simulação $\mathcal{C}_{ij}$, nos cenários de menor e maior perturbação	77
5.6	Gráfico de Nyquist do bloco dinâmico linear, em ambas as situações de perturbação consideradas	80
5.7	Módulo da resposta em frequência e limite superior do erro de modelagem provenientes do ensaio com baixo nível de perturbação	81
5.8	Limite superior do erro de modelagem no domínio da frequência na condição de perturbação mais intensa	82
5.9	Estimativa da função estática não-linear	83
5.10	Diagrama ilustrativo da plataforma híbrida (HIL)	84
5.11	Processo de neutralização de pH	86
5.12	Dados usados na estimação dos parâmetros do modelo de atrito de Karnopp	92
5.13	Comportamento da estimativa dos parâmetros de atrito de Karnopp	92
5.14	Desvio-padrão σ e coeficiente de correlação R^2 das estimativas	93
5.15	Validação cruzada dos modelos de Karnopp e de Kano	94
5.16	Relação entrada-saída da válvula real e reproduzida pelos modelos de Karnopp e de Kano	94
5.17	Comportamento do erro de simulação $\mathcal{C}_i$ em função do parâmetro S	98
5.18	Estimativas da curva estática nos diferentes cenários de perturbação	99
5.19	Dados da plataforma HIL com sinais de excitação de diferentes magnitudes	101
5.20	Estimativas do comportamento não-linear ao variar a magnitude do sinal de excitação $d(k)$	102

5.21	Dados de experimentos gerados ao variar o ganho do controlador da malha de pH	103
5.22	Ajuste da curva estática dos experimentos realizados com diferentes sintonias para o controlador AIC	105
5.23	Comparação do pH reproduzido por modelos do processo linear e não-linear	108
B.1	Módulo de $\hat{G}(e^{j\omega})$ e limite superior do erro de modelagem dos modelos identificados ao avaliar o efeito de perturbações	128
B.2	Comportamento de $ \hat{G}(e^{j\omega}) $ e de $\Delta(\omega)$ referente aos modelos estimados adotando sinais de teste com magnitudes distintas	129
B.3	Relação entre $ \hat{G}(e^{j\omega}) $ e $\Delta(\omega)$ dos modelos estimados ao variar a sintonia do controlador	130

Lista de Tabelas

4.1	Classificação do modelo do bloco linear baseada no limite superior do erro de modelagem	70
5.1	Comparativo dos algoritmos de estimação de parâmetros	78
5.2	Estimativa da função estática não-linear da planta simulada	83
5.3	Constantes de equilíbrio das reações químicas	87
5.4	Concentração molar e invariantes de reação dos reagentes	88
5.5	Valores dos parâmetros da planta de neutralização de pH	89
5.6	Especificações da válvula de controle que integra a plataforma HIL . .	90
5.7	Estimativas dos parâmetros dos modelos de Karnopp e Kano	93
5.8	Cenários utilizados para avaliar o efeito de perturbações	97
5.9	Síntese dos resultados provenientes dos diferentes cenários de perturbação	97
5.10	Resultados obtidos ao variar a magnitude do sinal de teste	100
5.11	Efeito da sintonia do controlador nos modelos estimados.	104
5.12	Comparativo entre os resultados obtidos ao representar o processo por uma estrutura linear e não-linear	106

Simbologia e abreviações

Operadores e notação

$(\cdot) \rightarrow \mathcal{G}(0, P)$: a distribuição da variável aleatória $(\cdot)$ tende a gaussiana com média nula e covariância P ;

$(\cdot)^{-1}$: inversão matricial;

$(\cdot)^T$: transposição do vetor (ou matriz) denotado(a) por $(\cdot)$;

$\dot{x}$: derivada temporal de x ;

$\hat{\theta}^{(i)}$: estimativa de θ na i -ésima iteração;

$\hat{X}$: estimativa de X ;

$\hat{y}(k|\theta, \eta)$: predição de $y(k)$ para valores fixos de θ e η , baseada em medidas até o instante $k - 1$;

$[X_{\min}; X_{\max}]$: intervalo de valores compreendidos entre $X_{\min}$ e $X_{\max}$;

$\{z(k)\}_{k=1}^N$: sequência $z(k)$ para valores de k variando de 1 até N ;

$\arg \min V(\cdot)$: argumento que minimiza a função $V(\cdot)$;

$\max(\cdot)$: valor máximo;

$\min(\cdot)$: valor mínimo;

$E(\cdot)$: operador de esperança matemática;

$P(\cdot)$: probabilidade do evento $(\cdot)$ ocorrer;

q : operador de deslocamento, $q^{-1}y(k) = y(k - 1)$;

$\|\cdot\|_2$: norma euclidiana;

$\text{cov}(X, Y)$: covariância das variáveis X e Y ;

$\text{std}(\cdot)$: desvio-padrão de uma variável;

$\mathbb{R}^N$: espaço euclidiano de dimensão N ;

$\mathbb{Z}^+$: conjunto de números inteiros positivos;

Símbolos gerais

- (S^*, J^*) : parâmetros do modelo de atrito que proporcionam o melhor ajuste do conjunto atrito + processo aos dados observados;
- (S_i, J_j) : valores candidatos ao par (S, J) indexados por i e j , respectivamente;
- α : constante que relaciona t_s e $\bar{T}_{sw}$;
- $\bar{T}_{sw}$: tempo médio entre cada comutação do sinal GBN;
- ΔS : intervalo de valores candidatos a S adotado na concepção de $\mathcal{D}_S$;
- $\Delta(\omega)$: limite superior do erro de modelagem do bloco $G(q)$ na frequência ω ;
- $\epsilon(k)$: erro de predição de um passo à frente do sinal intermediário $w(k)$;
- η : parâmetros do bloco $\mathcal{N}$ que integra a estrutura de Wiener;
- γ : constante usada para gerar o ponto de partida do algoritmo de busca por (S^*, J^*) ;
- $\hat{u}_{ij}(k)$: estimativa de $u(k)$ a partir do par (S_i, J_j) ;
- κ : ganho estacionário do bloco dinâmico linear aproximado por uma estrutura FIR;
- $\mathcal{B}(\omega)$: *bias* do modelo estimado na frequência ω ;
- $\mathcal{C}_{ij}$: erro de simulação livre do conjunto atrito + processo com o par (S_i, J_j) ;
- $\mathcal{D}$: conjunto de valores candidatos para os parâmetros (S, J) do modelo de atrito;
- $\mathcal{D}_S, \mathcal{C}_i$: análogos a $\mathcal{D}$ e $\mathcal{C}_{ij}$, para $J = 0$;
- $\mathcal{F}_2$: índice que quantifica o desempenho do modelo da associação atrito + processo em reproduzir um conjunto de observações;
- $\mathcal{L}$: bloco dinâmico linear;
- $\mathcal{N}$: bloco que representa uma função estática não-linear;
- $\mathcal{R}$: vetor de restrições empregado na estimativa de ξ ;
- $\mathcal{V}(\omega)$: variância de $\hat{G}(q)$ na frequência ω ;
- ω : frequência [rad/s];
- ω_c : frequência de corte;
- Φ_d, Φ_e, Φ_u e Φ_v : espectro dos sinais $d(k)$, $e(k)$, $u(k)$ e $v(k)$, respectivamente;
- Φ_{ue} : densidade de potência do espectro cruzado de $u(k)$ e $e(k)$;
- Ψ : matriz de regressores usada na estimativa inicial de θ ;
- ρ_v : constante usada para ajustar a relação sinal-ruído nas simulações;
- σ_e^2 : variância de $e(k)$;

- τ_{mf} : especificação da constante de tempo da malha fechada;
- τ_p : aproximação da constante de tempo dominante do processo;
- θ : parâmetros do bloco $\mathcal{L}$;
- ϑ : parâmetros da função $w(k) = f^{-1}(y(k))$;
- $\xi, \psi(k)$: vetores de parâmetros e de regressores usados na estimativa inicial de $G(q)$;
- $d(k)$: sinal de teste no instante k ;
- $e(k)$: variável aleatória branca de média nula;
- $f(w(k), \eta)$: função estática não-linear do modelo de Wiener;
- $f^{-1}(y(k), \vartheta)$: mapeamento de $y(k)$ em $w(k)$, também expresso por $g(y(k), \vartheta)$;
- $G(q), H(q)$: funções que representam o comportamento dinâmico do processo;
- $G(e^{j\omega}), H(e^{j\omega})$: resposta em frequência de $G(q)$ e $H(q)$;
- $G_0(q), H_0(q)$: valores reais das funções de transferência $G(q)$ e $H(q)$;
- K_p : ganho estático do processo em torno do ponto de operação;
- l, n, n_c : ordem do bloco $\mathcal{L}$, do modelo FIR e da média móvel usada para representar as perturbações do processo, respectivamente;
- $l_{\max}, m_{\max}$: valores máximos de l e m considerados nos cálculos de V_{oe} e V'_{oe} ;
- m : grau das *splines* cúbicas;
- N : número de observações em um conjunto de dados;
- p : multiplicador de Lagrange;
- p_{sw} : probabilidade do sinal GBN comutar o valor em um instante específico;
- $r(k)$: sinal de referência da malha de controle no instante k ;
- $R^{(i)}$: direção de busca do algoritmo de otimização na i -ésima iteração (seção 3.2.5);
- r_N : índice de correlação linear usado na avaliação do ajuste da curva estática;
- S° : valor de S que proporcionam o menor $\mathcal{C}_i$;
- T_s : período de amostragem;
- t_s : tempo de acomodação;
- $u(k)$: entrada do modelo do processo no instante k ;
- $V(\theta, \vartheta)$: critério usado para estimar os parâmetros do modelo de Wiener;
- $v(k)$: sinal de perturbações do processo no instante k ;
- $V_{oe}(m)$: critério usado para selecionar o grau m da função $f^{-1}(y(k), \vartheta)$;
- $V'_{oe}(l, n_c)$: critério adotado para a escolha das estruturas de $G(q)$ e $H(q)$;

$w(k)$: sinal intermediário entre os blocos interconectados no instante k ;
 $w_1, \dots, w_m$: extremidades dos trechos do polinômio seccionado $f(w(k), \eta)$;
 $Y, \bar{Y}$: vetor de dados de saída e a média dos valores em Y ;
 $y(k)$: saída do modelo do processo no instante k ;
 $y_1, \dots, y_m$: extremidades dos trechos do polinômio seccionado $f^{-1}(y(k), \vartheta)$;
 $z(k)$: saída do bloco que representa o atrito na válvula no instante k ;
 $\hat{w}'(k), \hat{w}^*(k)$: estimativas de $w(k)$, calculadas a partir de $\hat{g}(y(k), \vartheta)$ e do modelo FIR, respectivamente;

Símbolos relacionados aos modelos de atrito

δv : variável auxiliar que denota valores candidatos a DV ;
 λ : vetor de parâmetros do modelo de atrito de Karnopp;
 $\mathcal{S}, \mathcal{Z}$: parametrizações para criar δv ;
 $\phi(k)$: vetor de regressores no instante k ;
 $\sigma(s)$: Desvio-padrão de $\hat{\lambda}$ para um valor específico de $\delta v(s)$;
 A_d : área do diafragma;
 DV : delimita a faixa na qual $\dot{x}(t)$ é considerada nula;
 F_{atrito} : força de atrito;
 F_c : coeficiente de atrito de Coulomb;
 F_{ext} : força externa aplicada por um atuador pneumático;
 F_{mola} : força imposta pela mola;
 F_s : atrito seco (força de *breakaway*);
 F_v : coeficiente de atrito viscoso linear;
 J : parâmetro do modelo de Kano que reproduz o efeito de *slip-jump*;
 k_b : instantes em que a válvula está prestes a se mover (*breakaway*);
 k_{mola} : constante de elasticidade da mola;
 m : massa das partes móveis (haste + obturador) da válvula;
 $R^2(s)$: Coeficiente de correlação de $\hat{\lambda}$ para um valor específico de $\delta v(s)$;
 S : banda-morta da histerese no modelo de Kano;
 v_s : velocidade de Stribeck;
 $x(t)$: posição da haste no instante t ;

Abreviações e acrônimos

ARMAX: modelo auto-regressivo, de média móvel com entradas exógenas (*autoregressive moving average model with exogenous inputs*);

FIR: resposta finita ao impulso (*finite impulse response*);

FPE: critério baseado no erro de predição (*final prediction error*);

GBN: sinal binário generalizado (*generalized binary signal*);

HIL: plataforma híbrida composta por elementos reais e simulados (*hardware in the loop*);

LVDT: transdutor diferencial de variação linear, usado para medir a posição da haste da válvula (*linear variable differential transformer*);

MISO: sistema com múltiplas entradas e uma única saída (*multiple input and single output*);

MSE: erro médio quadrático (*mean square error*);

NARMAX: modelo não-linear auto-regressivo, de média móvel com entradas exógenas (*nonlinear autoregressive moving average model with exogenous inputs*);

PI: controlador com ações proporcional e integral;

op: saída do controlador;

pv: variável controlada;

dB: decibels;

Sumário

1	Introdução	1
1.1	Motivação	1
1.2	Objetivos	4
1.3	Contribuições propostas	4
1.4	Publicações	5
1.5	Estrutura do texto	6
2	Modelagem do atrito em válvulas de controle	9
2.1	Válvulas de controle de processos	10
2.2	Modelos de atrito	12
2.2.1	Modelos estáticos	12
2.2.2	Modelos dinâmicos	15
2.2.3	Modelos empíricos	16
2.3	Métodos para a estimação dos parâmetros de modelos de atrito	22
2.3.1	Identificação do modelo de Karnopp	24
2.4	Considerações finais	28
3	Identificação de modelos de processos não-lineares	31
3.1	Modelos de blocos interconectados	32
3.2	Identificação de modelos de Wiener	35
3.2.1	Enunciado do problema	37

3.2.2	Parametrização dos blocos $\mathcal{L}$ e $\mathcal{N}$	40
3.2.3	Estimativa inicial de $G(q)$ e $f^{-1}(\cdot)$	42
3.2.4	Redução do modelo	46
3.2.5	Otimização numérica	47
3.2.6	Estimação do bloco estático não-linear	48
3.3	Comentários finais	48
4	Identificação dos parâmetros do conjunto atrito + processo	51
4.1	Descrição do problema	51
4.2	Projeto do sinal de excitação	54
4.3	Estimação dos parâmetros	58
4.3.1	Procedimento baseado em busca exaustiva	58
4.3.2	Estimação dos parâmetros usando otimização	60
4.4	Técnicas de validação do modelo	64
4.4.1	Validação por simulações	64
4.4.2	Validação baseada nos erros de modelagem	65
4.5	Fechamento do capítulo	70
5	Resultados e discussões	73
5.1	Identificação de uma planta simulada	73
5.1.1	Experimento para coleta de dados	75
5.1.2	Estimação dos parâmetros	77
5.1.3	Validação dos modelos	79
5.2	Planta híbrida de neutralização de pH	84
5.2.1	Descrição da planta de neutralização de pH	85
5.2.2	Estimação dos parâmetros do modelo de Karnopp	90
5.2.3	Identificação da planta de neutralização	95

6	Conclusões e sugestões para trabalhos futuros	109
6.1	Conclusões gerais	109
6.2	Sugestões para trabalhos futuros	111
	Referências Bibliográficas	113
A	Mínimos quadrados com restrições de igualdades	123
B	Gráficos do limite superior de erro da plataforma HIL	127

Capítulo 1

Introdução

Este capítulo introdutório se inicia com uma discussão a respeito dos aspectos mais relevantes que motivaram a escolha do tema de pesquisa. Em seguida (seção 1.2), são descritos os objetivos que nortearam o desenvolvimento deste trabalho. As contribuições propostas e alguns dos resultados já publicados são expostos, respectivamente, nas seções 1.3 e 1.4. Por fim, a estrutura de capítulos do documento é apresentada na seção 1.5.

1.1 Motivação

Em qualquer processo produtivo, eficiência e confiabilidade são sempre características desejáveis, no sentido de proporcionar produtos de boa qualidade pelo menor custo possível, através de procedimentos que levem em conta aspectos como a segurança das pessoas e impactos ao meio ambiente. Uma típica planta industrial possui centenas ou até milhares de malhas de controle [21], e o desempenho destas malhas está diretamente relacionado aos custos de produção e à qualidade do produto final.

Diversos trabalhos apontam para o problema de desempenho em malhas de controle de processos industriais [12, 67]. Em um estudo realizado por Desborough e Miller [25], baseado nos dados de 26 mil malhas de controle, instaladas em indústrias de diversos segmentos, concluiu-se que 68% destas malhas oferecem oportunidades

de melhorias significativas. Portanto, surge a seguinte questão: o que pode ser feito para reverter este quadro?

De acordo com Jelali [44], os fatores que comprometem o desempenho de malhas de controle são: erros de projeto, mau-funcionamento de sensores e atuadores, estrutura ou sintonia de controladores inadequada e falta de manutenção.

Os atuadores mais encontrados nas malhas regulatórias das plantas industriais são as válvulas de controle de acionamento pneumático, que apresentam características não-lineares (histerese, zona-morta, entre outras) provocadas pelo atrito entre as partes móveis da válvula. Este atrito pode causar efeitos indesejáveis como erros estacionários e comportamento oscilatório. Muitos pesquisadores consideram o atrito em válvulas de controle um dos fatores mais relevantes para problemas de desempenho em malhas de controle [21, 44, 50, 67]. Em outro estudo, realizado em indústrias canadenses [12], foi mostrado que 30% das malhas oscilavam devido a problemas relacionados às válvulas de controle.

A rigor, o funcionamento de qualquer válvula é afetado por atrito, por menor que seja tal influência. Quando o atrito é muito elevado, um procedimento imediato de manutenção pode ser necessário. Por outro lado, em situações menos severas, é possível usar compensadores específicos (baseados em modelos) para diminuir a perda de desempenho da malha, até que uma parada seja programada. Portanto, algoritmos para quantificação de atrito podem gerar uma classificação de prioridade de manutenção, de acordo com a intensidade da não-linearidade da válvula [22]. Nos casos em que o procedimento de manutenção não for viável (ou estritamente necessário), os mesmos modelos usados para quantificar o atrito podem ser usados na sua compensação. Por este fato, segundo Desborough e Miller [25], um método confiável e não-intrusivo, capaz de quantificar o atrito em válvulas de controle operando em malha fechada é altamente necessário na indústria de processos.

O problema de desempenho devido à sintonia e/ou estrutura inadequada de controladores, em geral, ocorre porque os controladores são sintonizados de forma

excessivamente conservadora para assegurar a estabilidade da malha de controle [44], pois não se dispõe de informações do processo, ou então, a sintonia é baseada em modelos grosseiros. Portanto, para que os controladores possam ser melhor sintonizados, são necessários modelos capazes de reproduzir com fidelidade a dinâmica dos processos envolvidos.

Srinivasan *et al.* [73] propuseram um método de otimização, baseado na busca exaustiva de parâmetros, capaz de identificar um modelo linear para o processo e de quantificar o atrito proveniente da válvula de controle. À luz do que foi dito anteriormente, este método contempla duas das principais fontes de variabilidade, pois a quantificação do atrito pode ser usada na solução de problemas provocados por não-linearidades da válvula, ao passo que o modelo do processo pode ser aplicado na re-sintonia do controlador. Entretanto, esta metodologia emprega uma estrutura para modelar o atrito que não é capaz de reproduzir adequadamente características não-lineares importantes de uma válvula de processo [23, 31].

Controladores lineares são sintonizados para trabalhar em torno de uma faixa limitada de operação, e eventuais mudanças nestas condições acabam comprometendo o desempenho do sistema. Em se tratando de processos que apresentem não-linearidades mais severas, esta influência é ainda mais significativa. Para processos cujo comportamento mude significativamente de acordo com o ponto de operação, uma alternativa é recorrer a controladores não-lineares, desde que um modelo que incorpore tais não-linearidades possa ser obtido.

Atualmente, há técnicas bem estabelecidas para a obtenção de modelos lineares [58]. Em contrapartida, a identificação de modelos não-lineares vem sendo assunto de intensa pesquisa [60]. Apesar disto, são raros os trabalhos que tratem o caso específico em que os modelos sejam estimados a partir de dados de operação em malha fechada. Todavia, algoritmos de identificação em malha fechada são de grande interesse prático, pois entre outras vantagens, oferecem maior segurança para a realização dos experimentos de coleta de dados [30, 89] e, muitas vezes, proporcionam modelos mais apropriados para o projeto de controladores [42].

1.2 Objetivos

Em linhas gerais, este trabalho visa estender a metodologia originalmente concebida por Srinivasan *et al.* [73], de modo que uma estrutura mais propícia seja utilizada para modelar o atrito, e que um modelo não-linear seja empregado na representação da dinâmica do processo. Além disso, deseja-se que o método de estimação de parâmetros seja aplicável a dados obtidos com o processo operando em malha fechada, de modo que o projeto do experimento para a coleta de dados também deve considerar esta questão.

Tais metas podem ser desmembradas nos seguintes objetivos específicos:

- Seleção de uma estrutura conveniente para modelar certas não-linearidades provenientes do atrito em válvulas de controle;
- Escolha de uma estrutura de modelo capaz de incorporar boa parte de eventuais comportamentos não-lineares referentes ao processo;
- Desenvolvimento de um algoritmo para a estimação de parâmetros de modelos não-lineares, baseado em dados obtidos com o processo operando em malha fechada;
- Estudo de ferramentas para a validação dos modelos obtidos.

1.3 Contribuições propostas

A primeira contribuição originada deste trabalho é uma metodologia para a quantificação do atrito em válvulas de controle usando o modelo de Karnopp [51]. Este método supõe que a posição da haste da válvula seja conhecida. Apesar desta suposição não ser comumente encontrada em casos práticos, os resultados obtidos ao aplicar o método em uma válvula real [70] mostraram que uma estrutura mais simples poderia ser utilizada para representar o fenômeno de atrito em válvulas de controle.

Outra contribuição a ser mencionada é o desenvolvimento de um algoritmo para estimar os parâmetros de modelos de Wiener (construídos a partir da associação de um bloco dinâmico linear seguido por uma função estática não-linear). Além de se basear em um critério considerado mais apropriado para o caso em que os dados sejam adquiridos em malha fechada [38, 91], com o intuito de facilitar o uso do algoritmo, são incorporadas funcionalidades que tornam a seleção da estrutura do modelo praticamente automática.

Por fim, o procedimento de identificação do conjunto atrito + processo, em que um modelo de Wiener é usado para incorporar a dinâmica do processo e o atrito na válvula de controle é quantificado através de um modelo constituído por dois parâmetros, é considerado como uma terceira contribuição deste trabalho.

1.4 Publicações

Até este momento foram publicados quatro artigos em congressos (três internacionais e um nacional). O primeiro deles [69], cujo título é “*Comparision between two friction model parameter estimation methods applied to control valves*” apresenta o algoritmo de estimação de parâmetros do modelo de atrito de Karnopp (subseção 2.3.1) e foi publicado nos anais do 8th *IFAC Symposium on Dynamics and Control of Process Systems, DYCOPS* ocorrido em Cancún, México, no ano de 2007.

A aplicação deste algoritmo a uma válvula real gerou o trabalho intitulado “*Karnopp friction model identification for a real control valve*” [70], apresentado em 2008, no 17th *IFAC World Congress*, realizado em Seul (Coreia do Sul). Os resultados obtidos nesse trabalho motivaram a elaboração do artigo “Comparação e equivalência dos modelos de atrito de Kano e Karnopp aplicados a válvulas de controle” [79], publicado nos anais do *XVII Congresso Brasileiro de Automatica*, realizado em 2008, na cidade de Juiz de Fora - MG. Os artigos citados anteriormente referem-se, basicamente, à modelagem e à quantificação do atrito em válvulas de controle.

Os resultados da aplicação do método de identificação do conjunto atrito + processo em dados simulados foram redigidos na forma de um artigo “*Valve friction and nonlinear process model closed-loop identification*” [71], publicado no 7th *International Symposium on Advanced Control of Chemical Processes, ADCHEM* ocorrido em julho deste ano, em Istambul (Turquia).

A aplicação do algoritmo de estimação de parâmetros do modelo de Wiener a dados gerados em malha fechada por uma planta simulada de neutralização de pH, deve gerar um artigo visando o *XVIII Congresso Brasileiro de Automatica CBA*, a ser realizado em 2010, em Bonito - MS.

Também pretende-se publicar os resultados obtidos ao identificar o conjunto atrito + processo em uma plataforma híbrida no 9th *IFAC Symposium on Dynamics and Control of Process Systems*, a ser realizado em julho de 2010, em Leuven (Bélgica). Esse trabalho também está sendo redigido na forma de um artigo a ser enviado a um periódico da área de controle de processos.

1.5 Estrutura do texto

O capítulo 2 trata da modelagem do atrito em válvulas de controle. As principais estruturas de modelos de atrito, assim como suas respectivas vantagens e desvantagens são discutidas. Um método para a estimação dos parâmetros do modelo de atrito de Karnopp, aplicado a uma válvula de controle pneumático é apresentado.

A identificação “caixa-preta” de modelos de Wiener é abordada no capítulo 3. Inicialmente são apresentados os principais modelos não-lineares construídos a partir da associação de blocos dinâmicos lineares e estáticos não-lineares. Em seguida, apresenta-se uma revisão bibliográfica a respeito das abordagens existentes para a estimação dos modelos de Wiener. Enfim, é proposto um algoritmo para determinar a estrutura e estimar os parâmetros deste modelo.

A quantificação do atrito e a identificação do modelo do processo, com dados

de malha fechada, é tratada no capítulo 4. Primeiramente discute-se o projeto de um sinal de excitação. Em seguida, são apresentados dois algoritmos para estimar os parâmetros do conjunto atrito + processo. Para auxiliar na validação dos modelos estimados, descreve-se um método para inferir limites para o erro de modelagem no domínio da frequência.

No capítulo 5 os métodos de identificação propostos são aplicados a uma planta simulada não-linear sujeita a atrito. Também estuda-se a identificação de uma plataforma híbrida, composta por uma válvula real integrada a uma planta simulada de neutralização de pH. Os resultados são avaliados em relação à robustez diante de perturbações, à influência da magnitude do sinal de excitação e à sintonia do controlador.

As principais conclusões do trabalho são resumidas no início do capítulo 6. Posteriormente, são citadas algumas propostas de continuidade ao tema explorado neste trabalho.

O documento inclui ainda dois apêndices. No primeiro é derivada a expressão usada no cálculo da estimativa inicial do modelo de Wiener. No apêndice B são apresentados os gráficos de limite superior do erro de modelagem empregados na validação dos modelos durante a identificação da plataforma híbrida.

Capítulo 2

Modelagem do atrito em válvulas de controle

O fenômeno de atrito desperta o interesse de pesquisadores e engenheiros desde o século XVI, quando o primeiro modelo matemático para representá-lo foi proposto por Leonardo da Vinci. Apesar de ser objeto de estudo há muito tempo, esse fenômeno ainda não é completamente conhecido pela comunidade científica, tanto que novas estruturas de modelos ainda têm sido propostas nos últimos anos [27, 55, 75].

O atrito pode ser descrito como uma força tangencial que surge em oposição à movimentação, ou mesmo tendência de movimento, entre dois corpos cujas superfícies estejam em contato [4]. Quando presente em dispositivos atuadores, o atrito pode ocasionar erros estacionários, ciclos limite e o indesejável fenômeno de “escorregamento”, também conhecido por *stick-slip*, que é provocado pelo fato da força de atrito ser maior durante uma situação de repouso do que em situações de movimento [16].

Em estratégias que visem compensar os efeitos do atrito, sem recorrer a malhas de controle com ganho elevado [26], um modelo adequado é imprescindível para prever e compensar os efeitos provocados pelo atrito. Além disso, modelos de atrito podem ser usados na análise de estabilidade, na previsão de ciclos limite de malhas de controle e no desenvolvimento de simuladores de processos.

Justificada a importância dos modelos de atrito, antes de iniciar uma revisão

acerca das estruturas de modelo disponíveis, faz-se necessária uma breve descrição das válvulas de controle, cujo comportamento não-linear causado pela presença de atrito, é uma das principais fontes de variabilidade nas malhas de controle [21].

2.1 Válvulas de controle de processos

Basicamente, as válvulas de controle são dispositivos capazes de variar a restrição à passagem de um fluido em resposta a um sinal de comando. As válvulas de controle são compostas por dois conjuntos principais: corpo e atuador. O primeiro, em destaque no canto inferior direito da Figura 2.1, abriga a sede da válvula. A vazão do fluido é alterada de acordo com a posição de um obturador em relação à sede da válvula.

Figura 2.1: Corpo e atuador de uma válvula de controle com atuador pneumático (adaptado de manuais técnicos [28, 29]).

O conjunto atuador é que proporciona a força motriz para movimentar o obturador. No caso dos atuadores pneumáticos, os mais usuais por questões de custo e simplicidade, o sinal de controle é convertido em pressão de ar e age sobre um diafragma em oposição à força elástica de uma mola (Figura 2.1). Portanto, como o diafragma está ligado ao obturador através de uma haste, o desequilíbrio entre as forças da mola e a produzida pelo ar resulta em movimentos do obturador.

Para evitar vazamento de fluido entre a haste e o corpo da válvula, são utilizados anéis circulares de vedação, também chamados de gaxetas*, que por exercerem contato com as partes móveis (haste e obturador), provocam atrito entre as superfícies.

Ao considerar o atrito provocado pelo contato das gaxetas de vedação com a haste da válvula, o balanço de forças baseado na segunda lei de Newton é dado por:

$$m \cdot \ddot{x}(t) = \sum \text{Forças}$$

$$m \frac{d^2}{dt^2} x(t) = F_{ext}(t) + F_{mola}(t) - F_{atrito}(t), \quad (2.1)$$

em que m é a massa das partes móveis; $x(t)$ é a posição da haste; $F_{ext}(t) = A_d \cdot P_{ext}(t)$ é a força externa aplicada, $P_{ext}(t)$ é a pressão de ar que atua sobre o diafragma e A_d é a área do diafragma; $F_{mola}(t) = -k_{mola} \cdot x(t)$ é a força da mola e k_{mola} é a constante de elasticidade da mola; $F_{atrito}(t)$ é a força de atrito.

Uma vasta quantidade de estruturas, cujo propósito é modelar dispositivos sujeitos a atrito, pode ser encontrada na literatura. A próxima seção apresenta uma discussão a respeito das estruturas disponíveis para tratar o problema da modelagem das não-linearidades presentes em válvulas de controle devidas ao atrito.

*Em geral, utilizam-se gaxetas de Teflon®, que apresentam menos atrito quando comparadas às de grafite; entretanto, as primeiras não são adequadas em aplicações em que fluidos de temperatura elevada escoem pela válvula.

2.2 Modelos de atrito

Há alguns anos, um grupo de pesquisadores [6] publicou uma revisão completa sobre a modelagem do atrito num contexto geral. Recentemente, Garcia [31] avaliou o desempenho de diversas estruturas empregadas na simulação de válvulas de controle com diferentes níveis de atrito.

De acordo com Altpeter [4], os modelos de atrito podem ser agrupados em duas classes: estáticos e dinâmicos. Recentemente, modelos voltados a válvulas de controle, pertencentes a uma terceira classe, conhecida por empíricos ou *data-driven* foram propostos [21, 50, 74]. Nos primeiros, a força de atrito é modelada explicitamente, ao passo que a classe de modelos empíricos, faz uso de relações matemáticas para representar comportamentos específicos causados pelo atrito, num dado dispositivo. Algumas das características, assim como as principais estruturas propostas de cada classe, são apresentadas a seguir.

2.2.1 Modelos estáticos

Nos modelos estáticos, como o próprio nome diz, a força de atrito é dada por um mapeamento estático da velocidade da haste e da força motriz ($F_{ext} + F_{mola}$). A forma mais usual de realizar este mapeamento é através da superposição de três componentes:

- Atrito de Coulomb, F_c , que representa uma força constante que atua em oposição à direção do movimento, e independe da magnitude da velocidade. Acredita-se que este seja o principal componente de atrito em válvulas de controle [48].
- Atrito viscoso linear, $F_v \cdot \dot{x}(t)$, onde $\dot{x}(t)$ é a velocidade (da haste, no contexto de válvulas). Este componente surge quando há deslocamento entre duas superfícies lubrificadas, e a sua magnitude aumenta linearmente com a velocidade, de acordo com o coeficiente F_v .

- Atrito estático, F_s , também conhecido como atrito seco ou *stiction*, (junção das palavras *static* e *friction*), quantifica o atrito na iminência de movimento. Em outras palavras, a força necessária para iniciar o movimento deve superar o atrito estático.

Uma estrutura de modelo distinta pode ser obtida de acordo com a forma como os componentes descritos anteriormente forem combinados. A forma geral de descrever a força de atrito numa válvula de controle, conhecida como modelo clássico [6], é dada por:

$$F_{atrito}(F_{ext}(t), \dot{x}(t), x(t)) = \begin{cases} F_{ext}(t) - k_{mola} \cdot x(t) & , \text{ se } \dot{x}(t) = 0; \\ F_{din}(t) & , \text{ caso contrário.} \end{cases} \quad (2.2)$$

A primeira linha em (2.2) representa a força de atrito em situações de repouso, enquanto a segunda caracteriza o regime dinâmico, no qual $F_{din}(t)$ é dado por:

$$F_{din}(t) = \text{sgn}(\dot{x}(t)) \cdot F_c + F_v \cdot \dot{x}(t) + \text{sgn}(\dot{x}(t)) \cdot (F_s - F_c) \cdot e^{-\left(\frac{\dot{x}(t)}{v_s}\right)^2}, \quad (2.3)$$

em que a função $\text{sgn}(\dot{x}(t))$ é definida como:

$$\text{sgn}(\dot{x}(t)) \triangleq \begin{cases} -1 & , \text{ se } \dot{x}(t) < 0; \\ 0 & , \text{ se } \dot{x}(t) = 0; \\ 1 & , \text{ se } \dot{x}(t) > 0. \end{cases}$$

e v_s , denominado velocidade de Stribeck, é um fator usado para ajustar a atenuação do termo relacionado a F_s em (2.3). Este termo visa quantificar o comportamento de uma parcela da força de atrito, denominada atrito de Stribeck[†], que decresce monotonicamente quando ocorre movimento. Para capturar este comportamento, valores empíricos são atribuídos a v_s .

[†]O atrito de Stribeck foi verificado experimentalmente no início do século XX, pelo cientista alemão Richard Stribeck (1861-1950), enquanto investigava a relação da força de atrito com a velocidade do movimento [4].

O fenômeno de atenuação do atrito, quando um corpo passa do regime estático para o dinâmico, é chamado *stick-slip*. No caso das válvulas de controle, como consequência deste fenômeno, toda vez que a haste iniciar um movimento, ou seja:

$$|F_{ext}(t) - k_{mola} \cdot x(t)| > F_s$$

a movimentação ocorrerá na forma de um salto, denominado *slip-jump* [21].

Vale ressaltar que, ao utilizar o modelo (2.2) em simuladores, a comutação entre os regimes estático e dinâmico depende da condição $\dot{x}(t) = 0$. No entanto, detectar esta condição não é trivial, pois $\dot{x}(t)$ é calculado numericamente a partir de valores da posição. Para resolver este problema, Karnopp [51] propôs a criação de uma faixa de valores, dentro da qual a velocidade do movimento seja considerada nula. Com este artifício, o modelo (2.2) é reescrito como:

$$F_{atrito}(t) = \begin{cases} F_c \cdot \text{sgn}(\dot{x}(t)) + F_v \cdot \dot{x}(t) & , \text{ se } |\dot{x}(t)| \geq DV \\ F_{ext}(t) - k_{mola} \cdot x(t) & \begin{cases} , \text{ se } |\dot{x}(t)| < DV \text{ e} \\ |F_{ext}(t) - k_{mola} \cdot x(t)| \leq F_s \end{cases} \\ F_s \cdot \text{sgn}(F_{ext}(t) - k_{mola} \cdot x(t)) & \begin{cases} , \text{ se } |\dot{x}(t)| < DV \text{ e} \\ |F_{ext}(t) - k_{mola} \cdot x(t)| > F_s, \end{cases} \end{cases} \quad (2.4)$$

em que DV delimita a faixa na qual a velocidade é considerada nula. Além dos regimes dinâmico e estático, descritos respectivamente na primeira e segunda linhas de (2.4), o modelo de Karnopp [51] também distingue a situação de iminência de movimento, que ocorre quando $|F_{ext}(t) - k_{mola} \cdot x(t)| > F_s$ enquanto $|\dot{x}(t)| < DV$.

Para ilustrar a diferença entre o comportamento da força de atrito em função da velocidade, reproduzido pelos modelos clássico e de Karnopp, as expressões (2.2) e (2.4) foram simuladas utilizando os seguintes parâmetros:

$$F_c = 600\text{N}, \quad F_v = 2200 \frac{\text{N}\cdot\text{s}}{\text{m}}, \quad F_s = 720\text{N}, \quad v_s = 10^{-3} \frac{\text{m}}{\text{s}} \quad \text{e} \quad DV = 0,6v_s.$$

Os resultados da simulação são mostrados na Figura 2.2. Note que, para o modelo de Karnopp, quanto maior a diferença entre F_s e F_c , maior será a descontinuidade nos extremos do regime estático, onde $|\dot{x}(t)| = DV$.

Figura 2.2: Comportamento da força de atrito em função da velocidade do movimento, reproduzido pelos modelos clássico e de Karnopp.

2.2.2 Modelos dinâmicos

Modelos dinâmicos, que recebem esta nomenclatura por possuírem parâmetros variantes no tempo em suas estruturas, foram criados para permitir que comportamentos específicos, não capturados pelos apresentados na seção 2.2.1, pudessem ser representados. Por exemplo, os modelos estáticos quantificam a força de atrito, admitindo apenas movimento uniforme. Todavia, certos efeitos do atrito são verificados somente quando a velocidade não é constante [4, 6]. Através de ensaios em que a velocidade varia significativamente, verifica-se que o mapeamento $F_{ext} = f(\dot{x}(t))$ apresenta comportamento não-linear na forma de histerese [16].

Sabe-se ainda que, a rigor, a força de atrito apresenta um comportamento elástico quando a força aplicada é menor do que a necessária para sair de uma situação de repouso [6]. Logo, se uma força menor do que a necessária para vencer o atrito é aplicada, ocorre um ínfimo deslocamento, e se esta força voltar à condição inicial, o mesmo vale para a posição. Este regime, denominado pré-deslocamento (do Inglês, *pre-sliding*) [75], não consegue ser reproduzido pelos modelos estáticos.

Além disso, a partir de dados experimentais, verificou-se que a força necessária

para iniciar um movimento, pode ser função da taxa de variação da força motriz [46], ao invés de constantes como em (2.3). Dentre as estruturas de modelos dinâmicos capazes de incorporar cada um dos comportamentos descritos anteriormente, podem-se citar: Lund-Grenoble (LuGre) [16], Leuven [75], elastoplástico [27], Maxwell Slip [55], entre outros.

Apesar de ampliar a gama de fenômenos representáveis, a classe de modelos dinâmicos necessita de sofisticados métodos para a estimação de seus parâmetros, pois possuem variáveis de estado que não são mensuráveis [66], e são descritos por estruturas mais complexas, se comparadas aos modelos estáticos. Por outro lado, os modelos baseados em mapeamentos estáticos, especialmente o modelo de Karnopp, têm apresentado resultados satisfatórios, quando utilizado para modelar [70] ou para compensar [48] o fenômeno de atrito em válvulas de controle.

Uma provável justificativa para tais resultados pode ser o compromisso entre simplicidade estrutural e capacidade de representar fenômenos proporcionados pelo modelo de Karnopp. Ademais, num trabalho recente [31], mostrou-se através de simulações que os modelos de Karnopp e de LuGre reproduzem de forma equivalente os comportamentos de histerese e *slip-jump* em uma válvula de controle.

Tais fatos levam a crer que a estrutura do modelo de Karnopp é suficiente para representar o atrito em válvulas de controle. Além disso, a dificuldade de se estimar os parâmetros de modelos dinâmicos a partir de dados de operação em malha fechada, dificulta a aplicação desta classe de modelos em muitas situações práticas [66]. Todavia, a título de sugestão, são indicadas algumas referências onde os modelos dinâmicos são tratados com maior profundidade [4, 6, 16, 27, 35, 55].

2.2.3 Modelos empíricos

Conforme dito anteriormente, ao invés de quantificar diretamente a força de atrito, como é feito pelos modelos vistos até este momento, a classe de modelos

empíricos visa representar a relação[‡] entrada-saída de um determinado dispositivo. No caso das válvulas de controle, o sinal de entrada é a força aplicada $F_{ext}(k)$ para movimentar a haste e o sinal de saída é a sua posição $x(k)$. Tais modelos são uma forma alternativa de representar comportamentos não-lineares de válvulas sujeitas a atrito.

A grande motivação para utilizar os modelos empíricos é que estes possuem poucos parâmetros, que estão diretamente relacionados com os fenômenos estudados e, em geral, são facilmente identificáveis a partir de ensaios específicos [21, 50], desde que os sinais de entrada e saída possam ser medidos. Outra vantagem dos modelos empíricos em relação aos apresentados anteriormente, é que como são descritos por estruturas mais simples, exigem menos esforço computacional durante simulações.

Dentre os comportamentos não-lineares encontrados em válvulas de controle provocados pelo atrito, o mais comum é o de histerese [48]. Como as forças de atrito se manifestam em direção oposta à do movimento, toda vez que uma mudança no sentido for necessária, a força motriz deve exceder não apenas o atrito estático no novo sentido, mas também a força de atrito presente anteriormente. Analogamente, é o mesmo comportamento não-linear provocado por folga nas engrenagens de sistemas mecânicos.

Um modelo empírico para reproduzir a histerese em uma válvula de controle pode ser expresso através de:

$$x(k) = \begin{cases} F_{ext}(k) - \mathcal{H} & \begin{cases} , \text{ se } F_{ext}(k) > F_{ext}(k-1) \\ \text{ e } F_{ext}(k) > x(k-1) + \mathcal{H} \end{cases} \\ F_{ext}(k) + \mathcal{H} & \begin{cases} , \text{ se } F_{ext}(k) < F_{ext}(k-1) \\ \text{ e } F_{ext}(k) < x(k-1) - \mathcal{H} \end{cases} \\ x(k-1) & , \text{ caso contrário,} \end{cases} \quad (2.5)$$

em que $F_{ext}(k-1)$ e $x(k-1)$ são, respectivamente, a força externa e a posição da haste no instante anterior a k e $2\mathcal{H}$ denota a banda morta da histerese. Da primeira

[‡]Normalmente os modelos empíricos são baseados em sinais amostrados [21, 50, 74], portanto a dependência de t (tempo contínuo) é substituída pela de k (tempo discreto).

linha de (2.5), note que para a haste se movimentar no sentido ascendente, ou seja $x(k) > x(k-1)$, duas condições devem ser satisfeitas: (1) a força no instante atual deve ser maior do que no anterior e (2) a força externa deve ultrapassar a histerese provocada pelo atrito. O movimento descendente é expresso de forma análoga na segunda linha de (2.5). Por outro lado, se ambas as condições necessárias para haver movimentação da haste não forem satisfeitas, então $x(k) = x(k-1)$.

A Figura 2.3 mostra a resposta da posição $x(k)$ para uma excitação senoidal, cuja amplitude varia entre 20% e 80%, quando a válvula está sujeita à histerese reproduzida por (2.5) com $\mathcal{H} = 10\%$. Os segmentos $\overline{AB}$ e $\overline{CD}$, de comprimento $2\mathcal{H}$, surgem devido às reversões no sentido do movimento.

Figura 2.3: Entrada e saída reproduzida pelo modelo de histerese (2.5), parametrizada por $\mathcal{H} = 10\%$ (esquerda) e a correspondente relação $x(k) = f(F_{ext}(k))$ (direita).

Outra não-linearidade presente em válvulas de controle, principalmente nas que utilizam gaxetas de grafite, é o *slip-jump*. Com o intuito de modelar este comportamento, provocado pelo atrito estático, Stenman *et al.* [74] propuseram a seguinte estrutura:

$$x(k) = \begin{cases} x(k-1) & , \text{ se } |F_{ext}(k) - x(k-1)| \leq d_{sj}; \\ F_{ext}(k) & , \text{ caso contrário,} \end{cases} \quad (2.6)$$

em que d_{sj} representa o *slip-jump*. Este modelo foi empregado com êxito em técnicas

de detecção do atrito estático [73, 74]. Entretanto, para fins de predição da posição da haste em válvulas de controle, esta estrutura pode proporcionar resultados pobres, especialmente em situações onde a influência da histerese é significativa.

Um modelo composto por dois parâmetros (S e J), capaz de reproduzir tanto o comportamento não-linear da histerese, quanto o *slip-jump* provocado pelo atrito estático, foi proposto recentemente [20], onde o parâmetro S expressa a banda-morta da histerese e J quantifica o *slip-jump*, toda vez que a haste sai de uma situação de repouso. Todavia, verificou-se através de simulações, que esta proposta inicial funciona bem apenas para uma classe restrita de sinais [31]. Kano *et al.* [50] reformularam o modelo supracitado, de modo a remover a restrição quanto ao sinal de entrada. A relação entre a força aplicada e a posição da haste reproduzida por este modelo pode ser explicada através da Figura 2.4.

Figura 2.4: Entrada e saída reproduzida pelo modelo empírico [50], onde $S = 15\%$ e $J = 4\%$ (esquerda), e a correspondente relação $x(k) = f(F_{ext}(k))$ (direita).

Inicialmente, a força aplicada é 50% da força máxima e a haste está em repouso na posição $x(0) \approx 0,45$, ilustrada pelo ponto C na Figura 2.4. Decorridos 15s de simulação, $F_{ext}(k)$ é incrementado linearmente, porém a haste só se movimenta quando a força aplicada é superior ao atrito estático, no ponto D . A partir deste ponto o atrito estático desaparece, provocando um desequilíbrio no balanço de forças

que rege a posição da haste $x(k)$ e então ocorre o *slip-jump* até o ponto E . Portanto, o comprimento do segmento $\overline{DE}$ corresponde a J . Este mesmo comportamento se repete quando $F_{ext}(k)$ atinge 0,75 e volta ao regime estático por 10s.

O efeito do parâmetro S só é notado quando há mudança de sentido. Note que quando $F_{ext}(k) = 0,9$ (indicado pelo ponto F na Figura 2.4), a força aplicada começa a ser reduzida e a haste se mantém parada, até que em G , onde $F_{ext}(k) = 0,75$, a diferença de forças entre os pontos F e G , $\Delta F_{ext(FG)}$, ultrapassa o valor da histerese, e a haste salta para H . Logo: $\overline{FG} = S$. Ainda há uma nova parada (I) e outro *slip-jump* (pontos J , K), antes da próxima mudança de sentido, quando $\Delta F_{ext(LA)} > S$, resultando no “escorregamento” da posição de A para B .

O comportamento não-linear proveniente da superposição da histerese e do *slip-jump* no modelo empírico de dois parâmetros (S e J), que a partir deste ponto será referenciado como modelo de Kano, é descrito através de um fluxograma ilustrado na Figura 2.5.

Antes de mais nada, é necessário transformar a força externa aplicada, $F_{ext}(k)$, de unidade de engenharia para porcentagem, pois o modelo trabalha apenas com valores normalizados. Além dos parâmetros S e J , o modelo possui três variáveis internas: stp que assume os valores 0 ou 1 para indicar as situações de movimento ou repouso, respectivamente; u_s que armazena o valor de $F_{ext}(k)$ no exato instante em que a haste sai de uma situação de movimento para repouso e $d_{aux} = \pm 1$ que é usado para denotar o sentido da força de atrito.

O primeiro bloco de decisão é utilizado para impedir que valores espúrios de $F_{ext}(k)$ sejam processados, impondo 0 e 1 como limites inferior e superior de saturação da entrada. Depois disto, é calculada a variação da força externa $\Delta F_{ext}(k)$, em relação ao instante anterior. E então, se $F_{ext}(k)$ mudar de sentido a partir de uma situação de movimento, u_s é atualizado com o valor da entrada no instante anterior e stp passa a indicar situação de repouso, para evitar que haja uma mudança de sentido sem passar pelo regime estático.

Na sequência, um bloco faz o teste da variável stp e se a haste estiver em movimento, o valor de $x(k)$ é atualizado através da expressão:

$$x(k) = F_{ext}(k) - \frac{d_{aux}(S - J)}{2}. \quad (2.7)$$

Caso contrário, ou seja, se for indicada uma situação de repouso ($stp = 1$), há três possibilidades:

1. Houve mudança no sentido de $F_{ext}(k)$, superior à banda-morta de histerese, fazendo com que a haste se mova em oposição ao último deslocamento, ou seja, o sinal da variável d_{aux} é invertido antes de atualizar $x(k)$, de acordo com a expressão (2.7);

Figura 2.5: Fluxograma do modelo empírico de Kano com parâmetros S e J [50].

2. A partir do repouso, a força externa aplicada é suficiente para vencer o atrito estático e provocar o deslocamento no mesmo sentido da última movimentação. Neste caso, o novo valor da posição é calculado a partir de (2.7);
3. A haste permanece imóvel, logo $x(k) = x(k - 1)$.

Por fim, sempre que a posição da haste for atualizada, a variável *stp* passa a indicar movimento.

O modelo S e J considera, implicitamente, que a força de atrito independe da velocidade do movimento. Apesar de esta hipótese não ser verdadeira, conforme verificado em testes práticos [4], no caso das válvulas de controle de acionamento pneumático, considerar a força de atrito constante no regime dinâmico pode ser razoável, pois aparentemente, o atrito viscoso é desprezível se comparado à magnitude de outras forças presentes no balanço (2.1), como a força da mola (F_{mola}) e o atrito de Coulomb (F_c) [70].

2.3 Métodos para a estimação dos parâmetros de modelos de atrito

Métodos clássicos de estimação de parâmetros, baseados em regressão linear, não podem ser diretamente aplicados à identificação de modelos de atrito, pois as estruturas apresentadas anteriormente são não-lineares nos parâmetros. Portanto, são necessários algoritmos específicos para estimar os parâmetros dos modelos de atrito.

Altpeter [4], por exemplo, dividiu em etapas o procedimento de identificação do modelo de LuGre. A cada etapa, o sistema em questão é submetido a um sinal de entrada que excite comportamentos específicos, de modo que o(s) parâmetro(s) correspondentes sejam estimados. Entretanto, por requerer excitações específicas em cada etapa, em situações onde apenas dados de operação em malha fechada estejam disponíveis, este método não é aplicável.

Rizos e Fassois [68] propuseram uma formulação, a partir do modelo Maxwell Slip [55], através da qual a força aplicada a um dispositivo sujeito a atrito é representada por uma regressão não-linear da posição. Com isso, se força aplicada e posição forem medidos, os parâmetros do modelo de atrito são estimados usando um algoritmo de duas fases. Na primeira, uma otimização baseada em algoritmos genéticos é usada para buscar, no espaço de parâmetros, regiões de mínimos globais ou locais. Na segunda, utiliza-se o algoritmo Simplex de Nelder-Mead [84] para localizar o melhor modelo, dentre os obtidos no passo anterior. Além de se tratar de um método complexo, a variância do modelo estimado devido ao elevado número de parâmetros envolvidos pode ser uma desvantagem a se considerar.

Cheok *et al.* [18] sugeriram alterações no algoritmo Simplex de otimização não-linear, para acelerar a taxa de convergência e para aumentar as chances de encontrar o mínimo global. Este algoritmo foi usado na identificação dos parâmetros de um servomotor, onde o atrito foi parametrizado usando o modelo de Karnopp [51].

Os parâmetros de modelos de atrito também podem ser estimados aplicando o método do balanço harmônico com o auxílio de funções descritivas. Na proposta de Besançon-Voda e Blaha [11], o sistema deve operar em malha fechada e um bloco não-linear do tipo relê + histerese é adicionado ao ramo de realimentação, para garantir a oscilação da malha. Dois pares distintos de amplitude do relê e histerese são ajustados para que a malha apresente ciclos-limite de mesma amplitude e frequência. Os valores de tais pares são então usados, numa expressão de balanço harmônico, para calcular diretamente o coeficiente de atrito de Coulomb inerente ao sistema. Em um trabalho recente, Kim e Chung [52] estenderam essa metodologia, de modo a estimar os parâmetros do modelo de Karnopp. Apesar de proporcionar resultados satisfatórios, esses métodos requerem a adição de elementos na malha de controle, o que é indesejável ou até inviável em plantas industriais.

Uehara [78] avaliou três técnicas de quantificação de atrito que aproximam a não-linearidade na válvula por uma histerese (subseção 2.2.3). Na primeira, proposta por Kano *et al.* [50], a estimativa do nível de atrito é baseada na inspeção

visual da posição da haste e da saída do controlador (op). Na segunda técnica, idealizada por Choudhury *et al.* [22], o atrito é caracterizado ao ajustar uma elipse ao gráfico que expressa a variável controlada (pv) em função da op . O terceiro método, desenvolvido por Hägglund [39], estima a histerese a partir da pv , do ganho estático do processo e dos parâmetros do controlador. Embora apenas informações disponíveis na maioria das situações práticas sejam necessárias, estes métodos dependem de tratamentos nos dados (filtragem e decimação) que devem ser ajustados caso a caso.

Em outro trabalho, Ravanbod-Shirazi e Besançon-Voda [66] desenvolveram um método de identificação para um atuador pneumático, no qual os parâmetros do modelo de atrito de Karnopp, com exceção do atrito estático F_s , são estimados usando uma regressão linear. Romano e Garcia [69] propuseram alterações nesse método com a finalidade de aplicá-lo em válvulas de controle. A rigor trata-se de duas modificações: (1) inclusão do termo relativo à força da mola, $-k_{mola} \cdot x(k)$, na etapa que utiliza regressão linear e (2) proposta de uma forma alternativa para estimar F_s através de medidas em instantes de iminência de movimento. Tal método é descrito a seguir.

2.3.1 Identificação do modelo de Karnopp

Baseando-se no modelo de Karnopp descrito na seção 2.2.1, quando a velocidade da haste é maior do que DV , o balanço de forças (2.1) em instantes k é dado por:

$$F_{ext}(k) = m \cdot \ddot{x}(k) + F_v \cdot \dot{x}(k) + F_c \cdot \text{sgn}(\dot{x}(k)) + k_{mola} \cdot x(k). \quad (2.8)$$

Note que (2.8) é linear em relação aos parâmetros m , F_v , F_c e k_{mola} . Logo, se os valores de $\ddot{x}(k)$, $\dot{x}(k)$, $x(k)$ e $F_{ext}(k)$ forem conhecidos, ou puderem ser estimados[§],

[§]Um algoritmo simples e eficiente para o cálculo numérico de derivadas pode ser encontrado em Aguirre [1].

pode-se escrever o seguinte modelo de regressão linear:

$$\hat{F}_{ext}(k) = \phi^T(k) \cdot \lambda, \quad (2.9)$$

em que o vetor de parâmetros λ e o vetor de regressores $\phi(k)$ são dados por (2.10) e (2.11), respectivamente:

$$\lambda = \begin{pmatrix} m & F_v & F_c & k_{mola} \end{pmatrix}^T \quad (2.10)$$

$$\phi(k) = \begin{pmatrix} \ddot{x}(k) & \dot{x}(k) & \text{sgn}(\dot{x}(k)) & x(k) \end{pmatrix}^T. \quad (2.11)$$

Uma forma de estimar o vetor de parâmetros (2.10) é minimizando o erro quadrático entre a força externa medida e a estimada (2.9):

$$\hat{\lambda} = \arg \min_{\lambda} \sum_k \left(F_{ext}(k) - \phi^T(k) \cdot \lambda \right)^2$$

cuja solução é dada por [58]:

$$\hat{\lambda} = \left(\sum_k \phi(k) \cdot \phi^T(k) \right)^{-1} \cdot \sum_k \phi(k) \cdot F_{ext}(k). \quad (2.12)$$

Entretanto, os períodos em que a expressão (2.8) é válida, *a priori*, não são conhecidos, pois a velocidade limite DV é uma incógnita do modelo. Para lidar com este problema, define-se a variável $\delta v(s)$, tal que:

$$\delta v(s) = \frac{s}{\mathcal{Z}} \cdot \max(|\dot{x}(k)|), \quad (2.13)$$

sendo: $\mathcal{Z} \gg 1$ e $\mathcal{S} < \mathcal{Z}$, para $s = 1, 2, \dots, \mathcal{S}$. Em palavras, $\delta v(s)$ assume $\mathcal{S}$ valores distintos, obrigatoriamente menores do que a maior velocidade verificada nos dados experimentais. A escolha do par de constantes $(\mathcal{S}, \mathcal{Z})$ determina quão pequena é a diferença entre dois valores consecutivos de $\delta v(s)$, ou seja:

$$\Delta \delta v \triangleq \delta v(s+1) - \delta v(s).$$

Para cada valor de $\delta v(s)$, os valores do vetor de regressores $\phi(k)$ e da força externa aplicada $F_{ext}(k)$ são escolhidos a partir dos dados experimentais, de modo

que a condição $|\dot{x}(k)| > \delta v(s)$ seja satisfeita e então $\hat{\lambda}$ é calculado através de (2.12). Portanto, os valores da variável auxiliar $\delta v(s)$ são usados como valores candidatos a DV .

O comportamento da estimativa do vetor de parâmetros, à medida que s aumenta, pode ser caracterizado em duas fases [66]. Primeiramente, para $\delta v(s) \ll DV$, os valores de $\hat{\lambda}$ variam significativamente para diferentes valores de $\delta v(s)$, pois dados correspondentes a instantes em que o balanço de forças (2.8) não se aplica, são usados na estimativa. Por outro lado, à medida que s aumenta e $\delta v(s) \approx DV$, as estimativas aproximam-se dos valores reais dos parâmetros e se mantêm quase constantes, mesmo quando $\delta v(s)$ ultrapassa o valor ideal de DV . Entretanto, devido à presença de ruído em dados experimentais, valores muito grandes de s podem fazer com que poucos pontos sejam incorporados a $\phi(k)$ e a $F_{ext}(k)$, resultando em estimativas incoerentes de λ [70].

Com o propósito de validar $\hat{\lambda}$, define-se o desvio-padrão das estimativas por:

$$\sigma(s) = \sqrt{\frac{\sum_{k=1}^{n_s} \left(F_{ext}^s(k) - \hat{F}_{ext}^s(k) \right)^2}{n_s - n_p}}, \quad (2.14)$$

em que $F_{ext}^s(k)$ e $\hat{F}_{ext}^s(k)$ são, respectivamente, a força externa medida e a sua estimativa calculada usando (2.9), para cada valor distinto de $\delta v(s)$; n_s é o número de pontos em $F_{ext}^s(k)$ e n_p é a dimensão de λ ($n_p = 4$, neste caso). Como o desvio-padrão, $\sigma(s)$, é proporcional à variabilidade de $\hat{F}_{ext}^s(k)$, a região onde $\sigma(s)$ é mínimo, deve corresponder ao valor de s onde $\hat{\lambda}$ foi escolhido.

Outra ferramenta que pode ser usada na validação do vetor de parâmetros estimados é o coeficiente de correlação, $R^2(s)$, que é calculado através de [66]:

$$R^2(s) = \frac{\sum_{k=1}^{n_s} \left(\hat{F}_{ext}^s(k) - \bar{F}_{ext}^s(k) \right)^2}{\sum_{k=1}^{n_s} \left(F_{ext}^s(k) - \bar{F}_{ext}^s(k) \right)^2}, \quad (2.15)$$

sendo que $\overline{F}_{ext}^s(k)$ é a média temporal de $F_{ext}^s(k)$. À medida que a estimativa da força externa aproxima-se do valor experimental, o coeficiente de correlação assume valores cada vez mais próximos de 1.

Apesar de ser possível estimar a velocidade limite DV [66], o efeito deste parâmetro é desprezível na simulação de modelos de válvulas de controle [51, 69]. Portanto, usualmente adota-se valores arbitrários entre 10^{-4} e 10^{-2} m/s para este parâmetro.

Finalmente, o coeficiente referente ao atrito estático, F_s , deve ser obtido de modo a minimizar o erro quadrático em relação à velocidade, ou seja:

$$\hat{F}_s = \arg \min_{F_s} \left(\sum_k \left(x(k) - \hat{x}(\hat{\lambda}, F_s, k) \right)^2 \right) \quad (2.16)$$

sujeito a: $\hat{F}_c \leq \hat{F}_s \leq \max(|F_{ext}(k) - k_{mola} \cdot x(k)|)$.

O termo $\hat{x}(\hat{\lambda}, F_s, k)$ denota a posição da haste obtida através de simulações do modelo de Karnopp com os parâmetros previamente estimados, e com a incógnita F_s , que é limitada pelos valores do coeficiente de atrito de Coulomb F_c e da máxima força motriz aplicada ao diafragma da válvula. Portanto, deve-se aplicar algum algoritmo de otimização não-linear, como por exemplo o Simplex de Nelder-Mead [84], para resolver (2.16).

Para evitar que a otimização (2.16) resulte em um mínimo local, é desejável que se disponha de uma estimativa inicial do coeficiente de atrito estático próxima do valor “ótimo”. Uma forma de obter tal estimativa é utilizar o balanço de forças (2.1), nos instantes k_b em que a haste está prestes a se mover. Sob tais condições, denominadas situações de *breakaway*, de (2.4) segue que:

$$F_s \cdot \text{sgn}(F_{ext}(k_b) - k_{mola} \cdot x(k_b)) = F_{ext}(k_b) - k_{mola} \cdot x(k_b). \quad (2.17)$$

Os instantes de iminência de movimento k_b podem ser facilmente reconhecidos por inspeção visual dos gráficos de força externa e de posição da haste. As situações

de *breakaway* também podem ser identificadas por valores elevados da aceleração $\ddot{x}(k)$.

Uma vez que o valor de k_{mola} foi previamente estimado por (2.12), pode-se calcular uma estimativa inicial para F_s , a partir de (2.17). Com o intuito de reduzir eventuais erros devido à imprecisão na leitura de $F_{ext}(k_b)$ e $x(k_b)$, é conveniente determinar o coeficiente de atrito estático usando a média dos valores obtidos nos diferentes instantes em que a haste está prestes a se mover.

Embora este procedimento seja baseado na inspeção visual do comportamento da haste, que pode levar a estimativas imprecisas, resultados experimentais mostraram que o valor do atrito estático dado por (2.17) é bem próximo do valor calculado por métodos alternativos [79]. Este fato sugere que muitas vezes, a otimização não-linear (2.16) pode ser evitada.

Como qualquer outro método de identificação, este algoritmo requer que os dados sejam persistentemente excitantes, para garantir a inversão matricial em (2.12). Além disso, recomenda-se um sinal de excitação cuja amplitude varie dentro de uma ampla faixa de valores [18, 66], para que informações do comportamento não-linear da válvula também estejam presentes nos dados.

Os resultados da aplicação deste método para estimar os parâmetros de atrito de uma válvula real são mostrados na subseção 5.2.2.

2.4 Considerações finais

Em situações onde a posição da haste e a força aplicada ao atuador da válvula sejam conhecidos (ou puderem ser estimadas), o modelo obtido através deste método pode ser usado no projeto de compensadores de atrito baseados em modelo ou no diagnóstico de falhas. Entretanto, quando esta hipótese não for verdadeira, surge a necessidade de uma nova estratégia para identificação do modelo de atrito.

Existem alguns métodos de identificação de atrito baseados nos modelos empí-

ricos, que são capazes de estimar a histerese [7, 39] ou o atrito estático [73, 74] de uma válvula de controle, sem a informação da posição da haste ou da força aplicada ao atuador. O problema de tais métodos é serem baseados em modelos que contemplem um único parâmetro.

Os métodos de detecção [22] e quantificação [23, 45] de atrito propostos na literatura permitem que os parâmetros S e J do modelo empírico de Kano, apresentado na seção 2.2.3, sejam estimados sem a disponibilidade de $x(k)$ ou $F_{ext}(k)$. Todavia, nestes trabalhos modelos de estruturas lineares são considerados para representar a dinâmica do processo. Esta hipótese pode ser razoável se o processo operar numa faixa restrita. Entretanto, diante de perturbações significativas ou quando o processo apresentar comportamento altamente não-linear, estes métodos podem ter o desempenho deteriorado [73].

No capítulo 4 é proposta uma extensão para os métodos de identificação existentes [23, 45, 73], através da qual é possível tanto estimar os parâmetros do modelo de atrito S e J , como identificar um modelo dinâmico não-linear para o processo, usando a estrutura Wiener de modelos de blocos interconectados.

Capítulo 3

Identificação de modelos de processos não-lineares

Nas últimas décadas, a identificação de sistemas usando estruturas de modelos lineares foi intensamente explorada, tanto que atualmente existem ferramentas padronizadas e bem estabelecidas para tratar tal problema. Em muitas situações práticas, aproximações por estruturas lineares em torno de uma condição de trabalho proporcionam resultados satisfatórios. Entretanto, modelos lineares não são capazes de representar certos regimes dinâmicos característicos de sistemas não-lineares. Além disso, a maioria dos sistemas reais são inerentemente não-lineares, o que acaba comprometendo, de alguma forma, a eficácia dos métodos de identificação linear, principalmente em situações onde o sistema em questão possua uma ampla região de operação.

Há uma grande variedade de estruturas de modelos não-lineares, dentre as quais podem-se citar: a série de Volterra [13], os modelos NARMAX (do Inglês, *Nonlinear Auto Regressive Moving Average with eXogenous variables*) polinomiais e racionais [1], as estruturas baseadas em rede neurais artificiais [40] e os modelos de blocos interconectados [14]. Uma boa revisão a respeito das estruturas disponíveis para representar sistemas não-lineares foi publicada há alguns anos por Pearson [64].

Este capítulo é dedicado a apresentar algumas das estruturas de modelos de blocos interconectados para a identificação “caixa-preta” de sistemas não-lineares. A partir da seção 3.2 apenas o modelo de Wiener é considerado, pois este será

empregado na proposta de uma metodologia para a identificação de modelos de processos sujeitos a não-linearidades, provocadas pelo atrito em válvulas de controle.

3.1 Modelos de blocos interconectados

Em muitos casos, sistemas não-lineares podem ser representados através da interconexão de blocos dinâmicos lineares $\mathcal{L}$ e de blocos estáticos não-lineares $\mathcal{N}$. Esta classe de estruturas, conhecida por modelos de blocos interconectados, tem sido utilizada com sucesso na identificação de modelos para diversas aplicações práticas: colunas de destilação [63], processos de neutralização de pH [49, 41], processos biológicos [87], reatores de polimerização [17], fenômenos de histerese [43], célula combustível [47], entre outras.

Os modelos de blocos interconectados possuem diversas vantagens em relação a outras representações de sistemas não-lineares: (1) são facilmente utilizados no projeto de controladores [17, 41]; (2) a análise da estabilidade destes modelos pode ser feita através do bloco dinâmico linear [24]; (3) a forma direta como o conhecimento prévio do sistema pode ser incorporado; (4) utilizam algoritmos de estimação simples, que demandam pouco esforço computacional.

De acordo com a forma como os bloco(s) estático(s) e dinâmico(s) são combinados, surgem diferentes estruturas de modelos. A mais comum é a de Hammerstein, em que uma função estática não-linear precede um bloco dinâmico linear, $\mathcal{N} \rightarrow \mathcal{L}$, conforme ilustrado na Figura 3.1, onde $u(k)$, $w(k)$, $v(k)$ e $y(k)$ representam os sinais de entrada, intermediário entre $\mathcal{N}$ e $\mathcal{L}$, das perturbações e da saída do modelo, respectivamente.

Como a não-linearidade atua somente sobre o sinal de entrada $u(k)$, a dinâmica dos modelos de Hammerstein não depende do ponto de operação [24]. Por outro lado, o mapeamento entre valores estacionários da entrada $\bar{u}$ e da saída $\bar{y}$ do modelo pode ser não-linear. Portanto, tais modelos podem exibir multiplicidade de entrada,

Figura 3.1: Diagrama de blocos de um modelo de Hammerstein.

ou seja, para um dado $\bar{y}$ pode haver mais de um $\bar{u}$, desde que a função do bloco $\mathcal{N}$ não seja invertível [65].

Outra forma muito usada para representar sistemas não-lineares, usando a conexão dos blocos $\mathcal{L}$ e $\mathcal{N}$, é ligando a saída de um bloco dinâmico linear a um bloco não-linear ($\mathcal{L} \rightarrow \mathcal{N}$). Esta estrutura, cujo diagrama de blocos está ilustrado na Figura 3.2, é conhecida como modelo de Wiener. Note que o termo de perturbações $v(k)$ é adicionado na saída do bloco $\mathcal{L}$. Com isso, as perturbações também estão sujeitas à curva de ganho do processo, portanto, quanto maior o ganho, maior será a sensibilidade de $y(k)$ a eventuais distúrbios. Apesar de não modelar adequadamente o ruído de medição, esta suposição é mais coerente, do ponto de vista de processos industriais, pois a influência das perturbações, normalmente, é mais significativa do que a proporcionada por ruídos de medição [89].

Aguirre *et al.* [3] mostraram que os autovalores do modelo de Wiener linearizado em torno de um ponto de operação $(\bar{u}, \bar{y})$ dependem explicitamente dos valores estacionários da entrada e da saída. Logo, diferentemente dos modelos de Hammerstein, a dinâmica desses depende do ponto de operação. Ademais, o ganho em estado estacionário dos modelos de Wiener, também varia com o par $(\bar{u}, \bar{y})$.

Ao combinar os modelos de Hammerstein e de Wiener, ou seja, se o bloco dinâmico for envolvido por dois blocos estáticos não-lineares ($\mathcal{N}_1 \rightarrow \mathcal{L} \rightarrow \mathcal{N}_2$),

Figura 3.2: Estrutura do modelo de Wiener.

obtém-se a estrutura chamada Hammerstein-Wiener [90], ilustrada na Figura 3.3(a). Apesar de não ter recebido a mesma atenção que receberam os modelos compostos pela combinação de dois blocos, do ponto de vista prático, uma boa motivação para usar esta representação é que $\mathcal{N}_1$ poderia reproduzir o comportamento não-linear de um dispositivo atuador, enquanto o bloco $\mathcal{N}_2$ seria usado para incorporar as não-linearidades do processo.

Figura 3.3: Modelos de Hammerstein-Wiener (a) e de Lur'e (b).

Há ainda outros modelos de blocos interconectados, como por exemplo, o modelo Wiener-Hammerstein [14], no qual um bloco estático não-linear é envolto por dois dinâmicos lineares ($\mathcal{L}_1 \rightarrow \mathcal{N} \rightarrow \mathcal{L}_2$); e o modelo de Lur'e* [64], em que uma função estática não-linear é incluída num ramo de realimentação, ao redor de um bloco dinâmico linear, como mostrado na Figura 3.3(b).

Como será visto mais adiante, na metodologia de identificação do conjunto atrito + processo, o atrito proveniente da válvula de controle é modelado por uma das estruturas apresentadas no capítulo 2. Portanto, a finalidade da estrutura de blocos

*A estrutura do modelo de Lur'e é capaz de exibir multiplicidade de saída, ou seja, um valor fixo da entrada $\bar{u}$ poder proporcionar mais de um valor estacionário na saída $\bar{y}$, desde que a função estática não-linear no ramo de realimentação não seja invertível [65].

interconectados é de representar apenas o modelo do processo, ou seja, estimar os parâmetros de um modelo dinâmico não-linear “caixa-preta”, que relacione a variável manipulada com a saída do processo. O modelo de Wiener (Figura 3.2) é considerado para esse propósito, por apresentar algumas vantagens, quando comparado a outras estruturas de blocos interconectados:

- É capaz de incorporar o comportamento não-linear em que as características dinâmicas do modelo variem com o ponto de operação;
- Em relação aos modelos de Hammerstein-Wiener e de Lur’e, apresenta uma estrutura que demanda algoritmos mais simples e eficientes para a estimação de parâmetros;
- O conjunto atrito + processo, que será tratado no capítulo 4, pode ser visto como um modelo Hammerstein-Wiener (Figura 3.3(a)).

3.2 Identificação de modelos de Wiener

Um dos primeiros trabalhos de que se tem ciência a respeito da identificação de modelos de Wiener foi realizado por Billings e Fakhouri [14]. Este método resulta em modelos não-paramétricos por meio da análise de funções de correlação, não sendo apropriado para aplicações práticas, pois requer sinais específicos de excitação (ruído branco). Além disso, a variância das estimativas costuma ser elevada.

A maior dificuldade que surge ao estimar os parâmetros de modelos de blocos interconectados é que os sinais intermediários não são conhecidos [24]. No caso do modelo de Wiener, este sinal é a saída do bloco dinâmico linear, $w(k)$, conforme a Figura 3.2. No algoritmo desenvolvido por Pearson e Pottmann [65], admite-se que a curva estática do sistema $\bar{y} = f(\bar{u})$ seja conhecida *a priori*. Deste modo, os parâmetros do bloco dinâmico são estimados impondo restrições ao método clássico usado em modelos ARX (do Inglês, *Auto Regressive with eXogenous variables*).

Entretanto, a característica estática de um sistema muitas vezes não é conhecida, ou então, não é viável realizar ensaios para coletar dados em estado estacionário ao longo de toda a faixa de interesse. Para lidar com tais fatos, Coelho [24] propôs duas formas para obter a curva estática a partir de dados dinâmicos: (1) um modelo NARMAX é identificado e a função $\bar{y} = f(\bar{u})$ é determinada usando conceitos de agrupamentos de termos e de pontos fixos de modelos não-lineares [1]; (2) a característica estática é estimada diretamente por um algoritmo iterativo, baseado na análise do coeficiente de agrupamento de termos.

Métodos iterativos também podem ser usados na identificação do modelo $\mathcal{L} \rightarrow \mathcal{N}$. Neste caso, numa primeira iteração, admite-se que exista e sejam conhecidos os parâmetros do mapeamento $w(k) = g(y(k))$. Logo, o sinal $w(k)$ é calculado e o bloco $\mathcal{L}$ é estimado. Em seguida, a partir desta estimativa e do sinal de entrada $u(k)$, determina-se o sinal intermediário, que é usado para identificar os parâmetros do bloco estático não-linear. Este procedimento é repetido até que não haja variação significativa nos parâmetros. Apesar de proporcionar resultados satisfatórios [82, 89], ainda não se demonstrou a convergência dos algoritmos que utilizam esta metodologia.

Westwick e Verhaegen [85] estenderam o algoritmo MOESP (do Inglês, *Multivariable Output-Error State sPace*), pertencente ao esquema de identificação por subespaços, para estimar os parâmetros do modelo de Wiener. Apesar de não-iterativo e eficiente em termos de esforço computacional, Forssell [30] mostrou que o método de subespaços não é consistente quando aplicado à identificação de modelos com dados de malha fechada.

Para evitar os problemas de mínimos locais e de convergência dos algoritmos iterativos, Kalafatis *et al.* [49] parametrizaram o bloco dinâmico linear através de um filtro para amostragem de frequência ou FSF (do Inglês, *Frequency Sampling Filter*), de modo que os parâmetros pudessem ser determinados analiticamente, desde que o sistema não esteja sujeito a perturbações (introduzidas na saída do bloco linear) ou ruídos de medição.

Estruturas de modelos mais complexas exigem métodos iterativos para a estimação de parâmetros, portanto, há o risco de se obter um mínimo local. Por outro lado, ao adotar uma estrutura menos “flexível”, que requeira uma quantidade maior de parâmetros, pode-se obter um estimador linear e, conseqüentemente, reduzir (ou até eliminar) o problema de mínimos locais, sob pena de aumentar a variância da estimativa[†].

Hagenblad [36] propôs uma metodologia de múltiplas etapas para lidar com estas questões. Primeiramente, o bloco dinâmico foi parametrizado por um modelo de resposta impulsiva finita, ou simplesmente FIR (do Inglês, *Finite Impulse Response*), de ordem elevada, de modo que uma estimativa inicial fosse obtida a partir de uma regressão linear. Em seguida, esta estimativa é reduzida para uma estrutura de função de transferência racional. Tjörnström [76] mostrou que a variância da estimativa usando esta estratégia de dois passos é menor ou igual à que seria obtida por um procedimento direto. Por fim, utiliza-se o método de Gauss-Newton [58] para obter numericamente os parâmetros do modelo de Wiener, a partir da estimativa obtida na etapa anterior.

A seguir é descrito um procedimento para a estimação de parâmetros de modelos de Wiener a partir de dados de operação em malha fechada, sujeitos a perturbações não-lineares (Figura 3.2).

3.2.1 Enunciado do problema

No diagrama da Figura 3.4, $G(q, \theta)$ e $H(q, \theta)$ são blocos dinâmicos lineares e invariantes no tempo, parametrizados por θ e q denota o operador de deslocamento, tal que:

$$q^{-2} \cdot u(k) = u(k - 2).$$

[†]A variância das estimativas é proporcional à quantidade de parâmetros e à intensidade de perturbações no processo [88].

A função estática não-linear, cuja parametrização é feita através do vetor η , é representada por $f(\cdot, \eta)$ e $K(q)$ é um controlador linear. Os sinais $r(k)$, $u(k)$, $y(k)$, $v(k)$ e $e(k)$ denotam, respectivamente, o valor de referência, a entrada, a saída, as perturbações do processo e um ruído branco de média nula e variância σ_e^2 . Portanto, trata-se de um processo modelado por uma estrutura de Wiener operando em malha fechada.

Neste trabalho pretende-se estudar uma forma de identificar os parâmetros do modelo de Wiener empregando um método de identificação direta, ou seja, apenas os dados de entrada e saída do processo são usados para estimar θ e η .

Figura 3.4: Esquema para identificação em malha fechada.

Primeiramente é necessário escolher um critério através do qual θ e η serão estimados. Em uma abordagem muito comum na literatura [8, 49, 85, 86], as perturbações do processo $v(k)$ são desconsideradas, de modo que o preditor de $y(k)$ seja dado por:

$$\hat{y}(k|\theta, \eta) = f(G(q, \theta)u(k), \eta). \quad (3.1)$$

Deste modo, θ e η são calculados através de:

$$(\hat{\theta}, \hat{\eta}) = \arg \min_{\theta, \eta} \sum_k (y(k) - f(G(q, \theta)u(k), \eta))^2. \quad (3.2)$$

Apesar de este critério ser adotado até mesmo em pacotes comerciais [61, 62], Hagenblad *et al.* [38] mostraram que o estimador (3.2) é polarizado caso haja dependência entre $u(k)$ e $v(k)$. Portanto, não seria conveniente utilizar esta abordagem no presente contexto, pois, devido ao efeito da realimentação (Figura 3.4), sempre haverá tal correlação em alguma intensidade.

Caso a função estática $f(\cdot, \eta)$ seja invertível, pode-se adotar o seguinte critério:

$$V(\theta, \vartheta) = \sum_k \epsilon^2(k, \theta, \vartheta) \quad (3.3)$$

em que $\epsilon(k, \theta, \vartheta)$ é dado por:

$$\begin{aligned} \epsilon(k, \theta, \vartheta) &= H^{-1}(q, \theta) (f^{-1}(y(k), \vartheta) - G(q, \theta)u(k)) \\ &= H^{-1}(q, \theta) (w(k, \vartheta) - G(q, \theta)u(k)). \end{aligned} \quad (3.4)$$

Note que o mapeamento de $y(k)$ em $w(k)$, expresso por $w(k) = f^{-1}(y(k), \vartheta)$, é parametrizado através do vetor ϑ .

Em alguns trabalhos que tratam da identificação de modelos de blocos interconectados [19, 90, 91] argumenta-se que, ao minimizar o critério (3.3), é possível calcular estimativas consistentes para o cenário representado na Figura 3.4. Além disso, minimizar $V(\theta, \vartheta)$ é equivalente a obter os parâmetros que maximizem a função de máxima verossimilhança de $y(k)$ [38].

Outra forma de interpretar o critério (3.3) é considerar que $V(\theta, \vartheta)$ quantifica o erro de predição do sinal intermediário $w(k)$. Basta verificar que:

$$\begin{aligned} \epsilon(k, \theta, \vartheta) &= f^{-1}(y(k), \vartheta) - \hat{w}(k|\theta, \vartheta) \\ &= w(k, \vartheta) - \hat{w}(k|\theta, \vartheta), \end{aligned} \quad (3.5)$$

em que o preditor de um passo à frente[†] do sinal intermediário é dado por [58]:

$$\hat{w}(k|\theta, \vartheta) = H^{-1}(q, \theta)G(q, \theta)u(k) + (1 - H^{-1}(q, \theta)) f^{-1}(y(k), \vartheta). \quad (3.6)$$

Uma vez escolhido o critério a ser usado para estimar os parâmetros do modelo de Wiener, a seguir são parametrizados os blocos $G(q, \theta)$, $H(q, \theta)$, $f(w(k), \eta)$ e o mapeamento $f^{-1}(y(k), \vartheta)$.

[†]Expressão usada para calcular $w(k, \vartheta)$ a partir dos dados observados até o instante $k - 1$.

3.2.2 Parametrização dos blocos $\mathcal{L}$ e $\mathcal{N}$

Em muitos casos, a função estática não-linear é aproximada por polinômios [24, 82]. Todavia, sabe-se que os polinômios de ordem elevada podem apresentar comportamento oscilatório e tornar a estimação de parâmetros mal-condicionada, além de apresentar resultados insatisfatórios ao extrapolar certos trechos de dados [89]. Para evitar estes problemas, muitas vezes [36, 81] utilizam-se funções lineares por partes.

Os problemas apresentados por polinômios não ocorrem com as transformações lineares por partes, no entanto, existe a desvantagem de a derivada destas funções não ser contínua na transição entre trechos adjacentes. Além disso, pode ser necessário um elevado número de parâmetros para aproximar determinadas funções. Uma alternativa interessante é modelar o bloco não-linear com *splines* cúbicas, que são polinômios de terceira ordem seccionados por trechos, contidos no intervalo $[w_{\min}; w_{\max}]$ e definidos pelos pontos $\{w_1, \dots, w_m\}$, tal que:

$$w_{\min} = w_1 < w_2 < \dots < w_{m-1} < w_m = w_{\max}. \quad (3.7)$$

Para qualquer $w \in [w_{\min}; w_{\max}]$, uma *spline* cúbica, cujos trechos sejam especificados por (3.7), pode ser representada através de [56]:

$$y(k) = f(w(k), \eta) = \sum_{i=2}^{m-1} f_i |w(k) - w_i|^3 + f_m + f_{m+1} w(k), \quad (3.8)$$

em que m é o grau da *spline* cúbica e $f_2, \dots, f_{m+1}$ são números reais que a definem. Portanto, uma *spline* de grau m é constituída de m parâmetros, desde que os pontos $\{w_1, \dots, w_m\}$ sejam conhecidos *a priori*. A escolha destes pontos é discutida na subseção 3.2.6. Vale ressaltar ainda que as derivadas de primeira e segunda ordem da função descrita em (3.8) são contínuas [56].

Neste trabalho, admite-se que o mapeamento não-linear $y(k) = f(w(k), \eta)$ é monotônico e invertível, ou seja, existe uma função $w(k) = g(y(k), \vartheta)$. Logo, é

válido representar o sinal intermediário em função da saída do processo:

$$w(k) = g(y(k), \vartheta) = \sum_{i=2}^{m-1} g_i |y(k) - y_i|^3 + g_m + g_{m+1}y(k), \quad (3.9)$$

em que os valores de y_i (para $i = 1, \dots, m$) satisfazem a condição:

$$y_{\min} = y_1 < \dots < y_i < \dots < y_m = y_{\max}.$$

A não ser que haja algum conhecimento prévio acerca do mapeamento (3.9), uma escolha razoável é admitir $y_{\min}$ e $y_{\max}$ como os valores mínimo e máximo contidos nos dados experimentais, e os demais valores de y_i distribuídos uniformemente no intervalo $[y_{\min}; y_{\max}]$.

A parte dinâmica do modelo de Wiener é representado por uma estrutura ARMAX (do Inglês, *Auto Regressive Moving Average with eXogenous variables*). A principal motivação para esta escolha é o fato de a estrutura ARMAX ser a forma mais simples de parametrizar os modelos do processo e das perturbações de forma independente. Logo, $G(q, \theta)$ e $H(q, \theta)$ em (3.4) são dados por:

$$G(q, \theta) = \frac{b_1 q^{-1} + \dots + b_l q^{-l}}{1 + a_1 q^{-1} + \dots + a_l q^{-l}} = \frac{B(q)}{A(q)} \quad (3.10)$$

$$H(q, \theta) = \frac{1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c}}{1 + a_1 q^{-1} + \dots + a_l q^{-l}} = \frac{C(q)}{A(q)}. \quad (3.11)$$

Portanto, de acordo com as parametrizações (3.8)-(3.11) segue que:

$$\theta = (b_1, \dots, b_l, a_1, \dots, a_l, c_1, \dots, c_{n_c})^T \quad (3.12)$$

$$\eta = (f_2, \dots, f_{m+1})^T \quad (3.13)$$

$$\vartheta = (g_2, \dots, g_{m+1})^T. \quad (3.14)$$

Ao substituir (3.10) e (3.11) em (3.4) resulta:

$$\begin{aligned} \epsilon(k, \theta, \vartheta) &= \frac{A(q)}{C(q)} \left(g(y(k), \vartheta) - \frac{B(q)}{A(q)} u(k) \right) \\ &= \frac{1}{C(q)} (A(q)g(y(k), \vartheta) - B(q)u(k)). \end{aligned} \quad (3.15)$$

Para a estrutura adotada, o erro de predição $\epsilon(k, \theta, \vartheta)$ é não-linear em relação aos parâmetros θ e ϑ . Logo, o critério (3.3) deve ser minimizado numericamente. Todavia, $V(\theta, \vartheta)$ pode apresentar diversos mínimos locais [89] e o algoritmo de otimização numérica não necessariamente converge para o mínimo global. Há, basicamente, duas formas de lidar com este problema: (1) efetuar a busca a partir de diversas condições iniciais; ou (2) obter uma estimativa inicial que esteja na região de atração do mínimo global.

A primeira abordagem demandaria elevado esforço computacional. Por outro lado, alguns autores [37, 49] mostraram que é possível calcular uma estimativa inicial do modelo de Wiener, de forma não-iterativa. Como mostrado na subseção 3.2.3, ao aproximar $G(q, \theta)$ por uma estrutura FIR, estimativas do bloco dinâmico linear e do mapeamento entre a saída $y(k)$ e o sinal intermediário $w(k)$ são obtidas a partir da solução de um problema de programação quadrática.

Em seguida, o modelo FIR é reduzido para a estrutura definida em (3.10). Esta etapa, assim como uma forma de determinar a ordem l do modelo são tratadas na subseção 3.2.4. O método empregado para minimizar numericamente $V(\theta, \vartheta)$ é discutido na subseção 3.2.5. Por fim, uma forma de calcular (3.13), que parametrize o bloco $\mathcal{N}$, é apresentada na subseção 3.2.6.

3.2.3 Estimativa inicial de $G(q)$ e $f^{-1}(\cdot)$

Para obter uma estimativa inicial dos parâmetros, consideram-se as seguintes suposições, referentes ao modelo de Wiener:

1. O bloco com dinâmica linear $G(q, \theta)$ é estável em malha aberta;
2. As perturbações no processo são dominantes em relação a eventuais ruídos de medição, de modo que os mesmos possam ser desprezados. Em outras palavras, a estrutura do modelo de Wiener apresentada na Figura 3.2 é apropriada para representar o sistema em questão.

De acordo com a primeira suposição, aproxima-se o bloco dinâmico linear por um modelo FIR, de ordem suficientemente elevada para incorporar a dinâmica do sistema. Logo, o sinal intermediário $w(k)$, ilustrado na Figura 3.4, é dado por:

$$\begin{aligned} w(k) &= \beta_1 u(k-1) + \dots + \beta_n u(k-n) + v(k) \\ &= \sum_{i=1}^n \beta_i u(k-i) + v(k). \end{aligned} \quad (3.16)$$

Ao substituir $w(k)$ de (3.9) em (3.16) e isolando o termo $v(k)$, referente às perturbações, resulta:

$$v(k) = \sum_{i=2}^{m-1} g_i |y(k) - y_i|^3 + g_m + g_{m+1} y(k) - \sum_{i=1}^n \beta_i u(k-i). \quad (3.17)$$

Sejam os vetores de regressores $\psi(k)$ e de parâmetros ξ :

$$\begin{aligned} \psi(k) &\triangleq (|y(k) - y_2|^3, \dots, |y(k) - y_{m-1}|^3, 1, y(k), \\ &\quad -u(k-1), \dots, -u(k-n))^T \end{aligned} \quad (3.18)$$

$$\xi \triangleq (g_2, \dots, g_{m-1}, g_m, g_{m+1}, \beta_1, \dots, \beta_n)^T. \quad (3.19)$$

De (3.18) e (3.19), pode-se reescrever (3.17):

$$v(k) = \psi^T(k) \cdot \xi. \quad (3.20)$$

Uma forma de estimar ξ é minimizando a norma ℓ_2 de $v(k)$, que representa a diferença entre a saída do modelo FIR e de $w(k)$ dado por (3.9). Portanto, para uma sequência de N pontos:

$$\hat{\xi} = \arg \min_{\xi} \sum_{k=n+1}^N (\psi^T(k) \cdot \xi)^2. \quad (3.21)$$

Deve-se ressaltar que o estimador (3.21) é polarizado, caso $v(k)$ não seja ruído branco [34, 36]. Entretanto, apesar de $v(k)$ raramente aproxima-se de ruído branco, pois é usado para representar distúrbios do processo, tal polarização pode ser desprezada desde que o vetor ξ gere estimativas iniciais de $G(q)$ e $f^{-1}(\cdot)$ que estejam na região de atração do mínimo global da função objetivo (3.3).

Antes de resolver (3.21), um aspecto comum na identificação de modelos de blocos interconectados deve ser considerado: como não se conhece $w(k)$, o ganho estacionário entre $\bar{u}$ e $\bar{y}$ pode ser distribuído arbitrariamente entre os blocos da associação (neste caso, $\mathcal{L}$ e $\mathcal{N}$). Portanto, a não ser que alguma restrição seja imposta, o problema de estimação admite infinitas soluções.

A forma mais comum de evitar esse problema é impor o valor de um dos elementos de ξ , definido em (3.19). Por exemplo, assim como em outros trabalhos [49, 82, 91], admite-se $g_{m+1} = 1$. Esta abordagem possui a vantagem de ser facilmente implementada, pois, depois de reformular (3.21) devido à restrição em g_{m+1} , a estimação dos parâmetros torna-se um problema de mínimos quadrados lineares.

Outra maneira de garantir a unicidade da representação é através da imposição do ganho estacionário κ do modelo FIR. Tal restrição é expressa através de:

$$\kappa = \sum_{i=1}^n \beta_i = \left(\underbrace{0 \dots 0}_m \underbrace{1 \dots 1}_n \right) \cdot \xi = \mathcal{R} \cdot \xi. \quad (3.22)$$

Como (3.22) é linear em relação aos parâmetros, o vetor ξ que resolve o problema (3.21), sujeito a (3.22), é dado por[§]:

$$\hat{\xi} = \kappa \left(\Psi^T \Psi \right)^{-1} \mathcal{R}^T \left(\mathcal{R}^T \left(\Psi^T \Psi \right)^{-1} \mathcal{R} \right)^{-1}, \quad (3.23)$$

em que a matriz de regressores Ψ é definida por:

$$\Psi \triangleq (\psi(n+1) \dots \psi(N))^T.$$

Apesar de (3.23) envolver mais cálculos do que um problema típico de mínimos quadrados sem restrição, Hagenblad [36] verificou que os tempos de computação para obter $\hat{\xi}$, através de qualquer uma das abordagens citadas, são da mesma ordem de grandeza. Também há uma motivação natural para impor (3.22): como o propósito do bloco $\mathcal{L}$ é incorporar as características dinâmicas, assim como o comportamento estacionário deve ser representado por (3.9), é conveniente admitir $\kappa = 1$.

[§]A expressão (3.23) é derivada no apêndice A usando o método de multiplicadores de Lagrange.

Outro detalhe relacionado à estimativa inicial dos parâmetros é a montagem da matriz Ψ . Esta questão envolve a escolha da ordem do modelo FIR, n , e do grau m do polinômio seccionado que parametriza o mapeamento $w(k) = g(y(k), \vartheta)$. Quanto maior o valor de n , melhor será a aproximação da dinâmica do sistema. Entretanto, esta escolha requer que mais dados estejam disponíveis para que uma estimativa razoável seja obtida. Desde que o intervalo de amostragem tenha sido escolhido apropriadamente[¶], valores de n entre 30 e 50 proporcionam bons resultados práticos [36, 89].

A seleção do grau m da função $w(k) = g(y(k), \vartheta)$ pode ser realizada de modo a minimizar o seguinte critério [89]:

$$V_{oe} = \frac{N + \lambda_p}{N - \lambda_p} \cdot \sum_{k=n+1}^N (\hat{w}'(k) - \hat{w}^*(k))^2, \quad (3.24)$$

em que λ_p é a quantidade de parâmetros, enquanto $\hat{w}'(k)$ e $\hat{w}^*(k)$ são as estimativas do sinal intermediário, calculadas a partir de $\hat{g}(y(k), \vartheta)$ e do modelo FIR, respectivamente. Logo, pode-se reescrever (3.24) como:

$$\begin{aligned} V_{oe}(m) &= \frac{N + (n + m)}{N - (n + m)} \cdot \sum_{k=n+1}^N \left(\hat{g}(y(k), \vartheta) - \sum_{i=1}^n \beta_i q^{-i} u(k) \right)^2 \\ &= \underbrace{\frac{N + (n + m)}{N - (n + m)}}_{\diamond} \cdot \sum_{k=n+1}^N \left(\psi^T(k) \cdot \hat{\xi} \right)^2. \end{aligned} \quad (3.25)$$

O termo identificado por $\diamond$ em (3.25) tem a função de penalizar modelos sobreparametrizados, ao passo que o restante da expressão leva em conta o ajuste do modelo aos dados. A eficácia deste critério foi comparada com abordagens baseadas no erro de predição, como o FPE (do Inglês, *Final Prediction Error*) [88, 89] e mostrou-se mais robusto em situações práticas.

Uma vez obtida uma aproximação do bloco dinâmico linear através de uma estrutura FIR, o próximo passo rumo à identificação do modelo de Wiener é transformar esta estimativa inicial em uma estrutura de função de transferência racional,

[¶]Em Aguirre [2] é descrito um método prático, baseado em funções de correlação, para determinar o intervalo de amostragem de sinais.

conforme a parametrização definida em (3.10) e (3.11). Este procedimento, assim como uma forma de obter a ordem l do modelo reduzido, são apresentados na próxima subseção.

3.2.4 Redução do modelo

O problema de redução de modelo consiste em, a partir de um modelo de ordem n representado por $G_n(q)$, obter uma descrição $G_l(q)$, de ordem l , que minimize a diferença $G_n(q) - G_l(q)$, de acordo com um determinado critério [76, 83]. No presente contexto, pode-se usar o modelo FIR e gerar uma estimativa do sinal intermediário $\hat{w}^*(k)$ para uma dada entrada $u(k)$. Portanto, de acordo com (3.10) a (3.12), o modelo FIR é transformado em uma estrutura ARMAX resolvendo o seguinte problema:

$$\hat{\theta} = \arg \min_{\theta} \sum_{k=l+1}^N \left(\frac{1}{C(q, \theta)} (A(q, \theta) \hat{w}^*(k) - B(q, \theta) u(k)) \right)^2. \quad (3.26)$$

Note que o argumento quadrático é idêntico a $\epsilon(k)$ em (3.15), a não ser por $\hat{w}^*(k)$, ao invés de $f^{-1}(y(k), \vartheta)$. Embora a resolução de (3.26) envolva uma regressão não-linear, este problema pode ser facilmente resolvido usando o estimador estendido de mínimos quadrados [1].

Com relação à escolha da ordem l do modelo reduzido, pode-se empregar um critério semelhante ao usado na determinação de m :

$$V'_{oe}(l, n_c) = \frac{N + (2l + n_c + m)}{N - (2l + n_c + m)} \cdot \sum_{k=l+1}^N \left(\hat{w}^*(k) - \frac{B(q, \theta)}{A(q, \theta)} u(k) \right)^2. \quad (3.27)$$

Assim como em (3.25), este critério leva em consideração tanto o ajuste do modelo aos dados, como o aspecto de sobreparametrização. Como a ordem de $C(q)$ também não é conhecida, na prática, n_c pode ser escolhido por tentativa-e-erro, admitindo valores de forma crescente e verificando o comportamento de (3.27).

3.2.5 Otimização numérica

Conforme dito anteriormente (subseção 3.2.2), a estimação de parâmetros do modelo de Wiener, para a estrutura adotada, é um problema de mínimos quadrados não-linear.

Os métodos de busca iterativa baseados na informação do gradiente da função objetivo $V(\theta, \vartheta)$, de forma geral, são expressos por [58]:

$$\begin{pmatrix} \hat{\theta} \\ \hat{\vartheta} \end{pmatrix}^{(i+1)} = \begin{pmatrix} \hat{\theta} \\ \hat{\vartheta} \end{pmatrix}^{(i)} - \mu^{(i)} \cdot (R^{(i)})^{-1} \cdot \nabla V(\hat{\theta}^{(i)}, \hat{\vartheta}^{(i)}), \quad (3.28)$$

em que $(\cdot)^{(i)}$ denota o valor de $(\cdot)$ na i -ésima iteração, $R^{(i)}$ é uma matriz que determina a direção de busca, $\mu^{(i)}$ é uma constante positiva usada para garantir que:

$$V(\hat{\theta}^{(i+1)}, \hat{\vartheta}^{(i+1)}) < V(\hat{\theta}^{(i)}, \hat{\vartheta}^{(i)})$$

e $\nabla V(\hat{\theta}^{(i)}, \hat{\vartheta}^{(i)})$ é o gradiente da função objetivo em relação aos parâmetros.

A escolha mais simples para $R^{(i)}$ é admiti-la uma matriz identidade. Com isso, (3.28) torna-se o método de “descida mais íngreme” (do Inglês, *steepest-descent*). No entanto, este método é pouco eficiente próximo de uma região de mínimo [36, 58]. Para situações como esta, o método de Gauss-Newton, no qual:

$$R^{(i)} = \sum_k \nabla \epsilon(k, \hat{\theta}^{(i)}, \hat{\vartheta}^{(i)}) \nabla \epsilon^T(k, \hat{\theta}^{(i)}, \hat{\vartheta}^{(i)}) \quad (3.29)$$

é bem mais apropriado [89]. Vale ressaltar que como $V(\theta, \vartheta)$ é quadrática, então:

$$\nabla V(\hat{\theta}^{(i)}, \hat{\vartheta}^{(i)}) = 2 \sum_k \nabla \epsilon(k, \hat{\theta}^{(i)}, \hat{\vartheta}^{(i)}) \cdot \epsilon(k, \hat{\theta}^{(i)}, \hat{\vartheta}^{(i)}). \quad (3.30)$$

Portanto, a despeito da constante μ , a busca (3.28) depende apenas dos resíduos do modelo $\epsilon(k, \theta, \vartheta)$ e de suas primeiras derivadas em relação aos parâmetros, $\nabla \epsilon(k, \theta, \vartheta)$. Tipicamente, o método de Gauss-Newton converge mais rápido do que o de “descida mais íngreme”, com a vantagem (em relação ao método de Newton) de não requerer o cálculo da Hessiana de $V(\theta, \vartheta)$ (segunda derivada da função objetivo).

Na implementação do algoritmo de otimização, o gradiente $\nabla \epsilon(k, \theta, \vartheta)$ foi calculado numericamente, através de aproximações por diferenças.

3.2.6 Estimação do bloco estático não-linear

Uma vez determinados os parâmetros $(g_2, \dots, g_{m+1})$, $w(k)$ pode ser obtido diretamente a partir de (3.9). Logo, o vetor η que parametriza o bloco estático não-linear pode ser obtido da solução de um problema de mínimos quadrados linear:

$$\hat{\eta} = \arg \min_{\eta} \frac{1}{N} \sum_{k=1}^N (y(k) - \varphi^T(k)\eta)^2, \quad (3.31)$$

sendo que:

$$\varphi(k) \triangleq \left(|\hat{w}'(k) - w_2|^3, \dots, |\hat{w}'(k) - w_{m-1}|^3, 1, \hat{w}'(k) \right)^T.$$

Os valores dos pontos $\{w_1, \dots, w_m\}$ podem ser selecionados de duas formas: (1) como, neste momento, o mapeamento $w(k) = f^{-1}(y(k), \vartheta)$ é conhecido, os trechos da *spline* podem ser determinados por inspeção visual, de modo que mais pontos sejam alocados onde haja variações bruscas; ou (2) assim como na parametrização da função inversa $w(k) = f^{-1}(y(k), \vartheta)$, os valores de w_i , para $i = 1, \dots, m$, podem ser distribuídos uniformemente na faixa dos valores de $w(k)$.

Na presente formulação, adotou-se o mesmo grau em ambas as *splines*, (3.8) e (3.9); entretanto, os mesmos procedimentos apresentados anteriormente são aplicáveis, quando uma quantidade distinta de trechos for usada nas parametrizações de $y(k) = f(w(k), \eta)$ e de $w(k) = g(y(k), \vartheta)$.

3.3 Comentários finais

O fluxograma da Figura 3.5 ilustra as etapas envolvidas no procedimento para a estimação dos parâmetros de um modelo de Wiener. Note que para obter o grau m do polinômio seccionado, que representa a função $w = g(y(k), \vartheta)$, todos os valores do intervalo $[3; m_{\max}]$ são testados e, em seguida, o valor que minimiza o critério (3.25) é selecionado, assim como o correspondente valor do vetor de parâmetros $\hat{\xi}$.

De forma análoga, o intervalo $[1; l_{\max}]$ e a expressão (3.27) são considerados na escolha da ordem do modelo reduzido l . Na grande maioria dos casos, $m_{\max} = 6 \sim 8$ e $l_{\max} \approx 5$ são mais que suficientes.

Na subseção (3.2.3) admitiu-se que o sistema a ser identificado seja estável em malha aberta. A grande maioria dos processos industriais apresenta esta característica. No entanto, em situações em que esta suposição não seja válida, outra formulação deve ser usada para estimar os parâmetros do modelo de Wiener, dentre

Figura 3.5: Fluxograma do procedimento para a estimação do modelo de Wiener.

as quais pode-se citar a empregada em Vörös [82] ou em Zhu [89].

Também foi suposto que a função estática não-linear fosse invertível, de modo que o sinal intermediário $w(k)$ pudesse ser obtido a partir da saída $y(k)$. Esta hipótese é verificada em muitas situações práticas, dentre as quais pode-se citar: processos de neutralização de pH [41, 49], reatores químicos [17] e colunas de destilação [91]. No entanto, tal suposição faz com que o modelo de Wiener não apresente multiplicidade de entrada [65].

Capítulo 4

Identificação dos parâmetros do conjunto atrito + processo

Conforme dito anteriormente, o atrito em válvulas é um dos principais responsáveis por causar oscilações nas malhas de controle. Outro fator que compromete o desempenho dos processos industriais é a sintonia inadequada de controladores. Neste capítulo, é proposta uma metodologia para que seja possível, a partir de dados obtidos em malha fechada, quantificar o atrito em válvulas de controle e obter modelos de processos, que possam ser empregados na re-sintonia de controladores.

4.1 Descrição do problema

No capítulo 2 foi descrito um método para a estimação dos parâmetros do modelo de atrito de Karnopp, supondo que a posição da haste da válvula fosse conhecida. Em válvulas modernas, equipadas com atuadores “inteligentes”, esta suposição é correta. Entretanto, o percentual de válvulas deste tipo encontrado na prática é bem pequeno [23]; em especial se consideradas apenas as mais antigas, que potencialmente apresentam atrito elevado.

Por outro lado, o sinal enviado à válvula é registrado, na grande maioria dos casos. Além disso, a tarefa de quantificar o atrito em válvulas deve ser conduzida da forma menos intrusiva possível. Para lidar com estes aspectos, o método proposto a seguir visa quantificar o atrito em válvulas de controle, a partir de um conjunto

de dados compostos apenas pela saída do controlador op e pela variável controlada pv , medidos com o sistema operando em malha fechada. A ideia de uma malha de controle sujeita a atrito é ilustrada na Figura 4.1.

Figura 4.1: Malha de controle diante da presença de atrito.

Como dito anteriormente, interferências externas são indesejáveis, no entanto, note que uma excitação externa é somada ao ao valor de referência (Figura 4.1). Este sinal, cujo projeto é discutido na seção 4.2, é responsável por assegurar a identificabilidade dos parâmetros.

Srinivasan *et al.* [73] trataram do problema de identificar o conjunto atrito + processo, representando o atrito com o modelo de Stenman *et al.* [74], composto por um único parâmetro e o processo através de uma estrutura ARMAX [1, 58]. Entretanto, foi verificado por Garcia [31] que a estrutura não-linear adotada naquele trabalho não é capaz de capturar corretamente o comportamento do atrito em válvulas de controle.

Recentemente verificou-se que, no contexto de válvulas de controle, a contribuição do coeficiente de atrito viscoso linear F_v no modelo de Karnopp (subseção 2.2.1) é desprezível em situações práticas [70]. De acordo com Uehara *et al.* [79], nestes casos os modelos de Karnopp e de Kano [50] são equivalentes. Portanto, optou-se por utilizar a estrutura do modelo de Kano para representar o atrito de uma válvula instalada em uma malha de controle. Tal escolha é justificada pelo fato deste modelo, descrito na subseção 2.2.3, possuir uma estrutura mais simples, composta por apenas dois parâmetros (S, J) .

Outra contribuição deste trabalho é a de empregar uma estrutura não-linear

para representar a dinâmica do processo. Esta escolha visa proporcionar algumas vantagens relevantes:

- Impedir que não-linearidades do processo sejam equivocadamente incorporadas no modelo de atrito;
- Minimizar problemas de polarização na estimativa da dinâmica do processo;
- Tornar o método de identificação apropriado para situações em que os dados sejam obtidos com a malha operando em uma faixa ampla;
- Obter uma aproximação do comportamento estático do processo que pode, eventualmente, ser usado no projeto de um controlador não-linear.

Todavia, a adoção de uma estrutura mais complexa apresenta algumas desvantagens, que não podem ser desprezadas: (1) requer ensaios de maior duração, uma vez que a quantidade de informação a ser obtida dos dados também é maior; e (2) necessidade de algoritmos mais sofisticados.

A partir da malha de controle mostrada anteriormente (Figura 4.1), ao adotar a estrutura de Wiener (seção 3.2) e o modelo de Kano na parametrização dos modelos do processo e do atrito, respectivamente, resulta o diagrama da Figura 4.2. Com o intuito de manter a notação usada no capítulo 3, a posição da haste e a variável controlada pv são denotadas por $u(k)$ e $y(k)$, respectivamente; $d(k)$ é o sinal de excitação externa e $z(k)$ é a saída do controlador, ou seja, o sinal efetivamente aplicado à planta.

Figura 4.2: Diagrama de blocos do modelo da malha de controle.

Em suma, sejam as suposições:

1. O sinal $d(k)$ pode ser manipulado durante o experimento de coleta de dados;
2. Apenas o conjunto de dados formado por $\{z(k), y(k)\}_{k=1}^N$ está disponível para estimação e/ou validação.

O objetivo do restante do capítulo é propor um método para identificar os parâmetros do modelo de atrito (S, J) , do processo $(G(q), f(\cdot))$ e de perturbações $H(q)$, de acordo com a estrutura mostrada na Figura 4.2.

4.2 Projeto do sinal de excitação

A consistência dos algoritmos de identificação em malha fechada depende de um sinal de entrada que não seja gerado exclusivamente pela realimentação da saída [30, 58, 90]. Em outras palavras, uma entrada extra é necessária para garantir que os dados de um experimento em malha fechada sejam suficientemente informativos*. Por este motivo a excitação externa $d(k)$ é necessária no diagrama da Figura 4.2. Com isso, a questão que surge é: que sinal $d(k)$ é mais apropriado para resolver o problema descrito anteriormente?

Forssell [30] derivou as expressões do controlador e do sinal de teste ótimos para a identificação de sistemas lineares. No entanto, estas expressões dependem do modelo e de características das perturbações do processo. Com o intuito de tornar tais resultados mais adequados a aplicações industriais, Zhu e Van den Bosch [92] propuseram uma aproximação para a expressão do espectro ótimo $\Phi_r^*(\omega)$ do sinal de teste que é introduzido no valor de referência da variável controlada:

$$\Phi_r^*(\omega) \approx \zeta \sqrt{\Phi_v(\omega) \Lambda(\omega)}, \quad (4.1)$$

*Um experimento informativo produz dados a partir dos quais é possível determinar os parâmetros do sistema de forma única. Pode ser facilmente mostrado [30] que em um esquema de identificação em malha fechada, um sinal persistentemente excitante [58] na entrada da planta não implica em dados suficientemente informativos.

em que: ζ é uma constante usada para limitar o sinal de teste, $\Phi_v(\omega)$ é o espectro de perturbações e $\Lambda(\omega)$ é uma função de ponderação que reflete a aplicação do modelo.

Até onde se sabe, não há trabalhos que tratem de como projetar o espectro ótimo do sinal de excitação para a identificação de modelos de blocos interconectados, que operem em malha fechada. Entretanto, os resultados desenvolvidos para modelos lineares podem orientar o projeto do sinal de teste e, conseqüentemente, do seu espectro $\Phi_d(\omega)$ para o presente contexto. Por exemplo, de (4.1) pode-se dizer que:

- A densidade espectral de potência do sinal de teste deve concentrar-se na faixa de frequências em que as perturbações atuem com maior intensidade;
- Para que o modelo identificado atenda aos seus propósitos, o sinal de teste deve tentar reproduzir a excitação verificada em situações para as quais o modelo é utilizado, em outras palavras, $\Phi_d(\omega)$ deve ter mais potência nas frequências em que for necessário obter erros de modelagem menores.

Como praticamente todos os processos industriais possuem características de filtros passa-baixas, em geral, não há razões práticas para excitar o processo em altas frequências [80]. O GBN (do Inglês, *Generalized Binary Noise*) [77] é uma forma simples de gerar um sinal persistentemente excitante com espectro concentrado em baixas frequências. A cada intervalo de amostragem, este sinal comuta entre dois valores $\pm V_{GBN}$, de acordo com a seguinte regra de probabilidade:

$$\begin{aligned} P(d(k) = -d(k-1)) &= p_{sw} \\ P(d(k) = d(k-1)) &= 1 - p_{sw}, \end{aligned} \tag{4.2}$$

em que $P(\cdot)$ indica a probabilidade do evento $(\cdot)$ ocorrer. O tempo médio entre cada comutação $\bar{T}_{sw}$, em unidade de tempo, é dado pela razão do intervalo de amostragem T_s pela probabilidade de mudança de valor p_{sw} [89]:

$$\bar{T}_{sw} = \frac{T_s}{p_{sw}}. \tag{4.3}$$

A Figura 4.3 mostra realizações do sinal GBN composto por 800 amostras espaçadas de 1s e magnitude V_{GBN} unitária, para três valores de $\bar{T}_{sw}$. Note a forma como o espectro do sinal varia de acordo com a especificação de tempo médio em cada patamar.

Para guiar o projeto de sinais de excitação GBN, Tulleken [77] propôs uma faixa de valores sub-ótimos do tempo médio em cada patamar em função do tempo de acomodação t_s :

$$\bar{T}_{sw} = \frac{t_s}{\alpha} \quad (4.4)$$

onde $1 \leq \alpha \leq 2$, de modo que, quanto maior o valor de α , mais o modelo é focado em altas frequências.

A principal motivação para (4.4) vem do fato de os espectros de perturbações e do sinal de entrada no processo serem determinados pela dinâmica do processo [89]. Logo, o sinal de teste é projetado a partir de t_s que, quando não conhecido *a priori*, pode ser obtido através de ensaios preliminares. Para os casos em que o tempo de acomodação varie de acordo com o ponto de operação, recomenda-se obter

Figura 4.3: Densidade espectral de potência do GBN para diferentes $\bar{T}_{sw}$.

uma estimativa próxima do ponto de operação nominal do processo.

Como se sabe, sinais de entrada com maior potência e ensaios de maior duração contribuem para melhorar a estimativa dos parâmetros [9]. Todavia, por razões de segurança ou até econômicas, a magnitude do sinal de teste V_{GBN} e a duração N do experimento devem obedecer a certas restrições. Portanto, as escolhas de V_{GBN} e N devem ser consideradas caso a caso, para lidar com o conflito entre os requisitos teóricos e as restrições práticas.

Um algoritmo para gerar uma realização do GBN com N pontos é mostrado no fluxograma da Figura 4.4. Note que uma sequência de números independentes e uniformemente distribuídos entre $[0; 1]$, denominada R_i (para $i = 1, \dots, N$), é usada para determinar se deve ocorrer uma mudança de valor no sinal de teste $d(k)$.

Figura 4.4: Fluxograma do algoritmo para geração do sinal GBN.

4.3 Estimação dos parâmetros

O conjunto formado pelos modelos de atrito e do processo na Figura 4.2, pode ser visto como um modelo Hammerstein-Wiener (ver Figura 3.3(a) na seção 3.1). Logo, o problema de estimar os parâmetros do modelo de atrito e do processo, poderia ser tratado por um algoritmo de estimação de modelos de blocos interconectados, como por exemplo, o desenvolvido por Zhu [90], onde a identificação é realizada através de uma abordagem “caixa-preta”.

No entanto, neste caso, há conhecimento prévio a respeito da estrutura da não-linearidade que antecede o bloco dinâmico. Além disso, como os sinais entre os blocos não são conhecidos, a não ser que uma estrutura apropriada seja usada para representar o fenômeno de atrito, ganhos arbitrários seriam distribuídos entre o bloco linear e os não-lineares. Com isso, a quantificação do fenômeno de atrito na válvula seria, provavelmente, inconsistente.

4.3.1 Procedimento baseado em busca exaustiva

Suponha que, a princípio, os parâmetros (S, J) sejam conhecidos. Com isso, é possível obter uma estimativa da saída do modelo de atrito $\hat{u}(k)$, a partir do sinal de entrada $z(k)$:

$$\hat{u}(k) = \mathcal{N}_1(z(k), \hat{u}(k-1), S, J), \quad (4.5)$$

em que $\mathcal{N}_1(\cdot)$ é a transformação não-linear descrita no fluxograma da Figura 2.5.

Uma vez determinada a sequência $\hat{u}(k)$, o algoritmo de estimação de parâmetros do modelo de Wiener (descrito na seção 3.2) é executado usando o conjunto de dados $\{\hat{u}(k), y(k)\}_{k=1}^N$ para obter o modelo do processo. Todavia, o par (S, J) não é conhecido de antemão.

Uma forma de lidar com esta questão é através do seguinte algoritmo:

Algoritmo 4.1 *Estimação dos parâmetros do conjunto atrito + processo baseado em busca exaustiva.*

1. Criar um conjunto de valores candidatos $\mathcal{D}$ para o par (S, J) . Neste passo, deve-se ter em mente os seguintes aspectos: (1) O valor de J é limitado por S [21], pois na pior das hipóteses, $J = S$; (2) Uma condição necessária para que a estimação dos parâmetros de atrito seja viável, é que haja movimentação da haste da válvula. Admitindo-se que o sinal de teste tenha sido convenientemente escolhido, é razoável afirmar que a maior parte das inversões de sentido tenham sido causadas pelo sinal de teste. Logo, pode-se dizer que a amplitude máxima de $z(k)$ impõe um limite superior para o parâmetro S , ou seja:

$$\max(S) < \max(z(k)) - \min(z(k)). \quad (4.6)$$

Tais restrições resultam no lugar geométrico esboçado na Figura 4.5;

Figura 4.5: Lugar geométrico dos valores candidatos ao par (S, J) .

2. Escolher um par $(S_i, J_j) \in \mathcal{D}$;
3. Calcular a sequência de valores $\{\hat{u}_{ij}(k)\}_{k=1}^N$ a partir de (4.5);
4. Estimar os parâmetros dos modelos do processo e de perturbações, utilizando o algoritmo proposto na seção 3.2 (Figura 3.5);
5. Quantificar o erro de simulação livre através do critério:

$$\mathcal{C}_{ij}(S_i, J_j) = \frac{1}{N} \sum_{k=1}^N \left(y(k) - \hat{f} \left(\hat{G}(q) \hat{u}_{ij}(S_i, J_j, k) \right) \right)^2; \quad (4.7)$$

6. Até que todos os pares do conjunto $\mathcal{D}$ tenham sido testados, voltar ao passo (2);
7. Assim que todos os pares $(S_i, J_j) \in \mathcal{D}$ tenham sido testados, os parâmetros dos modelos de atrito (S^*, J^*) , do processo $(G(q), f(\cdot))$ e de perturbações $H(q)$ são os obtidos quando os índices i e j proporcionam o menor valor de $\mathcal{C}_{ij}$.

Este algoritmo é baseado no conceito de mínimos quadrados separáveis [32], pois o problema original é separado em estimativas de modelos de Wiener usando sequências $\{\hat{u}_{ij}(k), y(k)\}_{k=1}^N$, geradas por cada $(S_i, J_j) \in \mathcal{D}$ e uma busca pelos parâmetros do modelo de atrito que proporcionem o menor erro de simulação livre, quantificado através de $\mathcal{C}_{ij}$.

4.3.2 Estimação dos parâmetros usando otimização

Um aspecto importante em relação ao algoritmo 4.1 é a questão de mínimos locais. Para evitá-los, deve-se escolher uma resolução suficientemente pequena para os pares (S_i, J_j) . Por outro lado, diminuir a resolução dos parâmetros candidatos implica no aumento de elementos no conjunto $\mathcal{D}$, e conseqüentemente, no esforço computacional.

Uma forma de diminuir o esforço computacional é utilizar algum algoritmo de otimização não-linear para buscar (S^*, J^*) . Entretanto, a não ser que o algoritmo seja inicializado próximo do mínimo global, o problema de se obter um mínimo local ainda persiste. Recentemente, ao tratar da quantificação do atrito em válvulas empregando mínimos quadrados separáveis e algoritmos de busca, Jelali [45] empregou busca direta em padrões (do Inglês, *pattern search*). Neste caso, parâmetros estimados por métodos de terceiros [22] foram usados como condição inicial.

Resultados práticos [39, 70, 79] têm indicado que a aproximação do atrito em válvulas de controle por um bloco de histerese é bem razoável. Isto se deve ao fato de a histerese ser proeminente em relação a outros comportamentos não-lineares nas

válvulas [48]. Da definição do modelo de atrito de Kano (subseção 2.2.3), pode-se considerar apenas o parâmetro S como uma aproximação da histerese provocada pelo atrito, pois mudanças no sentido de movimentação da haste ocorrem apenas se o sinal de controle manifestar reversões de magnitude S .

Diante dessas premissas, ao desprezar o parâmetro J ($J = 0$), pode-se buscar o valor de S que minimize o erro de simulação (4.7) e, em seguida, usar este valor para determinar a condição inicial para um algoritmo de busca para o par (S, J) . Esta estratégia é descrita no algoritmo 4.2.

Algoritmo 4.2 *Estimação dos parâmetros dos modelos de atrito e de Wiener empregando otimização na busca do par (S, J) .*

1. Inicialmente, admitir $J = 0$ e criar um vetor $\mathcal{D}_S$ de valores candidados a S :

$$\mathcal{D}_S = \left(0, \Delta S, 2\Delta S, \dots, S_{\max}\right), \quad (4.8)$$

tal que $S_{\max} \leq \max(S)$;

2. Para cada $S_i \in \mathcal{D}_S$, calcular o erro de simulação livre $\mathcal{C}_i$, de forma análoga a (4.7);
3. Selecionar o valor de S_i que minimize o erro de simulação livre:

$$S^\circ = \arg \min_{S_i} \mathcal{C}_i(S_i); \quad (4.9)$$

4. Partindo da estimativa inicial: $(S_0, J_0) = (S^\circ + \gamma \cdot \Delta S, \gamma \cdot \Delta S)$, resolver o problema de otimização:

$$(S^*, J^*) = \arg \min_{(S, J)} \frac{1}{N} \sum_{k=1}^N \left(y(k) - \hat{f} \left(\hat{G}(q) \hat{u}(S, J, k) \right) \right)^2, \quad (4.10)$$

em que γ é um número inteiro positivo ($\gamma \in \mathbb{Z}^+$) usado como parâmetro de ajuste da condição inicial de (4.10), cujo valor é discutido adiante;

5. Estimar os parâmetros do modelo de Wiener com a sequência $\{\hat{u}^*(k), y(k)\}_{k=1}^N$, em que $\hat{u}^*(k)$ é a saída do modelo de atrito parametrizado por (S^*, J^*) .

Uma vez descrito o algoritmo 4.2, algumas questões relacionadas à otimização no passo (4) são tratadas a seguir.

Algoritmo de otimização e critérios de parada

O caráter descontínuo do modelo de atrito (ver Figura 2.4), faz com que o problema em (4.10) seja não-convexo. Portanto, algoritmos de otimização baseados em gradientes, classificados como métodos de busca local, não devem ser empregados na obtenção de (S^*, J^*) .

O algoritmo Simplex de Nelder-Mead [54, 84] é apropriado para problemas de identificação de atrito, pois não requer o cálculo numérico ou analítico de gradientes. No entanto, quando aplicado a problemas com múltiplos mínimos locais, não há garantias de que o valor obtido seja o mínimo global. Ao inicializar o algoritmo próximo da solução ótima, é possível evitar esse inconveniente.

Antes de tratar da inicialização do algoritmo, a questão de convergência é abordada. Neste caso, foram adotados dois critérios de parada: (1) quando o diâmetro do simplex[†] for menor que 10^{-3} , ou (2) quando o algoritmo atingir 20 iterações (passos de busca). Estes critérios levam em consideração mudanças no par (S, J) nas últimas iterações e o esforço computacional.

Estimativa inicial

A condição inicial $(S^\circ + \gamma \cdot \Delta S, \gamma \cdot \Delta S)$ é baseada na expressão (2.7) que é reescrita com a notação usada na Figura 4.2:

$$u(k) = z(k) - \frac{d_{aux}(S - J)}{2}. \quad (4.11)$$

[†]O diâmetro de um simplex é determinado pela maior distância entre dois vértices. Como o problema de otimização (4.10) é de dimensão 2, o simplex possui três vértices, cada qual representando um par (S, J) .

Ao desprezar J nos passos (1) - (3), pode-se considerar que $S^\circ \pm \Delta S \approx S - J$, que implica em:

$$S^\circ - \Delta S < S - J < S^\circ + \Delta S. \quad (4.12)$$

Para acelerar a convergência ao resolver (4.10) a escolha da condição inicial (S_0, J_0) é orientada por:

- A área em destaque na Figura 4.6 representa o espaço de busca gerado a partir das restrições $\max(J) = S$ e (4.12). Note que o espaço de valores candidatos é proporcional a ΔS , portanto, é desejável selecioná-lo de modo que $\Delta S \ll S$;
- Conforme dito anteriormente, em geral, a contribuição devido à histerese é mais influente do que o fenômeno de escorregamento (*slip-jump*) em válvulas de controle, logo, $S \gg J$.

Figura 4.6: Condição inicial para otimização do algoritmo 4.2 e lugar geométrico do par (S, J) sujeito a (4.12).

Sob tais argumentos, ao adotar J_0 como um múltiplo de ΔS e de (4.12), resulta:

$$(S_0, J_0) = (S^\circ + \gamma \cdot \Delta S, \gamma \cdot \Delta S). \quad (4.13)$$

A condição inicial para $\gamma = 1$ é indicada por “ $\odot$ ” na Figura 4.6. A linha pontilhada a partir de S° indica os pares (S, J) candidatos quando $\Delta S \rightarrow 0$.

Na prática, como geralmente $S \gg J$, é conveniente testar valores crescentes para γ . A cada valor testado, o desempenho do modelo obtido é verificado através de critérios pré-estabelecidos. As técnicas empregadas neste trabalho para avaliar os modelos estimados são discutidas na próxima seção.

4.4 Técnicas de validação do modelo

A etapa de validação é encarregada de verificar se o modelo estimado é apropriado para uma determinada aplicação. Para realizar esta verificação, pode-se considerar as seguintes questões [58]: (1) O modelo é “bom” o bastante para atender aos seus propósitos? (2) As simulações estão de acordo com os dados observados? (3) O modelo descreve o sistema real?

Geralmente, o método para responder a estas questões é confrontar o modelo estimado com informações do sistema real. Tais informações incluem conhecimento prévio do processo, dados experimentais e experiência adquirida ao utilizar o modelo. Em identificação de sistemas, a entidade mais natural para confrontar os modelos são os dados experimentais, portanto muitas vezes a etapa de validação se restringe a responder a questão (2).

Na identificação de processos industriais, perturbações não mensuráveis e a inexistência de estruturas ideais para reproduzir certos fenômenos, via de regra, geram erros de modelagem. Portanto, além do modelo que melhor descreva um conjunto de dados específico, uma descrição adequada dos erros de modelagem pode ser usada para verificar se tais erros são significativos para a aplicação do modelo.

4.4.1 Validação por simulações

A forma mais usual de se validar um modelo é através da comparação entre a simulação do modelo obtido com dados medidos. Entretanto, um cuidado básico precisa ser observado: ao realizar tal comparação não é aconselhável utilizar os

mesmos dados empregados na estimação dos parâmetros para validar o modelo. Não é surpreendente que um modelo possa ter um bom desempenho quando avaliado com o próprio conjunto de dados com que foi estimado. O teste real é verificar se ele é capaz de descrever o comportamento do processo com base em um conjunto novo de dados.

Portanto, na prática, o ideal é efetuar dois testes independentes, gerando-se, assim, dois conjuntos de dados. Um deles é usado na estimação dos parâmetros e outro na validação do modelo. Se dois testes distintos não puderem ser realizados, então se divide o único conjunto disponível em duas partes (não necessariamente iguais), sendo cada uma das partes usada numa etapa diferente [1].

Uma forma de quantificar a adequação das simulações em relação aos dados reais é através do índice de acerto $\mathcal{F}_2$ [58], expresso em porcentagem:

$$\mathcal{F}_2[\%] = \left(1 - \frac{\|\hat{Y} - Y\|_2}{\|Y - \bar{Y}\|_2} \right) \times 100, \quad (4.14)$$

em que $Y \in \mathbb{R}^N$ é o vetor de dados de saída, enquanto $\hat{Y}$ e $\bar{Y}$ denotam, respectivamente, o vetor da saída simulado a partir de $\{z(k)\}_{k=1}^N$ e a média de Y .

4.4.2 Validação baseada nos erros de modelagem

Os testes de simulação não indicam até que ponto eventuais falhas sejam críticas, ou mesmo se podem ser corrigidas. Por exemplo, se as simulações apresentarem erros significativos em relação aos dados observados, não é possível afirmar se tais erros são significativos para a aplicação do modelo. Neste caso, o propósito do modelo do processo é ser aplicado na re-sintonia de controladores, portanto, é conveniente obter uma descrição dos erros de modelagem no domínio da frequência.

A seguir é proposto um método para estimar limites superiores do erro de modelagem no bloco dinâmico linear. Esta metodologia, baseada no comportamento assintótico da estimativa de parâmetros [57, 58], além de indicar as frequências em

que o desempenho do modelo seja pior, também provê informações que podem ser usadas no re-projeto do sinal de teste e assim diminuir o erro de modelagem em frequências específicas.

Comportamento assintótico dos parâmetros estimados

Suponha um sistema estável em malha aberta, linear e invariante no tempo, cuja saída $w(k)$ possa ser expressa através de:

$$\begin{aligned} w(k) &= \sum_{i=1}^{\infty} g_i^0 \cdot u(k-i) + \sum_{i=0}^{\infty} h_i^0 \cdot e(k-i) \\ &= G_0(q)u(k) + H_0(q)e(k), \end{aligned} \quad (4.15)$$

em que $e(k)$ é ruído branco de média nula e variância σ_e^2 e os coeficientes da resposta impulsiva do processo e das perturbações são, respectivamente, g_i^0 e h_i^0 .

Ljung [57] estudou o comportamento assintótico da estimativa de modelos, obtidos através do método dos mínimos quadrados, nas seguintes condições:

$$n \rightarrow \infty, \text{ quando } N \rightarrow \infty$$

$$n^2/N \rightarrow 0, \text{ à medida que } N \rightarrow \infty$$

que expressam formalmente a condição de que, mesmo para valores assintóticos, a ordem do modelo n é pequena se comparada à quantidade de amostras N . Em tais condições, foram obtidos os seguintes resultados teóricos:

1. Tanto a estimativa da resposta em frequência do processo $\hat{G}(e^{j\omega})$, quanto a das perturbações $\hat{H}(e^{j\omega})$ são consistentes, isto é:

$$\begin{pmatrix} \hat{G}(e^{j\omega}) \\ \hat{H}(e^{j\omega}) \end{pmatrix} \rightarrow \begin{pmatrix} G_0(e^{j\omega}) \\ H_0(e^{j\omega}) \end{pmatrix} \quad (4.16)$$

com probabilidade 1;

2. O erro da resposta em frequência estimada segue assintoticamente uma distribuição de probabilidade gaussiana, com média nula:

$$\sqrt{\frac{N}{n}} \begin{pmatrix} \hat{G}(e^{j\omega}) - E(\hat{G}(e^{j\omega})) \\ \hat{H}(e^{j\omega}) - E(\hat{H}(e^{j\omega})) \end{pmatrix} \rightarrow \mathcal{G}(0, P(\omega)). \quad (4.17)$$

$E(\cdot)$ denota o operador esperança matemática. O valor assintótico da matriz de covariância $P(\omega)$ é dado por:

$$P(\omega) = \sigma_e^2 |H_0(e^{j\omega})|^2 \begin{pmatrix} \Phi_u(\omega) & \Phi_{eu}(\omega) \\ \Phi_{ue}(\omega) & \sigma_e^2 \end{pmatrix}^{-1} \quad (4.18)$$

em que $\Phi_u(\omega)$ é a densidade espectral do sinal de entrada, assim como $\Phi_{eu}(\omega)$ e $\Phi_{ue}(\omega)$ são as densidades do espectro cruzado de $e(k)$ e $u(k)$.

As expressões (4.16) e (4.17) são válidas quando $N \rightarrow \infty$ e $n \rightarrow \infty$, para dados obtidos em malha fechada (a situação de malha aberta pode ser vista como um caso especial), desde que o algoritmo de estimação de parâmetros não resulte em um mínimo local [57, 89].

Como a resposta em frequência do modelo estimado, $\hat{G}(e^{j\omega})$, depende dos dados experimentais, faz sentido analisar o erro do modelo através de uma abordagem probabilística [89]. Para tanto, definem-se o *bias* $\mathcal{B}(\omega)$ e a variância $\mathcal{V}(\omega)$ referente ao modelo do processo no domínio da frequência:

$$\mathcal{B}(\omega) = G_0(e^{j\omega}) - E(\hat{G}(e^{j\omega})) \quad (4.19)$$

$$\mathcal{V}(\omega) = E\left(\hat{G}(e^{j\omega}) - E(\hat{G}(e^{j\omega}))\right)^2. \quad (4.20)$$

Levando-se em conta (4.19) e (4.20), é possível representar o erro médio quadrático[‡] do modelo estimado através de:

$$\begin{aligned} MSE(\omega) &\triangleq E\left(G_0(e^{j\omega}) - \hat{G}(e^{j\omega})\right)^2 \\ &= E\left(G_0(e^{j\omega}) - E(\hat{G}(e^{j\omega})) + E(\hat{G}(e^{j\omega})) - \hat{G}(e^{j\omega})\right)^2 \\ \therefore MSE(\omega) &= \mathcal{B}^2(\omega) + \mathcal{V}(\omega). \end{aligned} \quad (4.21)$$

De acordo com Zhu [88], o *bias* está relacionado à incapacidade de um modelo de representar comportamentos estáticos e dinâmicos encontrados nos dados. Todavia, uma consequência importante de (4.16) é que $\mathcal{B}(\omega) \rightarrow 0$, conforme $n \rightarrow \infty$.

[‡] MSE , do Inglês *Mean Square Error*.

Em contrapartida, a contribuição do termo de variância no erro médio quadrático é causada por perturbações não modeladas. De (4.18) pode-se escrever:

$$P(\omega) \asymp \sigma_e^2 |H_0(e^{j\omega})|^2 \frac{\text{adj}(\Phi(\omega))}{\det(\Phi(\omega))},$$

sendo:

$$\Phi(\omega) = \begin{pmatrix} \Phi_u(\omega) & \Phi_{eu}(\omega) \\ \Phi_{ue}(\omega) & \sigma_e^2 \end{pmatrix}.$$

Logo,

$$P(\omega) \asymp \frac{\sigma_e^2 |H_0(e^{j\omega})|^2}{\Phi_u(\omega)\sigma_e^2 - |\Phi_{ue}(\omega)|^2} \begin{pmatrix} \sigma_e^2 & -\Phi_{ue}(\omega) \\ -\Phi_{eu}(\omega) & \Phi_u(\omega) \end{pmatrix}, \quad (4.22)$$

em que $\asymp$ simboliza assintoticamente igual; $\det(\cdot)$ e $\text{adj}(\cdot)$ denotam a matriz adjunta e o determinante, respectivamente. Ao considerar (4.17) e apenas o elemento referente à variância do modelo do processo em (4.22), resulta que:

$$\mathcal{V}(\omega) \asymp \frac{n}{N} \frac{(\sigma_e^2)^2 |H_0(e^{j\omega})|^2}{\Phi_u(\omega)\sigma_e^2 - |\Phi_{ue}(\omega)|^2} = \frac{n}{N} \frac{\Phi_v(\omega)\sigma_e^2}{\Phi_u(\omega)\sigma_e^2 - |\Phi_{ue}(\omega)|^2}, \quad (4.23)$$

em que $\Phi_v(\omega)$ denota a densidade espectral das perturbações: $v(k) = H_0(q)e(k)$.

A expressão (4.23) revela que a variância do modelo em uma dada frequência é proporcional à razão da ordem do modelo pela quantidade de amostras. Além disso, quanto maior a potência do sinal de entrada, menor será a variância do modelo, ao passo que o inverso ocorre para o espectro de perturbações.

Validação do bloco linear

No algoritmo proposto na seção 4.3 utilizou-se uma estrutura capaz de representar tanto o fenômeno de atrito quanto um processo com dinâmica não-linear. Ademais, como um modelo FIR de ordem suficientemente elevada foi usado durante a estimativa do bloco linear (passo 4 desse algoritmo), espera-se que o erro de modelagem devido à polarização (*bias*) seja desprezível, se comparado ao termo que expressa a variância do modelo. Logo, com probabilidade de 99,7%:

$$\left| G_0(e^{j\omega}) - \hat{G}(e^{j\omega}) \right| \leq 3\sqrt{\mathcal{V}(\omega)} = \Delta(\omega). \quad (4.24)$$

Voltando à estrutura da Figura 4.2, se for admitido que o modelo de atrito e a função não-linear $w(k) = f^{-1}(y(k))$ sejam perfeitamente conhecidos, a partir de (4.23), o limite superior $\Delta(\omega)$ para o erro de modelagem do bloco linear é dado por:

$$\Delta(\omega) = 3\sqrt{\frac{n}{N} \frac{\Phi_v(\omega)\sigma_e^2}{\Phi_u(\omega)\sigma_e^2 - |\Phi_{ue}(\omega)|^2}}. \quad (4.25)$$

Na prática, $\Phi_v(\omega)$, $\Phi_u(\omega)$, $\Phi_{ue}(\omega)$ e σ_e^2 são substituídos pelas estimativas:

$$\hat{\Phi}_v(\omega) = \sum_{\tau=-n}^n \left(\frac{1}{N} \sum_{k=1}^N \hat{v}(k)\hat{v}(k+\tau) \right) e^{-j\omega\tau} \quad (4.26)$$

$$\hat{\Phi}_u(\omega) = \sum_{\tau=-n}^n \left(\frac{1}{N} \sum_{k=1}^N \hat{u}(k)\hat{u}(k+\tau) \right) e^{-j\omega\tau} \quad (4.27)$$

$$\hat{\Phi}_{ue}(\omega) = \sum_{\tau=-n}^n \left(\frac{1}{N} \sum_{k=1}^N \hat{u}(k)\varepsilon(k+\tau) \right) e^{-j\omega\tau} \quad (4.28)$$

$$\hat{\sigma}_e^2 = \text{var}(\varepsilon(k)), \quad (4.29)$$

em que $\hat{u}(k)$ é dado por (4.5). Ademais, a estimativa do sinal de perturbações $\hat{v}(k)$ e o erro de predição $\varepsilon(k)$ são dados por:

$$\hat{v}(k) = \hat{f}^{-1}(y(k)) - \hat{G}(q)\hat{u}(k)$$

$$\varepsilon(k) = \hat{H}^{-1}(q)\hat{v}(k).$$

Zhu [88] utilizou a relação entre o erro de modelagem $\Delta(\omega)$ e o módulo da estimativa da resposta em frequência para classificar a qualidade do modelo, de acordo com o critério da Tabela 4.1. Por exemplo: seja um intervalo de frequências importantes para a aplicação do modelo delimitado por $[\omega_1; \omega_2]$. Se $\Delta(\omega)$ proporcionar erros inferiores a 30% do módulo da resposta em frequência de $\hat{G}(q)$, para $\omega \in [\omega_1; \omega_2]$, então este modelo é classificado como “bom”.

De acordo com simulações e experiências práticas, modelos classificados como *A* ou *B* são considerados apropriados para aplicações em controle de processos [88]. Por outro lado, modelos *C* e *D* demandam um re-projeto do ensaio de coleta de dados. Conforme (4.25), para obter um modelo com desempenho melhor há duas

Tabela 4.1: Classificação do modelo de acordo com a relação entre $\Delta(\omega)$ e $|\hat{G}(e^{j\omega})|$.

	Classificação	Limite superior do erro de modelagem
<i>A</i>	Muito bom	$\Delta(\omega) \leq 0,2 \hat{G}(e^{j\omega}) $
<i>B</i>	Bom	$0,2 \hat{G}(e^{j\omega}) < \Delta(\omega) \leq 0,5 \hat{G}(e^{j\omega}) $
<i>C</i>	Marginal	$0,5 \hat{G}(e^{j\omega}) < \Delta(\omega) \leq 0,9 \hat{G}(e^{j\omega}) $
<i>D</i>	Pobre	$\Delta(\omega) > 0,9 \hat{G}(e^{j\omega}) $

possibilidades: (i) realizar um ensaio de maior duração ou (ii) escolher um sinal $d(k)$ de modo que $\Phi_u(\omega)$ seja maior no intervalo $[\omega_1; \omega_2]$.

Durante o cálculo do limite superior do erro de modelagem, admitiu-se que os blocos não-lineares que envolvem o bloco dinâmico foram perfeitamente modelados. Uma forma de levar em conta os erros de tais blocos é majorar (4.24) com um fator maior do que 3 desvios padrão.

4.5 Fechamento do capítulo

A ideia de usar valores candidatos para inferir a saída do bloco que modela o atrito e, a partir deste sinal, identificar um modelo de processo, foi originalmente proposta por Srinivasan *et al.* [73]. Em relação a este trabalho, foram sugeridas as seguintes extensões:

- Modelo de atrito mais apropriado;
- Estrutura não-linear para o modelo do processo;
- Adição de um sinal extra $d(k)$ para garantir ensaios informativos, em malha fechada.

Paralelamente ao desenvolvimento desta pesquisa, Choudhury *et al.* [23] entenderam a proposta de Srinivasan *et al.* [73] utilizando um modelo de atrito mais

apropriado (também composto por dois parâmetros), similar ao modelo de Kano. Portanto, esta extensão pode ser considerada como um primeiro passo rumo às contribuições sugeridas nesta Tese.

Neste capítulo estudou-se como integrar as ideias de quantificação do atrito (capítulo 2) e de identificação de modelos de processos não-lineares (capítulo 3). Da primeira resultou a adoção de uma descrição apropriada para o atrito e nas restrições do espaço de parâmetros candidatos (S, J) . Por outro lado, as ideias relacionadas à identificação de modelos de Wiener foram trabalhadas, de modo que fosse possível obter um modelo de processo não-linear, sem nenhuma intervenção do usuário. Tal característica permite que S e J sejam determinados por busca exaustiva ou por otimização, de modo a minimizar a diferença entre a saída $y(k)$ e a estimativa $\hat{y}(k)$, calculada através do modelo do conjunto atrito + processo. Por fim, foi descrita uma forma de estimar um limite superior para o erro de modelagem, usando resultados da teoria assintótica [57].

Capítulo 5

Resultados e discussões

Neste capítulo pretende-se avaliar o desempenho da metodologia para identificação conjunta dos modelos de atrito e processo, apresentada no capítulo anterior. Primeiramente, o método é aplicado a um ambiente totalmente simulado (seção 5.1) com o intuito de analisar a influência de perturbações na estimação dos parâmetros. Na seção 5.2, utiliza-se uma plataforma híbrida, em que uma válvula de controle real é “interligada” ao simulador de uma planta de neutralização de pH. Esta plataforma é usada para testar a capacidade da estrutura proposta (modelo de atrito + Wiener) de modelar um processo altamente não-linear. Além disso, questões relacionadas à influência de perturbações, da magnitude do sinal de teste e da sintonia do controlador nos modelos identificados também são abordadas.

5.1 Identificação de uma planta simulada

Com o intuito de avaliar a exatidão das estimativas dos parâmetros dos modelos de atrito, do processo e do limite superior do erro de modelagem, obtidas através do método proposto neste trabalho, considera-se uma malha simulada cujo diagrama de blocos é mostrado na Figura 5.1.

O bloco que representa a parte dinâmica do modelo de Wiener é uma função de transferência em tempo contínuo, simulada usando um intervalo de integração de 10 ms. O sinal de controle $u(k)$ é reconstruído por um segurador de ordem zero,

Figura 5.1: Diagrama de blocos da malha simulada.

denotado por bloco S/H. A saída de $G(s)$ é amostrada com um período $T_s = 0,4s$, antes de ser somada a $v(k)$, que representa as perturbações do processo. O sinal $e(k)$ denota um ruído branco gaussiano de média nula e variância unitária. Mais especificamente, os valores numéricos dos blocos usados na simulação são:

$$G(s) = \frac{5}{s^3 + 11,5s^2 + 15,5s + 5} \quad (5.1)$$

$$H(q) = \frac{\rho_v}{(1 - 0,85q^{-1})(1 - 0,95q^{-1})} \quad (5.2)$$

$$y(k) = f(w(k)) = \frac{w(k)}{\sqrt{0,1 + 0,9w^2(k)}} \quad (5.3)$$

$$K(q) = \frac{1,25 - 1,094q^{-1}}{1 - q^{-1}}, \quad (5.4)$$

em que ρ_v é uma constante usada para ajustar a intensidade da perturbação $v(k)$.

O gráfico de (5.3) é mostrado na Figura 5.2.

Figura 5.2: Gráfico da função estática não-linear $y(k) = f(w(k))$ da planta simulada.

Os parâmetros do modelo de atrito são: $S = 10\%$ e $J = 2\%$. Além de terem sido empregados anteriormente [23] para simular o atrito em válvulas de controle, tais valores são coerentes com o atrito observado em situações reais [79].

5.1.1 Experimento para coleta de dados

Um sinal GBN com magnitude comutante entre $\pm 3,75 \times 10^{-2}$ e tempo médio em cada patamar de $\bar{T}_{sw} = 18 \times T_s$ é introduzido em $d(k)$. Conforme discutido na seção 4.2, o espectro de $d(k)$ é definido ao adotar $\alpha \approx 1,3$ em (4.4). Como o sinal de referência $r(k)$ é mantido constante em $\bar{r} = 0,75$, esta magnitude corresponde a uma variação de $\pm 5\%$ de $\bar{r}$.

As simulações foram realizadas em duas situações distintas, de acordo com a intensidade do sinal de perturbação $v(k)$. Na primeira, o valor de ρ_v é ajustado de modo a obter uma relação sinal-ruído de 2,45% (de variância) em $w(k)$ e 2,62% em $y(k)$. Em um segundo cenário, é utilizado um nível de perturbação de 9,43% em $w(k)$, que implica em uma relação de 9,88% na saída.

Para cada cenário de perturbação foram realizadas duas simulações com duração de 400s (1000 pontos). Os dados da primeira simulação são empregados na estimação dos parâmetros, ao passo que os da segunda são destinados à validação dos modelos. Os dados de entrada e saída do processo, $z(k)$ e $y(k)$, referentes às situações de baixo e alto nível de perturbação são mostrados*, respectivamente, nas Figuras 5.3 e 5.4. Apesar de não ser usado durante o procedimento de identificação, a saída do bloco parametrizado por S e J , $u(k)$, também é apresentada para que a contribuição do fenômeno de atrito possa ser explicitamente visualizada.

Das Figuras 5.3 e 5.4, verifica-se que na maior parte do tempo, a saída do controlador $z(k)$ excursiona dentro do intervalo $[0,25; 0,45]$. Por inspeção visual, percebe-se que a saída $y(k)$ é afetada por variações da entrada $z(k)$. Portanto, é razoável admitir que $\max(S) = 20\%$.

*Para melhor visualização, são exibidos apenas os ensaios usados na etapa de estimação.

Figura 5.3: Dados de entrada e saída da simulação com perturbações de baixa intensidade.

Figura 5.4: Dados da simulação para a estimação de parâmetros referentes ao cenário de perturbação elevada.

5.1.2 Estimação dos parâmetros

O procedimento para estimação de parâmetros do conjunto atrito + processo baseado em busca exaustiva (algoritmo 4.1) é aplicado aos conjuntos de dados provenientes de ambos cenários de perturbação. Uma resolução de 0,01 (1%) é empregada na geração de valores candidatos para o par (S, J) . Na identificação do modelo de Wiener, limitou-se a ordem do bloco dinâmico e o grau do polinômio seccionado em $l_{max} = 5$ e $m_{max} = 9$, respectivamente. A estimativa inicial do bloco dinâmico linear é obtida através de uma estrutura FIR de ordem $n = 40$ e o modelo de perturbações é parametrizado arbitrariamente com $n_c = 2$.

O comportamento do erro de simulação $\mathcal{C}_{ij}$, definido em (4.7) na seção 4.3, em função dos valores candidatos do modelo de atrito é mostrado nas curvas de nível da Figura 5.5. As superfícies mais escuras denotam os parâmetros de atrito que proporcionam os menores erros de simulação. O par $(\hat{S}, \hat{J})$ que minimiza $\mathcal{C}_{ij}$ é indicado por “ $\otimes$ ”. Note que, para ambos cenários de perturbação, as estimativas dos parâmetros do modelo de atrito são exatamente os empregados na simulação, ou seja, $S = 10\%$ e $J = 2\%$. Naturalmente, este resultado foi possível porque S e J pertencem ao conjunto de parâmetros admissíveis.

Figura 5.5: Curvas de nível do comportamento do erro de simulação $\mathcal{C}_{ij}$, nos cenários de menor e maior perturbação.

Também pode-se perceber na Figura 5.5 que os menores valores de $\mathcal{C}_{ij}$ são obtidos quando a diferença $S_i - J_j$ aproxima-se do valor usado na simulação: 0,08 (8%), desde que $J_j \leq 5\%$. Este resultado é explicado pelo fato de a diferença entre S e J determinar o comportamento dinâmico da posição da haste. Além disso, ao admitir um valor excessivamente elevado para J_j e mantendo constante a diferença $S_i - J_j$, o modelo de atrito pode deixar de apresentar algumas reversões de sentido presentes nos dados experimentais.

A seguir, o algoritmo de estimação baseado em otimização do par (S, J) , proposto na subseção 4.3.2, foi testado com o mesmo conjunto de dados (Figuras 5.3 e 5.4). Para a construção do vetor $\mathcal{D}_S$ (4.8), adota-se $\Delta S = 0,01$, assim como valores de 1 a 3 são atribuídos a γ para o cálculo da estimativa inicial (4.13). O desempenho dos modelos obtidos é quantificado pelo critério do erro de simulação $\mathcal{F}_2$, definido em (4.14).

A Tabela 5.1 mostra que os valores de $\hat{S}$ e $\hat{J}$ dependem da estimativa inicial a partir da qual tais parâmetros são calculados. No cenário de menor perturbação, por exemplo, ao admitir $\gamma = 1$ o algoritmo de otimização resulta em um mínimo local. Por outro lado, quando outras estimativas iniciais são testadas e os modelos obtidos são avaliados através de $\mathcal{F}_2$, o algoritmo de estimação baseado em otimização (com $\gamma = 2$) proporciona resultados equivalentes à estratégia de busca exaustiva.

Tabela 5.1: Comparativo do desempenho dos algoritmos de estimação de parâmetros.

Nível de perturbação	Algoritmo	$\hat{S}$ [%]	$\hat{J}$ [%]	$\mathcal{F}_2$ [%]	Pares (S, J) testados
Menor	Busca exaustiva	10	2	81,80	231
	Otimização ($\gamma = 1$)	9,23	0,99	80,01	35(21 + 14)
	Otimização ($\gamma = 2$)	9,97	1,88	81,70	38(21 + 17)
	Otimização ($\gamma = 3$)	9,92	1,84	81,63	51(21 + 30)
Maior	Busca exaustiva	10	2	65,21	231
	Otimização ($\gamma = 1$)	10,48	1,86	63,07	51(21 + 30)
	Otimização ($\gamma = 2$)	10,36	2,09	64,53	43(21 + 22)
	Otimização ($\gamma = 3$)	10,95	2,68	60,24	41(21 + 20)

Há ainda uma importante vantagem em se empregar otimização na busca pelo par (S, J) : o esforço computacional é drasticamente reduzido. Para este problema, enquanto a abordagem de busca exaustiva requer que 231 combinações[†] de modelo de atrito sejam testadas, durante a otimização foram testados, na pior das hipóteses, 51 pares (21 para o cálculo da estimativa inicial e 30 no algoritmo Simplex). Ao considerar os 93 pares testados com $\gamma = 1, 2$ e 3 , no cenário de maior perturbação, a redução é de quase 60%.

Outro aspecto relevante é que, de uma forma geral, os modelos obtidos no cenário de menor perturbação apresentam erros de simulação menores. Como mostrado na Tabela 5.1, o par $(\hat{S}, \hat{J})$ está próximo do valor verdadeiro, a despeito do nível de perturbação. Logo, a estimativa do modelo do processo pode estar comprometida. De qualquer forma, os dados de validação estão contaminados por ruído de processo de intensidade equivalente ao presente nos dados de estimação. Este fato explicaria, em parte, os menores valores de $\mathcal{F}_2$.

Neste caso, o modelo do processo (5.1)-(5.3) é perfeitamente conhecido. Portanto, é possível investigar a influência do nível de perturbação nas estimativas dos blocos $G(q)$ e $f(\cdot)$. Diante da vantagem relacionada ao esforço computacional, apenas o modelo do conjunto atrito + processo estimado por otimização (algoritmo 4.2) é considerado nas análises subsequentes.

5.1.3 Validação dos modelos

Para verificar se o bloco linear é capaz de incorporar a dinâmica do processo, a curva de Nyquist do modelo estimado, em cada situação de perturbação, é comparada com a proporcionada por $G(s)$, especificada em (5.1), após ser discretizada usando um segurador de ordem zero. A Figura 5.6 mostra que, para ambos níveis de perturbação, a resposta em frequência é bem reproduzida em baixas frequências. À

[†]No exemplo desta seção, admitiu-se $\max(S) = 20\%$ e uma resolução de 1% , portanto a quantidade de combinações é dada por: $\sum_{i=0}^{20} (i + 1) = 231$.

medida que a frequência aumenta, surgem divergências entre as curvas de Nyquist. Vale ressaltar que, como é de se esperar, para o cenário de maior perturbação, a diferença entre as curvas verdadeira e estimada é mais expressiva.

Figura 5.6: Gráfico de Nyquist do bloco dinâmico linear, em ambas as situações de perturbação consideradas.

A curva de Nyquist da estimativa inicial do bloco dinâmico também é mostrada na Figura 5.6 (traço-ponto). Ao comparar tais curvas com as obtidas dos modelos reduzidos $\hat{G}(q)$ (linha tracejada), fica clara a melhoria em reproduzir a resposta em frequência de $G(s)$, depois de o modelo FIR ser transformado em uma estrutura de função de transferência racional.

Uma vez reveladas divergências entre a resposta em frequência dos modelos estimado e simulado, como verificar se estas discrepâncias são erros de modelagem significativos? A Figura 5.7 apresenta o módulo da resposta em frequência do modelo real $G(s)$ e do estimado $\hat{G}(q)$, assim como o erro de modelagem e o limite superior do erro $\Delta(\omega)$ (calculado conforme descrito na subseção 4.4.2), para a condição de menor perturbação.

Primeiramente, verifica-se que o erro de modelagem só atinge uma magnitude superior a 50% de $|G(e^{j\omega})|$ próximo da frequência de Nyquist. Este resultado mostra que os erros apresentados não são relevantes. A Figura 5.7 mostra ainda que a estimativa do limite superior do erro de modelagem é coerente com os erros observados

em todo o espectro de frequências.

Figura 5.7: Módulo da resposta em frequência e limite superior do erro de modelagem provenientes do ensaio com baixo nível de perturbação.

Para classificar $\hat{G}(q)$ de acordo com o limite superior do erro de modelagem (Tabela 4.1), é necessário estipular uma faixa de frequências que reflita a aplicação do modelo. No contexto da identificação de sistemas lineares, quando o objetivo é aplicar o modelo na sintonia de controladores, essa faixa de frequências é determinada pela banda de passagem da malha fechada [59]. No entanto, como não há especificações prévias para a sintonia do controlador e por este trabalho tratar da identificação de processos não-lineares, considera-se a faixa $[0; \omega_c]^\dagger$, sendo que ω_c é a frequência na qual $|G(e^{j\omega})| = 0,707$ (-3dB). Neste caso, $\hat{G}(q)$ seria classificado como *A* (“Muito bom”), pois: $\Delta(\omega) < 0,2|\hat{G}(e^{j\omega})|$, para $\omega \in [0; \omega_c]$.

A análise do erro de modelagem no domínio da frequência também é conduzida para o modelo estimado a partir dos dados sujeitos ao nível de distúrbios maior. Os resultados na Figura 5.8 mostram que em boa parte do espectro:

$$|G(e^{j\omega}) - \hat{G}(e^{j\omega})| \ll |\hat{G}(e^{j\omega})|.$$

[†]Caso seja especificada uma resposta em malha fechada significativamente mais rápida do que a do processo em malha aberta, a classificação de $\hat{G}(q)$ deve ser baseada em uma faixa de frequências mais ampla.

Figura 5.8: Limite superior do erro de modelagem e $|\hat{G}(e^{j\omega})|$ na condição de perturbação mais intensa.

Além disso, o limite superior do erro de modelagem é maior do que o obtido no caso em que a perturbação é menor. Este resultado é de se esperar, pois $\Delta(\omega)$ é proporcional à densidade espectral dos resíduos do modelo. Por esta razão, se o mesmo critério usado para classificar o modelo referente ao cenário de menor perturbação fosse aplicado nesta situação, apesar de satisfatório $\hat{G}(e^{j\omega})$ seria classificado como *B* (“Bom”).

Por fim, o ajuste da função estática não-linear é analisado. Para que seja possível avaliar o efeito das perturbações, as estimativas do bloco estático nas condições de menor $\hat{f}_{(l)}(w)$ e de maior perturbação $\hat{f}_{(h)}(w)$ são mostradas no gráfico disposto do lado esquerdo da Figura 5.9. Nos dois cenários, apesar de se adotar uma estrutura diferente da usada nas simulações, apresentada em (5.3), o mapeamento $y = f(w)$ é estimado com boa exatidão.

Uma forma de quantificar o ajuste da curva estática do processo é através do índice de correlação linear [24] entre a saída calculada a partir de (5.3), Y^0 , e a calculada a partir da estimativa do bloco $f(\cdot)$, representada por $\hat{Y}$:

$$r_N [\%] = \frac{\text{cov}(Y^0, \hat{Y})}{\text{std}(Y^0) \cdot \text{std}(\hat{Y})} \times 100, \quad (5.5)$$

em que $\text{cov}(Y^0, \hat{Y})$ é a covariância entre Y^0 e $\hat{Y}$, ao passo que $\text{std}(\cdot)$ denota o desvio-padrão de uma variável. Quanto mais próximo de 100%, maior a associação linear entre Y^0 e $\hat{Y}$.

Figura 5.9: Estimativa da função não-linear (esquerda) e do ganho estático em função do ponto de operação (direita).

Ao comparar os valores de r_N obtidos para cada cenário de perturbação, dispostos na Tabela 5.2, confirma-se que, assim como o bloco linear, a estimativa da curva estática também é deteriorada diante de um nível maior de distúrbios. No entanto, o gráfico de $\hat{f}_{(h)}(w)$ mostra que tal discrepância é discreta, apesar da variância ser quase 4 vezes maior do que na situação de pouca perturbação.

Tabela 5.2: Estimativa da função estática não-linear.

Função estimada	r_N [%]
$\hat{f}_{(l)}(\cdot)$	99,9647
$\hat{f}_{(h)}(\cdot)$	99,8440

A influência da intensidade de $v(k)$ na estimativa do comportamento estático é mais evidente ao observar o ganho do processo em função do ponto de operação. Como mostrado na Figura 5.9, para o cenário de menor perturbação, o erro na

estimativa do ganho é, na pior das hipóteses (em $w = 0,344$), de 14%, ao passo que no cenário de maior perturbação o erro é de até 44% em $w = 0,344$.

Os resultados apresentados nesta seção sugerem que o método proposto para a identificação da associação entre modelos de atrito e de Wiener é apropriado, para este caso. A seguir utiliza-se um ambiente mais realista para testar a aplicabilidade desta metodologia.

5.2 Planta híbrida de neutralização de pH

Nesta seção, o método de identificação do conjunto atrito + processo é aplicado a uma plataforma híbrida, em que a válvula de controle seja real e o processo seja simulado. Esta configuração é conhecida por HIL (do Inglês, *Hardware In the Loop*). Um diagrama ilustrativo da configuração utilizada é apresentado na Figura 5.10.

Figura 5.10: Diagrama ilustrativo da plataforma HIL.

A válvula utilizada (modelo ET, fabricada pela *Fisher® International* [29]) conta com um sensor LVDT (do Inglês, *Linear Variable Differential Transformer*) para medir o deslocamento da haste $u(k)$ e um sensor de pressão, para que se possa inferir a força motriz que atua sobre o diafragma da válvula. Estes sensores são ligados a uma placa de aquisição de dados (modelo PCI-6229, *National Instruments®*). O sinal proveniente do sensor de posição é usado como entrada do modelo fenomenológico de uma planta de neutralização de pH, que é simulado em tempo real no ambiente Matlab\Simulink®. No mesmo ambiente, também é implementado o

controlador do processo, cuja saída $z(k)$ é ligada ao atuador pneumático da válvula através de um canal de saída analógico.

Processos de neutralização de pH apresentam não-linearidades severas, pois o ganho estacionário varia consideravelmente para pequenas mudanças nas condições de operação da planta. Além disso, sistemas que envolvam o controle de acidez são frequentes em indústrias químicas e biológicas [5]. Por estes motivos, processos de neutralização de pH têm sido muito usados como *benchmark* na avaliação de técnicas para identificação e controle de sistemas não-lineares [49, 72]. A seguir é descrita a planta de pH[§] que integra a plataforma HIL.

5.2.1 Descrição da planta de neutralização de pH

O processo consiste em três fluxos de entrada: um de solução ácida H_2CO_3 , um de solução tampão (*buffer*) NaHCO_3 e um de reagente alcalino NaOH representados, respectivamente, por q_1 , q_2 e q_3 . A variável controlada é o pH medido no tanque reator T1. A variável manipulada é a vazão q_3 de NaOH , controlada pelo sinal aplicado ao atuador da válvula AV. A solução alcalina é armazenada em um tanque pressurizado T2, cuja base está na mesma altura da válvula AV. Há ainda uma válvula LV que controla o fluxo de saída q_4 de modo a tentar manter o nível do tanque reator h_{T1} em um valor de referência h_{T1nom} . A Figura 5.11 mostra o diagrama da planta de neutralização que compõe a plataforma HIL.

Ao derivar o modelo da planta admite-se que: (1) os reagentes envolvidos são incompressíveis, (2) as massas específicas das soluções são idênticas, constantes e representadas por ρ , (3) os íons envolvidos são totalmente solúveis, (4) a mistura no tanque T1 é perfeita, (5) a pressão relativa P_{T2} no interior de T2 é praticamente constante, (6) as válvulas AV e LV têm característica inerente de vazão linear e (7) a

[§]A planta descrita na subseção 5.2.1 é baseada na proposta de Alvarez *et al.* [5], de modo que algumas alterações foram realizadas para tornar o modelo mais compatível com a planta piloto do Laboratório de Controle de Processos da Escola Politécnica da Universidade de São Paulo.

Figura 5.11: Planta de neutralização de pH.

dinâmica das reações químicas assim como a perda de carga nas tubulações podem ser desprezadas.

A vazão de base q_3 pode ser modelada (lei de Bernoulli) por:

$$\begin{aligned} q_3(t) &= K_{q_3} \cdot X_{AV}(t) \sqrt{\Delta P(t)} \\ &= K_{q_3} \cdot X_{AV}(t) \sqrt{P_{T2} + \rho \cdot g \cdot h_{T2}(t)}, \end{aligned} \quad (5.6)$$

em que K_{q_3} é uma constante de proporcionalidade, $0 \leq X_{AV} \leq 1$ é a abertura da válvula AV, g é a aceleração gravitacional e h_{T2} é o nível do tanque pressurizado.

Ao substituir (5.6) no balanço de massa do tanque T2 resulta:

$$\begin{aligned} \rho \cdot A_{T2} \cdot \dot{h}_{T2}(t) &= -\rho \cdot q_3(t) \\ \dot{h}_{T2}(t) &= -\frac{1}{A_{T2}} \left(K_{q_3} \cdot X_{AV}(t) \sqrt{P_{T2} + \rho \cdot g \cdot h_{T2}(t)} \right). \end{aligned} \quad (5.7)$$

A condição inicial da equação diferencial (5.7) é o nível máximo do tanque pressurizado, representado por h_{T2max} , cuja área A_{T2} é considerada constante.

O pH no tanque do reator é modelado segundo o método de invariantes de

reação[¶]. A partir desta abordagem, a concentração de íons de hidrogênio $[H^+]$ no tanque reator é determinada usando constantes de equilíbrio das reações químicas envolvidas (K_{a_1} , K_{a_2} e K_w) [5] e os invariantes de reação da solução neutralizada, W_{a_4} e W_{b_4} , através da seguinte função não-linear [41]:

$$[H^+] = W_{b_4} \left(\frac{K_{a_1} [H^+]^{-1} + 2K_{a_1} K_{a_2} [H^+]^{-2}}{1 + K_{a_1} [H^+]^{-1} + K_{a_1} K_{a_2} [H^+]^{-2}} \right) + W_{a_4} + \frac{K_w}{[H^+]}. \quad (5.8)$$

As reações químicas envolvidas no processo simulado, bem como as respectivas constantes de equilíbrio são mostradas na Tabela 5.3. Ao resolver $[H^+]$ em (5.8), obtém-se o pH da solução neutralizada:

$$\text{pH} = -\log_{10} ([H^+]). \quad (5.9)$$

Tabela 5.3: Constantes de equilíbrio das reações químicas [5].

Reação química	Constante de equilíbrio
$\text{H}_2\text{CO}_3 \xrightleftharpoons{K_{a_1}} \text{HCO}_3^- + \text{H}^+$	$K_{a_1} = 4,47 \times 10^{-7}$
$\text{HCO}_3^- \xrightleftharpoons{K_{a_2}} \text{CO}_3^{2-} + \text{H}^+$	$K_{a_2} = 5,62 \times 10^{-11}$
$\text{H}_2\text{O} \xrightleftharpoons{K_w} \text{OH}^- + \text{H}^+$	$K_w = 1,0 \times 10^{-14}$

As dependências temporais de W_{a_4} , W_{b_4} e, conseqüentemente, $[H^+]$ são omitidas em (5.8) e (5.9) por simplicidade de notação.

De forma análoga a (5.6), o fluxo q_4 é modelado através de:

$$q_4(t) = K_{q_4} \cdot X_{LV}(t) \sqrt{\rho \cdot g \cdot h_{T1}(t)}, \quad (5.10)$$

em que K_{q_4} é uma constante hidráulica e X_{LV} é a abertura da válvula de controle LV.

A abertura da válvula LV em relação ao valor nominal X_{LVnom} é controlada por um compensador PI, parametrizado por $K_{c(LIC)}$ e $T_{i(LIC)}$, de modo a manter o

[¶]A descrição do método de invariantes de reação está além do escopo deste trabalho. Os trabalhos de Campos [15], Gustafsson e Waller [33], Henson e Seborg [41] e Sotomayor [72] são alguns dos textos que tratam deste assunto.

nível do tanque reator em h_{T1nom} , a partir da condição inicial $X_{LV}(0) = X_{LVnom}$:

$$X_{LV}(t) = K_{c(LIC)} \left(h_{T1}(t) - h_{T1nom} + \frac{1}{T_{i(LIC)}} \int_0^t (h_{T1}(\tau) - h_{T1nom}) d\tau \right) + X_{LV}(0). \quad (5.11)$$

Do balanço de massa no tanque T1 resulta:

$$\dot{h}_{T1}(t) = \frac{1}{A_{T1}} (q_1(t) + q_2(t) + q_3(t) - q_4(t)), \quad (5.12)$$

em que A_{T1} é a área do tanque T1 e a condição inicial de (5.12) é dada por h_{T1nom} . Ao combinar o balanço de massa no tanque reator com o conceito de invariantes de reação segue que [15]:

$$\dot{W}_{a_4}(t) = \frac{1}{A_{T1}h_{T1}(t)} \sum_{i=1}^3 q_i(t) (W_{a_i} - W_{a_4}(t)) \quad (5.13)$$

$$\dot{W}_{b_4}(t) = \frac{1}{A_{T1}h_{T1}(t)} \sum_{i=1}^3 q_i(t) (W_{b_i} - W_{b_4}(t)). \quad (5.14)$$

Note que em (5.13) e (5.14) os invariantes de reação W_{a_i} e W_{b_i} das soluções nos fluxos q_i (para $i = 1, 2$ e 3) são considerados constantes, entretanto é possível simular mudanças na concentração molar dos reagentes ao variar tais valores. A Tabela 5.4 contém a concentração molar^{||} e os invariantes de reação dos fluxos q_1 , q_2 e q_3 .

Tabela 5.4: Concentração molar e invariantes de reação dos reagentes [5].

Fluxo	Concentração molar	Invariantes de reação
q_1	3×10^{-3} M de H_2CO_3	$W_{a_1} = 3 \times 10^{-3}$ M
		$W_{b_1} = 0$ M
q_2	3×10^{-2} M de $NaHCO_3$	$W_{a_2} = -3 \times 10^{-2}$ M
		$W_{b_2} = 3 \times 10^{-2}$ M
q_3	$3,05 \times 10^{-3}$ M de $NaOH$ e	$W_{a_3} = -3,05 \times 10^{-3}$ M
	5×10^{-5} M de $NaHCO_3$	$W_{b_3} = 5 \times 10^{-5}$ M

Por fim, a saída da planta de neutralização, representada por $y(t)$, é o pH medido que, em relação à expressão (5.9), inclui a dinâmica τ_{med} do elemento sensor

^{||}1M = $1 \frac{\text{mol}}{\text{m}^3}$.

AE:

$$y(t) = -\tau_{med} \dot{y}(t) + \text{pH}(t). \quad (5.15)$$

Os demais parâmetros empregados no simulador da planta de neutralização de pH, descrita através das expressões (5.6)-(5.15), são apresentados na Tabela 5.5. Cabe mencionar que as constantes hidráulicas K_{q3} e K_{q4} são obtidas ao substituir os valores nominais de operação da planta nas expressões (5.6) e (5.10).

O intervalo de integração usado na resolução das equações diferenciais do simulador é de 10ms. Com o mesmo período, o sinal do sensor de posição da válvula real é amostrado e atribuído a $X_{AV}(t)$, depois de convertido para p.u. Para ensaios em malha fechada, o sinal para o atuador da válvula AV é calculado pelo controlador AIC (ver Figura 5.11), em ambiente Matlab\Simulink®. Nos experimentos em

Tabela 5.5: Valores dos parâmetros da planta de neutralização de pH.

Descrição	Símbolo	Valor
Área da secção transversal do tanque T1	A_{T1}	0,1 m ²
Nível nominal do tanque T1	h_{T1nom}	0,1 m
Área da secção transversal do tanque T2	A_{T2}	0,5498 m ²
Nível máximo do tanque T2	h_{T2max}	1 m
Pressão interna no tanque T2	P_{T2}	$1,2 \times 10^6 \frac{\text{N}}{\text{m}^2}$
Massa específica das soluções	ρ	$1000 \frac{\text{kg}}{\text{m}^3}$
Aceleração gravitacional	g	$9,80665 \frac{\text{m}}{\text{s}^2}$
Constante hidráulica da válvula AV	K_{q3}	$1,5189 \times 10^{-7} \frac{\text{m}^4}{\text{s}\sqrt{\text{N}}}$
Constante hidráulica da válvula LV	K_{q4}	$1,5226 \times 10^{-5} \frac{\text{m}^4}{\text{s}\sqrt{\text{N}}}$
Posição nominal de AV	X_{AVnom}	0,6584 p.u.
Posição nominal de LV	X_{LVnom}	0,5 p.u.
Vazão nominal de ácido	q_{1nom}	$1,22 \times 10^{-4} \frac{\text{m}^3}{\text{s}}$
Vazão nominal de solução tampão	q_{2nom}	$6,4 \times 10^{-6} \frac{\text{m}^3}{\text{s}}$
Vazão nominal de base	q_{3nom}	$1,1 \times 10^{-4} \frac{\text{m}^3}{\text{s}}$
Constante de tempo do sensor de pH	τ_{med}	0,8s
pH nominal no tanque T1	pH_{nom}	7
Ganho do controlador LIC	$K_{c(\text{LIC})}$	0,1777
Tempo integral do controlador LIC	$T_{i(\text{LIC})}$	26,5s

malha aberta, o sinal de excitação é aplicado ao atuador pneumático da válvula AV via o conversor LY.

5.2.2 Estimação dos parâmetros do modelo de Karnopp

Antes de aplicar o método de identificação do conjunto atrito + processo, os parâmetros do modelo de atrito de Karnopp são estimados através do procedimento descrito na subseção 2.3.1, baseado na pressão do atuador e na posição da haste. Ao calcular o par (S, J) a partir do modelo de Karnopp usando expressões de equivalência [79]:

$$S = \frac{F_s + F_c}{\Delta P_{at} \cdot A_d} \quad (5.16)$$

$$J = \frac{F_s - F_c}{\Delta P_{at} \cdot A_d}, \quad (5.17)$$

os resultados obtidos de (5.16) e (5.17) servem de referência para avaliar o modelo de atrito estimado sem dispor de tais informações. A Tabela 5.6 contém algumas das especificações da válvula utilizada nos experimentos.

Devido ao elevado nível de ruído nos cálculos de $\hat{\ddot{x}}(k)$ e à ínfima influência exercida pelo termo $m \cdot \ddot{x}(k)$ no balanço de forças (2.8) [70], o vetor de parâmetros λ e o de regressores $\phi(k)$ em (2.12) são redefinidos:

$$\lambda = \begin{pmatrix} F_v & F_c & k_{mola} \end{pmatrix}^T \quad (5.18)$$

$$\phi(k) = \begin{pmatrix} \dot{x}(k) & \text{sgn}(\dot{x}(k)) & x(k) \end{pmatrix}^T. \quad (5.19)$$

Tabela 5.6: Especificações da válvula de controle [29] que integra a plataforma HIL.

Área do diafragma do atuador	A_d	$4,45 \times 10^{-2} \text{m}^2$
Excursão da haste	Δx	$2,8575 \times 10^{-2} \text{m}$
Faixa de pressão no atuador	ΔP_{at}	$1,37895 \times 10^5 \frac{\text{N}}{\text{m}^2}$
Massa das partes móveis	m	1,6kg
Material das gaxetas		grafite

A fim de coletar os dados de estimação envolvendo uma ampla região de operação, excitou-se o atuador da válvula com um sinal aleatório de distribuição uniforme entre 20% e 80% da faixa de valores admissíveis. Cada valor é mantido constante por 2s. Os sinais de pressão no atuador e de deslocamento da haste foram amostrados a uma frequência de 1kHz para que a velocidade $\dot{x}(k)$ pudesse ser calculada com maior exatidão.

A Figura 5.12 mostra a força motriz F_{ext} , a posição x da haste e a estimativa $\hat{x}$ da velocidade, calculada numericamente através do método descrito em Aguirre [1]. Em linhas gerais, cada janela de n_w pontos do vetor com medidas da posição da haste é aproximada por um polinômio $p_x(k)$ de ordem n_x . Com isso, a derivada numérica de $x(k)$ no ponto $k = k_0$ é dada por:

$$\hat{x}(k_0) \approx \left. \frac{d}{dk} p(k) \right|_{k=k_0}. \quad (5.20)$$

Neste caso, como a frequência de amostragem é elevada quando comparada à dinâmica do atuador, ajustou-se uma reta ($n_x = 1$) a cada janela de $n_w = 7$ pontos.

A variável auxiliar $\delta v(s)$ de valores candidatos a DV , definida em (2.13), é criada adotando $\mathcal{S} = 150$ e $\mathcal{Z} = 250$. A Figura 5.13 mostra o comportamento do vetor $\hat{\lambda}$ em função dos diferentes valores de δv . Note que as estimativas dos coeficientes referentes aos atritos de Coulomb F_c e viscoso F_v variam consideravelmente quando $\delta v(s) < 10^{-3}$ m/s, ao passo que permanecem praticamente constantes com os demais valores de $\delta v(s)$.

Na região em que a variação de $\hat{\lambda}$ é significativa o desvio-padrão $\sigma(s)$ é elevado e o coeficiente de correlação $R^2(s)$ é pequeno (Figura 5.14). Estes resultados revelam que os parâmetros corretos pertencem à região na qual $\delta v(s) > 10^{-3}$ m/s. A escolha exata de $\hat{\lambda}$ é realizada de modo a minimizar $\sigma(s)$ e maximizar $R^2(s)$. Este ponto é indicado por “ $\otimes$ ” nas Figuras 5.13 e 5.14.

Uma vez estimado o vetor λ , resta calcular o coeficiente de atrito estático F_s . Para tanto, os instantes k_b de iminência de movimento *breakaway* são identificados

Figura 5.12: Dados usados na estimação dos parâmetros do modelo de atrito de Karnopp.

Figura 5.13: Comportamento da estimativa dos parâmetros de atrito de Karnopp para diferentes valores de δv .

Figura 5.14: Desvio-padrão σ e coeficiente de correlação R^2 das estimativas para diferentes valores de δv .

dentre os dados da Figura 5.12. Em seguida, para cada k_b , F_s é resolvido na expressão (2.17), que descreve o balanço das forças que atuam na haste da válvula nos instantes de *breakaway*. A média das estimativas de F_s e os demais parâmetros do modelo de atrito de Karnopp são apresentados na Tabela 5.7. Também são mostrados S e J calculados a partir de (5.16) e (5.17).

Com o intuito de avaliar o desempenho do modelo de Karnopp e a validade do modelo de Kano “equivalente” um conjunto de dados diferente do empregado no cálculo das estimativas é usado para gerar simulações da posição da haste. A Figura 5.15 revela que ambos os modelos são capazes de reproduzir a relação entre pressão no diafragma e posição.

Além disso, a diferença entre as simulações $\hat{x}(k)$ é quase que imperceptível.

Tabela 5.7: Estimativas dos parâmetros do modelo de atrito de Karnopp e de Kano obtidos pelas expressões de equivalência.

Modelo de atrito	Parâmetros			
	$\hat{F}_v$	$\hat{F}_c$	$\hat{k}_{mola}$	$\hat{F}_s$
Karnopp	$1,282 \times 10^4 \text{ N}\cdot\text{s/m}$	688N	$2,07 \times 10^5 \text{ N/m}$	818N
Kano	$\hat{S}$		$\hat{J}$	
	24,54%		2,12%

Figura 5.15: Validação cruzada dos modelos de Karnopp e de Kano.

Isto implica que, para a válvula utilizada nos experimentos, é perfeitamente viável aproximar o modelo de Karnopp pelo de Kano. Na Figura 5.16 é mostrada a curva $x(k) = f(F_{ext}(k))$, também denominada curva de assinatura da válvula (do Inglês, *valve signature*). Note que tanto o modelo de Karnopp quanto o de Kano reproduzem com exatidão a histerese da válvula.

Devido aos resultados apresentados, as estimativas $\hat{S} = 24,54\%$ e $\hat{J} = 2,12\%$ (Tabela 5.7) são adotadas como referências nas comparações com as calculadas pelo método de identificação do conjunto atrito + processo analisado a seguir.

Figura 5.16: Relação entrada-saída (assinatura) da válvula real e reproduzida pelos modelos de Karnopp e de Kano.

5.2.3 Identificação da planta de neutralização

Para que se possa escolher o período de amostragem T_s e gerar o sinal de teste $d(k)$, deve haver uma estimativa, mesmo que grosseira, do tempo de acomodação t_s da planta. Esta informação pode ser levantada com ensaios preliminares, ou então, a partir do conhecimento prévio da planta. Ao realizar testes em malha aberta com variações na forma de degraus próximas do ponto de operação, verifica-se um ganho $K_p \approx 0,095$ e $t_s \approx 160s$. Se a resposta do processo for aproximada pela de um sistema linear de 1ª ordem, obtém-se uma constante de tempo de 40s.

Uma alternativa, caso não seja possível realizar experimentos em malha aberta, é inferir a constante de tempo dominante do processo, τ_p , através da razão entre os valores nominais do volume do tanque T1 e do somatório dos fluxos que entram no tanque:

$$\tau_p \approx \frac{A_{T1} \cdot h_{T1nom}}{q_{1nom} + q_{2nom} + q_{3nom}} = 42s. \quad (5.21)$$

Note que a constante de tempo τ_p , calculada em (5.21), é praticamente a mesma que foi estimada através da resposta ao degrau.

Segundo Tulleken [77], uma relação razoável para escolher o período de amostragem é adotar $T_s \approx 0,1\tau_p$. Logo, de acordo com (5.21), admite-se $T_s = 4s$. Durante a realização dos ensaios, um sinal GBN com tempo médio em cada patamar de 80s ($\bar{T}_{sw} = 20 \times T_s$) é somado à referência $r(k)$ que é mantida constante em $pH_{nom} = 7$ (ver Figura 5.10). O espectro do sinal de teste $d(k)$ é determinado ao escolher $\alpha = 2$ em (4.4), conforme discutido na seção 4.2. A magnitude de $d(k)$ é de $\pm 5\%$ pH_{nom} .

Com exceção do ensaio preliminar, todos os experimentos para a identificação da planta de neutralização são realizados em malha fechada. A malha de pH utiliza um controlador PI, representado por AIC na Figura 5.11, cuja sintonia é realizada através do método da Síntese Direta [10]. Este método permite expressar os parâmetros do controlador em função de uma especificação da constante de tempo da

malha fechada τ_{mf} :

$$K_{c(\text{AIC})} = \frac{\tau_p}{K_p \cdot \tau_{mf}} \quad (5.22)$$

$$T_{i(\text{AIC})} = \tau_p, \quad (5.23)$$

em que $K_{c(\text{AIC})}$ e $T_{i(\text{AIC})}$ são o ganho e o tempo integral do controlador AIC.

Nas primeiras análises, os dados são coletados com o controlador sintonizado com a especificação: $\tau_{mf} = 0,8\tau_p$. A duração de cada experimento é de 8000s (2000 pontos), de modo que uma das metades é destinada à estimação dos parâmetros e a outra é reservada para a validação.

Estudo do efeito de perturbações

Para avaliar o efeito de perturbações do processo nos parâmetros estimados, a plataforma HIL é identificada em quatro cenários sujeitos a diferentes distúrbios. Tais cenários são criados introduzindo ruído “colorido” nos valores nominais do fluxo q_1 e dos invariantes de reação W_{a_1} e W_{b_3} . A Tabela 5.8 apresenta o filtro e a variância do ruído branco $e(k)$ (de média nula) que gera o distúrbio somado a cada uma dessas variáveis.

Nos dois primeiros cenários, indicados por **1** e **2**, a vazão de ácido q_1 não sofre alterações durante o experimento. Por outro lado, em **3** e **4** são introduzidos distúrbios em q_1 , de modo que há um aumento de 100% na amplitude destas variações no cenário **4** em relação ao **3**. A intensidade das perturbações em W_{a_1} e W_{b_3} são multiplicadas respectivamente por cinco e por dois de **1** $\rightarrow$ **2**. Do cenário **3** para o **4**, os distúrbios em ambos invariantes de reação são multiplicados por dois. O intuito é simular situações sujeitas a níveis crescentes de perturbação.

Para dar uma ideia da magnitude dos distúrbios simulados, as variâncias de $e(k)$ no cenário **3** (Tabela 5.8) produzem oscilações em W_{a_1} , W_{b_3} e q_1 de até 2,15%, 10,65% e 3,75%, respectivamente. Cabe ressaltar que as variações em cada variável são geradas por realizações distintas de $e(k)$.

Tabela 5.8: Cenários utilizados para avaliar a influência de perturbações nas estimativas dos modelos.

Variável perturbada	Filtro	Variâncias de $e(k)$ em cada cenário			
		1	2	3	4
q_1	$\frac{0,2}{1-0,8q^{-1}}$	0	0	2×10^{-5}	4×10^{-5}
W_{a_1}	$\frac{0,1}{1-0,9q^{-1}}$	2×10^{-9}	1×10^{-8}	1×10^{-8}	2×10^{-8}
W_{b_3}	$\frac{0,0375}{(1-0,75q^{-1})(1-0,85q^{-1})}$	5×10^{-11}	1×10^{-10}	1×10^{-10}	2×10^{-10}

O algoritmo baseado em otimização para a busca do par (S, J) é utilizado na identificação do conjunto atrito + processo. Para tanto, os valores candidatos a S são limitados por $\max(S) = 0,3$. As demais parametrizações do algoritmo são:

$$\Delta S = 0,02, \quad l_{max} = 4, \quad m_{max} = 9 \quad \text{e} \quad n_c = 2.$$

A Tabela 5.9 apresenta os parâmetros de atrito, os índices $\mathcal{F}_2$ e r_N , bem como a classificação de $\hat{G}(q)$ obtidos nos diferentes cenários de perturbação. Entre parênteses, também é mostrada a porcentagem de $\Delta(\omega)$ em relação à $|\hat{G}(e^{j\omega})|$ no ponto mais crítico do intervalo $[0; \omega_c]$. Os gráficos de limite superior do erro de modelagem obtidos durante os teste com a plataforma HIL estão dispostos no apêndice B.

Tabela 5.9: Síntese dos resultados provenientes dos diferentes cenários de perturbação.

Cenário	$\hat{S}$ [%]	$\hat{J}$ [%]	$\mathcal{F}_2$ [%]	r_N [%]	Classificação de $\hat{G}(q)$
1	25,68	1,96	88,34	99,9537	A (13,55%)
2	25,67	1,94	87,08	99,9851	A (13,82%)
3	25,05	2,03	82,22	99,9852	B (31,77%)
4	24,50	2,23	72,89	99,9385	B (49,41%)

Assim como no exemplo numérico visto anteriormente (seção 5.1), em todos os cenários de perturbação considerados as estimativas de S e J estão bem próximas dos valores obtidos na subseção 5.2.2. Este resultado reforça a ideia de distúrbios externos não influenciarem a estimativa dos parâmetros do modelo de atrito. Conforme é mostrado na Figura 5.17, tal característica também é percebida quando é

admitido $J = 0$, nas primeiras etapas do algoritmo de estimação. Note que o mínimo do erro de simulação $\mathcal{C}_i$ ocorre para S entre 0,22 e 0,24, independentemente do nível de perturbação.

Figura 5.17: Comportamento do erro de simulação $\mathcal{C}_i$ calculado no passo (3) do algoritmo 4.2 em função do parâmetro S .

Outro aspecto a ser destacado é a deterioração do índice $\mathcal{F}_2$ à medida que a intensidade dos distúrbios aumenta. Este mesmo comportamento ocorre com a classificação do modelo do bloco linear, na região de frequências delimitada por $[0; \omega_c]$, segundo o critério do limite superior do erro de modelagem.

Em relação ao ajuste da curva estática, como é de se esperar, o menor coeficiente de correlação linear é obtido na situação de maior perturbação. No entanto, as estimativas $\hat{f}(\cdot)$ do comportamento não-linear nos cenários **2** e **3** são ligeiramente melhores do que a encontrada no **1**. A Figura 5.18 mostra $\hat{f}(\cdot)$ para cada caso, em comparação com a curva de titulação da planta de neutralização. Apenas a faixa em que há dados experimentais é analisada.

Embora tenha sido verificada a influência negativa dos distúrbios sobre a estimativa do modelo de Wiener, baseado nas ferramentas de validação discutidas na seção 4.4, pode-se dizer que modelos razoáveis do conjunto atrito + processo foram estimados em todos os cenários de perturbação. A seguir é estudado como a magnitude do sinal de teste pode influenciar os resultados.

Figura 5.18: Estimativas da curva estática nos diferentes cenários de perturbação.

Influência da magnitude do sinal de teste

Por um lado é vantajoso aumentar a magnitude do sinal de teste para elevar a relação sinal-ruído. Em contrapartida, ao excitar o processo em uma região mais ampla, os dados devem apresentar mais características não-lineares, podendo até tornar o problema de identificação mais difícil de ser resolvido. A fim de estudar o que ocorre com a qualidade dos modelos estimados quando se altera a magnitude de $d(k)$, adota-se o cenário de perturbação **3** descrito na Tabela 5.8. O processo é excitado com três intensidades: excitação natural, $\pm 2,5\% \text{ pH}_{nom}$ e $\pm 10\% \text{ pH}_{nom}$.

O algoritmo de estimação executado com as parametrizações empregadas anteriormente (pág. 97) proporciona os resultados resumidos na Tabela 5.10. Primeiramente destaca-se que, em qualquer um dos ensaios, os parâmetros do modelo de atrito não divergem mais de 4,5% em relação aos valores de referência: $S = 24,54\%$ e $J = 2,12\%$ (subseção 5.2.2). Portanto, pode-se dizer que as estimativas de S e J independem da magnitude do sinal de teste, resultando inclusive em “bons” valores mesmo no caso de excitação natural.

Tabela 5.10: Resultados obtidos ao variar a magnitude do sinal de teste $d(k)$.

Magnitude de $d(k)$	S [%]	J [%]	$\mathcal{F}_2$ [%]	r_N [%]	Classificação de $\hat{G}(q)$
0% pH_{nom}	25,35	2,17	40,85	97,5791	D (116,3%)
$\pm 2,5\%$ pH_{nom}	24,38	2,21	59,33	99,2826	C (60,55%)
$\pm 5\%$ pH_{nom}	25,05	2,03	82,22	99,9852	B (31,77%)
$\pm 10\%$ pH_{nom}	25,35	2,18	85,62	99,1942	A (14,04%)

Todavia o mesmo não se aplica à estimativa do modelo de Wiener. De acordo com os índices apresentados na Tabela 5.10, ao excitar a malha com um sinal de magnitude $\pm 2,5\%$ pH_{nom} (ou inferior), o desempenho de $\hat{G}(q)$ e de $\hat{f}(\cdot)$ é drasticamente comprometido. Este resultado é ainda mais explícito ao comparar a saída dos modelos estimados com o pH da plataforma HIL (Figura 5.19).

Há basicamente dois aspectos que justificam esta característica: em primeiro lugar, reduzir a magnitude do sinal $d(k)$ implica em piorar a relação sinal-ruído. Além disso, devido ao atrito elevado na válvula, variações menores na saída do controlador resultam em menos variações na entrada do processo, comprometendo a persistência de excitação [1, 58] dos dados utilizados na estimação do modelo de Wiener.

Como pode ser verificado na Figura 5.19, as variações no pH do processo são proporcionais à amplitude do sinal de teste $d(k)$. Em outras palavras, quanto menor a magnitude da excitação, menor a presença de informações dinâmicas nos dados experimentais. A classificação dos modelos $\hat{G}(q)$, baseada na relação entre $\Delta(\omega)$ e $|\hat{G}(e^{j\omega})|$, apresentada na Tabela 5.10, também reflete este comportamento.

Outro aspecto interessante é o experimento com $d(k) = \pm 10\%$ pH_{nom} gerar um índice r_N pior do que em outros casos onde a magnitude do sinal de teste é menor. Isto ocorre porque apenas a faixa onde há dados é considerada para o cálculo de r_N . Portanto, ensaios restritos a uma faixa menor podem, eventualmente, proporcionar um índice de correlação linear maior, apesar de não reproduzirem tão bem a curva estática em uma região específica. A Figura 5.20 confirma tal justificativa. Enquanto

Figura 5.19: Dados da plataforma HIL e estimativa do pH dos modelos obtidos com $d(k)$ de diferentes magnitudes: (a) 0% pH_{nom} (excitação natural), (b) $\pm 2,5\%$ pH_{nom} , (c) $\pm 5\%$ pH_{nom} e (d) $\pm 10\%$ pH_{nom} .

o ensaio gerado sem nenhum sinal de teste (excitação natural) revela pouco a respeito da curva estática, a estimativa gerada no experimento excitado com amplitude $\pm 10\%$ pH_{nom} é a que melhor aproxima a curva de titulação do processo, especialmente nas faixas onde são realizados os demais testes.

Em suma, os resultados obtidos ao variar a magnitude do sinal de teste sugerem que quanto maiores as variações em $d(k)$, melhor é a qualidade dos modelos estimados. Naturalmente, na prática tal escolha envolve a consideração de restrições operacionais.

Em alguns trabalhos que também identificam os modelos do atrito e do processo [23, 45, 73], argumenta-se que é possível quantificar o atrito em válvulas de controle sem recorrer a sinais de excitação externos. Os resultados apresentados

Figura 5.20: Estimativas do comportamento não-linear ao variar a magnitude de $d(k)$.

anteriormente revelam que embora os parâmetros S e J tenham sido estimados com boa exatidão, não há como validá-los nos casos onde a magnitude de $d(k)$ seja menor, pois o modelo do processo compromete a estimativa de $\hat{y}(k)$, empregada no cálculo do limite superior do erro de modelagem e nos índices $\mathcal{F}_2$ e r_N .

Efeito da sintonia do controlador nos parâmetros estimados

Segundo Uehara [78], alguns métodos de quantificação do atrito em válvulas de controle propostos na literatura [22, 39] são sensíveis à sintonia do controlador, pois estes métodos não são capazes de distinguir as oscilações provocadas pelo atrito das geradas por controladores mal sintonizados. Com o intuito de avaliar o algoritmo de estimação do conjunto atrito + processo em tais circunstâncias, o ganho $K_{c(AIC)}$, definido em (5.22), é variado adotando outras três especificações para a constante de tempo dominante de malha fechada: $\tau_{mf} = \tau_p$, $\tau_{mf} = 0,5\tau_p$ e $\tau_{mf} = 0,2\tau_p$. Os distúrbios do cenário **3** (Tabela 5.8) são adotados para simular perturbações do processo e a magnitude do sinal de excitação é mantida em $\pm 5\%$ pH_{nom} .

Como pode ser visto na Figura 5.21 o processo apresenta um comportamento mais oscilatório, à medida que a sintonia do controlador se torna mais agressiva. A estimativa do pH também é mostrada em cada caso, para que se possa visualizar o ajuste dos modelos aos dados experimentais.

Figura 5.21: Dados de experimentos gerados ao variar o ganho do controlador de acordo com as especificações: (a) $\tau_{mf} = \tau_p$, (b) $\tau_{mf} = 0,8\tau_p$, (c) $\tau_{mf} = 0,5\tau_p$ e (d) $\tau_{mf} = 0,2\tau_p$.

Os índices que medem o desempenho dos modelos estimados são apresentados na Tabela 5.11. De uma forma geral, as estimativas do par (S, J) são similares em relação às obtidas anteriormente. Embora a sintonia $\tau_{mf} = 0,2\tau_p$ resulte em um $\hat{J}$ mais distante (13,7%) do valor de referência (subseção 5.2.2), ressalta-se que esta divergência não é significativa, pois: (1) a contribuição de J é menor em relação à de S e (2) a diferença $S - J$ é equivalente às encontradas com as demais sintonias.

Não se pode negar que há uma variação mais pronunciada nos valores dos

Tabela 5.11: Efeito da sintonia do controlador nos modelos estimados.

Sintonia	S [%]	J [%]	$\mathcal{F}_2$ [%]	r_N [%]	Classificação de $\hat{G}(q)$
$\tau_{mf} = \tau_p$	25,48	2,20	81,17	99,9108	B (31,92%)
$\tau_{mf} = 0,8\tau_p$	25,05	2,03	82,22	99,9852	B (31,77%)
$\tau_{mf} = 0,5\tau_p$	26,19	2,12	87,90	99,6816	B (27,70%)
$\tau_{mf} = 0,2\tau_p$	23,59	1,83	72,23	99,8725	B (24,42%)

parâmetros de atrito do que a percebida nos testes realizados até este momento. Todavia, esta variação é bem menor do que a apresentada por outros métodos [22, 39], onde mudanças menores no ganho do controlador geram erros de até 38,55% [78]. Logo, os resultados levam a crer que o método de quantificação de atrito deste trabalho é mais robusto quando aplicado a malhas cujas oscilações também sejam devidas à sintonia inadequada do controlador.

No que se refere à estimativa do modelo do processo, a despeito da sintonia empregada, os índices $\mathcal{F}_2$ e r_N , assim como a classificação segundo o critério baseado no limite superior do erro de modelagem revelam resultados semelhantes. A seguir são discutidos alguns aspectos relacionados às informações da Tabela 5.11:

- Quando a sintonia é muito agressiva ($\tau_{mf} = 0,2\tau_p$), há degradação no índice $\mathcal{F}_2$. Isto pode ser explicado pelo fato do ensaio revelar poucas informações do regime estático (Figura 5.21). Com isso, a estimativa $\hat{f}(\cdot)$ da curva estática, mostrada na Figura 5.22, também é comprometida;
- Devido à amplitude de valores alcançados, visivelmente maior que as apresentadas pelos demais ensaios (Figura 5.22), o experimento cujo controlador é sintonizado a partir de $\tau_{mf} = 0,5\tau_p$ proporciona o pior índice de correlação linear r_N . Além disso, como pode ser verificado na Figura 5.21, as discrepâncias são maiores na região em que $\text{pH} > 8$, pois: (1) há poucas amostras e (2) a curva de titulação varia substancialmente. Entretanto, se considerada apenas a faixa onde a maioria dos dados está concentrada ($6,4 < \text{pH} < 8$), o ajuste da curva estática é tão bom quanto o obtido com outras sintonias.

Figura 5.22: Ajuste da curva estática dos experimentos realizados ao variar a sintonia do controlador AIC.

Vantagens da estrutura Hammerstein-Wiener

Como dito anteriormente (capítulo 4), a estrutura composta por um modelo de atrito parametrizado pelo par (S, J) e uma dinâmica linear é empregada por outros autores [23, 45] para representar a associação entre atrito e processo. A rigor, de acordo com a estrutura que denota o modelo do processo: linear ou não-linear (Wiener), esta associação pode ser vista como um modelo de Hammerstein ($\mathcal{N} \rightarrow \mathcal{L}$) ou Hammerstein-Wiener ($\mathcal{N}_1 \rightarrow \mathcal{L} \rightarrow \mathcal{N}_2$).

Além de prover uma estimativa do comportamento estático, argumentou-se na seção 4.1 que a adoção de uma estrutura de Wiener tem por objetivo impedir que não-linearidades comprometam a quantificação do atrito e minimizar problemas de polarização ao estimar o modelo do processo. Com o propósito de avaliar a eficácia desta estratégia, os valores de $\hat{S}$, $\hat{J}$ e do índice $\mathcal{F}_2$ obtidos ao modelar o processo com uma estrutura ARMAX são comparados aos verificados ao variar a magnitude do sinal de teste.

A ideia de variar o sinal de teste é efetuar a comparação nos seguintes contextos: (1) excitação natural que é de grande interesse prático, (2) diferentes relações sinal-ruído e (3) faixa de operação mais ampla, no caso de $\pm 10\%$ pH_{nom} . A Tabela 5.12 apresenta os resultados obtidos ao representar o processo por uma estrutura linear. Para facilitar as comparações, também são mostradas as estimativas $\hat{S}$, $\hat{J}$ e o índice $\mathcal{F}_2$ verificados para um modelo de processo não-linear (Tabela 5.10).

A análise dos resultados na Tabela 5.12 é sintetizada nos seguintes aspectos:

- Para o caso de excitação natural os níveis de atrito quantificados são notadamente distintos. Além da pequena relação sinal-ruído, outra provável justificativa é o fato do modelo de perturbações da estrutura ARMAX não ser apropriado para a planta considerada, pois sabe-se que a combinação de tais fatores causam estimativas polarizadas, principalmente se os dados forem coletados com o sistema operando em malha fechada [30, 58];
- Nos experimentos com excitação de $\pm 2,5\%$ e $\pm 5\%$ do pH nominal os parâmetros de atrito assemelham-se aos estimados nos demais testes. Entretanto, a diferença nos índices $\mathcal{F}_2$ mostra uma clara vantagem em modelar o processo com uma estrutura não-linear;
- Quando os dados são gerados com um sinal $d(k)$ de magnitude $\pm 10\%$ pH_{nom} , a região de operação aumenta consideravelmente. Neste caso, os dados contêm

Tabela 5.12: Comparativo entre os resultados obtidos ao representar o processo por uma estrutura linear (ARMAX) e não-linear (Wiener).

Magnitude de $d(k)$	Linear			Não-linear		
	S [%]	J [%]	$\mathcal{F}_2$ [%]	S [%]	J [%]	$\mathcal{F}_2$ [%]
0% pH_{nom}	16,35	2,10	9,171	25,35	2,17	40,85
$\pm 2,5\%$ pH_{nom}	23,53	1,92	46,18	24,38	2,21	59,33
$\pm 5\%$ pH_{nom}	24,70	2,05	58,96	25,05	2,03	82,22
$\pm 10\%$ pH_{nom}	19,97	2,04	50,65	25,35	2,18	85,62

mais informações acerca do comportamento não-linear da planta de neutralização de pH. Como a estrutura ARMAX não é capaz de reproduzir tais características, há uma degradação tanto no modelo de atrito quanto no do processo.

De fato, as comparações efetuadas revelam vantagens significativas em aproximar o conjunto atrito + processo por uma estrutura de Hammerstein-Wiener. A fim de mostrar tal vantagem graficamente, a Figura 5.23 apresenta o pH da plataforma HIL e os reproduzidos pelos modelos estimados quando o processo é representado por uma estrutura linear (ARMAX) e não-linear (Wiener). Todavia, como a curva estática da planta de neutralização de pH varia substancialmente, ressalta-se que no caso de processos em que a não-linearidade seja menos evidente é de se esperar que os proporcionados por modelos de Wiener tendam a ser menos expressivos.

Figura 5.23: Comparação do pH reproduzido por modelos do processo linear e não-linear identificados com as seguintes magnitudes de $d(k)$: (a) 0% pH_{nom} (excitação natural), (b) $\pm 2,5\%$ pH_{nom} , (c) $\pm 5\%$ pH_{nom} e (d) $\pm 10\%$ pH_{nom} .

Capítulo 6

Conclusões e sugestões para trabalhos futuros

Os aspectos mais relevantes referentes às análises realizadas foram apontados à medida que os resultados eram apresentados. Portanto, neste capítulo pretende-se apenas enfatizar as principais conclusões. Em seguida, também são sugeridas algumas possibilidades para dar continuidade ao tema abordado neste trabalho.

6.1 Conclusões gerais

Duas formas de estender o método de quantificação de atrito e identificação do processo desenvolvido por Srinivasan *et al.* [73] foram propostas, de modo que o conjunto atrito + processo é representado por um modelo de Hammerstein-Wiener. Na primeira os parâmetros de atrito (S e J) são determinados por busca exaustiva, ao passo que na segunda são calculados por otimização.

Os algoritmos desenvolvidos foram primeiramente testados com dados totalmente simulados e verificou-se que ambos proporcionam resultados similares. Entretanto, o algoritmo baseado em otimização demanda um esforço computacional significativamente menor. Em seguida, apenas a abordagem baseada em otimização foi usada em uma plataforma híbrida (HIL), na qual uma válvula de controle é integrada em tempo real com um modelo fenomenológico de uma planta não-linear.

A despeito de se variar a intensidade dos distúrbios do processo, a magnitude do sinal de teste ou a sintonia do controlador, as estimativas dos parâmetros do modelo de atrito são bem próximas de valores de referência, calculados por um método alternativo, baseado na informação da posição da haste da válvula. Este fato revela uma importante vantagem do algoritmo empregado em relação a outras técnicas de quantificação de atrito [22, 39, 50], cujos resultados são influenciados pela sintonia do controlador.

Com relação à identificação do modelo do processo, como é de se esperar, os resultados mostraram que as perturbações não medidas degradam a estimativa do modelo de Wiener, especialmente do bloco dinâmico linear. Também observou-se que a estimativa do modelo do processo melhora quando cresce a relação sinal-ruído nos dados.

Diferentemente de outros trabalhos [23, 45], no presente desenvolvimento é dada igual importância à identificação dos modelos de atrito e do processo. Todavia, mesmo que se esteja interessado apenas na quantificação do atrito, é vital que se obtenha um bom modelo do processo, caso contrário não é possível validar o atrito quantificado através de simulações. Este fato, associado aos resultados obtidos, revelam a importância de se utilizar uma excitação externa durante o experimento para coleta de dados, descartada pelos autores dos trabalhos supracitados.

Outra importante vantagem obtida ao modelar o processo com uma estrutura não-linear é permitir que o comportamento estacionário seja melhor estimado. Nos testes com a plataforma HIL verificou-se um “bom” ajuste da curva estática, desde que a malha de controle seja suficientemente excitada, em outras palavras, a relação sinal-ruído deve ser a maior possível. A estimativa do comportamento estático pode ser aplicada no projeto de estratégias de controle não-linear [33, 72].

A relação entre o módulo da resposta em frequência e a estimativa do limite superior do erro de modelagem usada como uma ferramenta complementar na análise e validação do bloco dinâmico linear. Tal ferramenta proporciona informações

relativas à qualidade do modelo no domínio da frequência, muito útil no projeto de sistemas de controle robustos. Outra aplicação é auxiliar na escolha de um novo sinal de teste capaz de excitar o sistema em frequências nas quais o modelo é deficiente. Contudo, a estimativa deste limite superior é baseada em diversas aproximações e, portanto, deve ser usada criteriosamente.

A princípio, os algoritmos desenvolvidos neste trabalho podem ser diretamente aplicados a situações reais. Entretanto, alguns aspectos práticos podem trazer certas dificuldades:

- Caso os dados da saída do controlador estejam armazenados em unidades de engenharia, é necessário normalizá-los, uma vez que a entrada do modelo de atrito admite apenas valores entre 0 e 1;
- Quando o processo apresentar atraso de transporte (tempo morto), o algoritmo proposto poderia ser executado admitindo atrasos arbitrários de n_k amostras;
- O período de amostragem deve ser suficientemente curto para que o sinal amostrado retenha as características dinâmicas do processo, se esta condição não for satisfeita, devido a limitações em sistemas de aquisição e armazenamento dos dados, os resultados podem ser severamente deteriorados.

6.2 Sugestões para trabalhos futuros

A seguir são sugeridas algumas perspectivas de extensão deste trabalho:

- Estudo de técnicas de compensação simultânea do atrito no atuador (válvula) e da característica não-linear do processo. Uma primeira ideia é integrar a estrutura de pré-alimentação para compensação de atrito proposta por [39] com estratégias de escalonamento de ganho. Técnicas baseadas em linearização por realimentação da saída também poderiam ser utilizadas;

- Em muitos casos, algumas das perturbações também são medidas. Em tais situações a generalização dos algoritmos estudados para o caso multivariável é de grande interesse. Para o caso da planta de neutralização, por exemplo, se o fluxo de solução ácida fosse medido, um modelo MISO (do Inglês, *Multiple Input and Single Output*) poderia ser estimado;
- No capítulo 4 o sinal de teste foi projetado considerando apenas a identificação do modelo do processo e por conveniência a excitação externa foi introduzida no valor de referência da malha. Até onde se sabe, o projeto de experimentos para a quantificação dos parâmetros de modelos de atrito é um problema em aberto;
- O conceito de separar a identificação de um sistema em um problema “caixa-branca” e outro “caixa preta” poderia ser aplicado a outras não-linearidades como zona-morta, saturação, entre outras;
- Por fim, uma extensão natural deste trabalho seria a aplicação dos algoritmos desenvolvidos a dados de processos reais.

Referências Bibliográficas

- [1] Aguirre, L. A. *Introdução à Identificação de Sistemas*. UFMG, Belo Horizonte, 2^a edição, 2004.
- [2] Aguirre, L. A. A nonlinear correlation function for selecting delay time. *Physics Letters A*, 203:88–94, 1995.
- [3] Aguirre, L. A.; Coelho, M. C. S.; Corrêa, M. V. On the interpretation and practice of dynamical differences between Hammerstein and Wiener models. *IEEE Proceedings, Part D: Control Theory & Applications*, 152(4):349–356, 2005.
- [4] Altpeter, F. *Friction modeling, identification and compensation*. PhD thesis, Ecole Polytechnique Federale de Lausanne, Department de Genie Mecanique, 1999.
- [5] Alvarez, H.; Londoo, C.; di Sciascio, F.; Carelli, R. pH Neutralization process as a benchmark for testing nonlinear controllers. *Industrial & Engineering Chemistry Research*, 40(11):2467–2473, 2001.
- [6] Armstrong-Héulovry, B.; Dupont, P.; Canudas de Wit, C. A survey of models, analysis tools and compensation methods for the control of machines with friction. *IEEE Transactions on Automatic Control*, 30(7):1083–1138, 1994.
- [7] Bai, E. W. Identification of linear systems with hard-input nonlinearities of known structure. *Automatica*, 38:853–860, 2002.
- [8] Bai, E. W. Frequency domain identification of Wiener models. *Automatica*, 39: 1521–1530, 2003.
- [9] Barenthin, M. *On input design in system identification for control*. PhD thesis, KTH School of Electrical Engineering, Stockholm, Sweden, 2006.

- [10] Bequette, B. W. *Process Dynamics: Modeling, Analysis, and Simulation*. Prentice Hall, Upper Saddle River, NJ, 2003.
- [11] Besançon-Voda, A.; Blaha, P. Describing function approximation of a two-relay system configuration with application to Coulomb friction identification. *Control Engineering Practice*, 10:655–668, 2002.
- [12] Bialkowski, W. L. Dreams vs. reality: A view from both sides of the gap. In *Control Systems*, p. 283–294, Wistler, BC, Canada, 1992.
- [13] Billings, S. A. Identification of nonlinear systems - A survey. *IEE Proceedings, Part D*, 127(6):272–285, 1980.
- [14] Billings, S. A.; Fakhouri, S. Y. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.
- [15] Campos, R. C. C. Projeto e construção de planta piloto de neutralização de pH e proposta de metodologia para incorporação de informações auxiliares na identificação NARX racional. Dissertação de mestrado, Centro Universitário do Leste de Minas Gerais, Cel. Fabriciano, MG, 2007.
- [16] Canudas de Wit, C.; Olsson, H.; Åström, K. J.; Lischinsky, P. A new model for control of systems with friction. *IEEE Transactions on Automatic Control*, 40:419–425, 1995.
- [17] Cervantes, A. L.; Agamennoni, A. E.; Figueroa, J. L. A nonlinear model predictive control system based on Wiener piecewise linear models. *Journal of Process Control*, 13:655–666, 2003.
- [18] Cheok, K. C.; Hu, H. -X.; Loh, N. K. Modeling and identification of a class of servomechanism systems with stick-slip friction. *Transactions of the ASME*, 110:324–328, 1988.
- [19] Chou, C. T.; Verhaegen, M. An indirect approach to closed-loop identification of Wiener models. In *Proceedings of the American Control Conference*, Hyatt Regency San Diego, California, USA, June 1999.

- [20] Choudhury, M. A. A. S.; Thornhill, N. F.; Shah, S. L. A data-driven model for valve stiction. In *Proceedings of the 5th IFAC Symposium on Advanced Control of Chemical Process (ADCHEM)*, Hong-Kong, January 11-14 2003.
- [21] Choudhury, M. A. A. S.; Thornhill, N. F.; Shah, S. L. Modeling valve stiction. *Control Engineering Practice*, 13:641–658, 2005.
- [22] Choudhury, M. A. A. S.; Thornhill, N. F.; Shah, S. L.; Shook, D. S. Automatic detection and quantification of stiction in control valves. *Control Engineering Practice*, 16:1395–1412, 2006.
- [23] Choudhury, M. A. A. S.; Jain, M.; Shah, S. L.; Shook, D. S. Stiction - definition, modelling, detection and quantification. *Journal of Process Control*, 18:232–243, 2008.
- [24] Coelho, M. C. S. Modelos de Hammerstein e de Wiener: conexões com modelos NARX e sua aplicação em identificação de sistemas não-lineares. Dissertação de mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2002.
- [25] Desborough, L.; Miller, R. Increasing customer value of industrial control performance monitoring - Honeywell's experience. In *Proceedings of the 6th International Conference on Chemical Process, CPC - VI*, p. 169–189, Tucson, USA, 2001.
- [26] Dupont, P. Avoiding stick-slip through PD control. *IEEE Transactions on Automatic Control*, 39:1094–1097, 1994.
- [27] Dupont, P.; Hayward, V.; Armstrong, B.; Altpeter, F. Single state elastoplastic friction models. *IEEE Transactions on Automatic Control*, 47(5):787–792, 2002.
- [28] *Type 657 Diaphragm Actuator, Sizes 30-70 and 87, Instruction Manual - Form 1900*. Fisher Controls International, February 2007.
- [29] *Design ET, EAT, and ETR Sliding-Stem Controls Valves, Product Bulletin - 51.1:ET*. Fisher Controls International, April 2007.
- [30] Forssell, U. *Closed-loop Identification: Methods, Theory, and Applications*. Licentiate thesis no. 566, Department of Electrical Engineering, Linköping University, Linköping, Sweden, Mar 1999.

- [31] Garcia, C. Comparison of friction models applied to a control valve. *Control Engineering Practice*, 16(10):1231–1243, 2008.
- [32] Golub, G. H.; Pereyra, V. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973.
- [33] Gustafsson, T. K.; Waller, K. V. Dynamic modeling and reaction invariant control of pH. *Chemical Engineering Science*, 38(3):389–398, 1983.
- [34] Gómez, J. C.; Baeyens, E. Hammerstein and Wiener model identification using rational orthonormal bases. *Latin American Applied Research*, 33(4):449–456, 2003.
- [35] Haessig, D. P.; Friedland, B. On the modeling and simulation of friction. *ASME Journal of Dynamic Systems, Measurement and Control*, 113:354–362, 1991.
- [36] Hagenblad, A. *Aspects of the Identification of Wiener Models*. Licentiate thesis no. 793, Department of Electrical Engineering, Linköping University, Linköping, Sweden, Nov 1999.
- [37] Hagenblad, A. Initialization and model reduction for Wiener model identification. In *Proceedings of the 7th Mediterranean Conference on Control and Automation (MED99)*, Haifa, Israel, 1999.
- [38] Hagenblad, A.; Ljung, L.; Wills, A. Maximum likelihood identification of Wiener models. *Automatica*, 44:2697–2705, 2008.
- [39] Hägglund, T. Automatic on-line estimation of backlash in control loops. *Journal of Process Control*, 17:489–499, 2007.
- [40] Haykin, S. *Redes Neurais: Princípios e Prática*. Bookman, Porto Alegre, 2^a edição, 2001.
- [41] Henson, M.; Seborg, D. Adaptive nonlinear control of a pH neutralization process. *IEEE Transactions on Control Systems Technology*, 2:169–182, 1994.
- [42] Hjalmarsson, H.; Gevers, M.; F., De Bruyne. For model-based control design, closed-loop identification gives better performance. *Automatica*, 32(12):1659–1673, 1996.

-
- [43] Hsu, J. T.; Ngo, K. D. T. A Hammerstein-based dynamic model for hysteresis phenomenon. *IEEE Transactions on Power Electronics*, 12:406–413, 1997.
- [44] Jelali, M. An overview of control performance assessment technology and industrial applications. *Control Engineering Practice*, 14:441–466, 2006.
- [45] Jelali, M. Estimation of valve stiction in control loops using separable least-squares and global search algorithms. *Journal of Process Control*, 18:632–642, 2008.
- [46] Johannes, V. I.; Green, M. A.; Brockley, C. A. The role of the rate of application of the tangential force in determining the static friction coefficient. *Wear*, 24:381–385, 1973.
- [47] Jurado, F. A method for identification of solid oxide fuel cells using a Hammerstein model. *Journal of Power Sources*, 154:145–152, 2006.
- [48] Kaiyhan, A.; Doyle III, F. J. Friction compensation for a process control valve. *Control Engineering Practice*, 8(7):799–812, 2000.
- [49] Kalafatis, A.; Wang, L.; Cluett, W. R. Identification of Wiener-type nonlinear systems in a noise environment. *International Journal of Control*, 66(6):923–941, 1997.
- [50] Kano, M.; Hiroshi, M.; Kugemoto, H.; Shimizu, K. Practical model and detection algorithm for valve stiction. In *Proceedings of the 7th IFAC Symposium on dynamics and control of process systems (DYCOPS)*, Cambridge, MA, USA, July 5-7 2004.
- [51] Karnopp, D. Computer simulation of stick-slip friction in mechanical dynamic systems. *Transactions of the ASME - Journal of Dynamic Systems, Measurement and Control*, 107(1):100–103, 1985.
- [52] Kim, M. S.; Chung, S. C. Friction identification of ball-screw driven servomechanisms through the limit cycle analysis. *Mechatronics*, 16:131–140, 2006.
- [53] Kirk, D. E. *Optimal control theory: An Introduction*. Prentice Hall, 1970.

- [54] Lagarias, J. C.; Reeds, J. A.; Wright, M. H.; Wright, P. E. Convergence properties of the Nelder-Mead Simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.
- [55] Lampaert, V.; Swevers, J.; Al-Bender, F. Modification of the leuven integrated friction model structure. *IEEE Transactions on Automatic Control*, 47:683–687, 2002.
- [56] Lancaster, P.; Šalkauskas, K. *Curve and Surface Fitting: An introduction*. Academic Press, London, 1986.
- [57] Ljung, L. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Transactions on Automatic Control*, 30(9):834–941, 1985.
- [58] Ljung, L. *System Identification: theory for the user*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1999.
- [59] Ljung, L. Identification for control - What is there to learn. Technical Report LiTH-ISY-R-1996, Dept of EE. Linköping University, Jan 1998.
- [60] Ljung, L. Perspectives on system identification. In *Proceedings of the 17th IFAC World Congress*, Seoul, Korea, July 6-11 2008.
- [61] Ljung, L. *System Identification Toolbox 7.1 - User's Guide*. The MathWorks, Inc., September 2007.
- [62] Ninness, B.; Wills, A. An identification toolbox for profiling novel techniques. In *Proceedings of the 16th IFAC symposium on system identification*, 2006.
- [63] Norquay, S. J.; Palazoglu, A.; Romagnoli, A. Application of Wiener model predictive control (WMPC) to an industrial C2-splitter. *Journal of Process Control*, 9:461–473, 1999.
- [64] Pearson, R. K. Selecting nonlinear model structures for computer control. *Journal of Process Control*, 13:1–26, 2003.
- [65] Pearson, R. K.; Pottmann, M. Gray-box identification of block-oriented nonlinear models. *Journal of Process Control*, 10:301–315, 2000.

- [66] Ravanbod-Shirazi, L.; Besançon-Voda, A. Friction identification using the Karnopp model, applied to an electropneumatic actuator. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 217(2):123–138, 2003.
- [67] Rinehart, N. F. A different perspective on plant reliability: Improve reliability by improving process control. In *Proceedings of the 61th Instrumentation Symposium for the Process Industries*, 2006.
- [68] Rizos, D. D.; Fassois, S. D. Maxwell slip model based identification and control of systems with friction. In *Proceedings of the 44th IEEE Conference on Decision and Control*, Seville, Spain, Dec 12-15 2005.
- [69] Romano, R. A.; Garcia, C. Comparison between two friction model parameter estimation methods applied to control valves. In *Proceedings of the 8th IFAC Symposium on dynamics and control of process systems (DYCOPS)*, Cancún, Mexico, June 6-8 2007.
- [70] Romano, R. A.; Garcia, C. Karnopp friction model identification for a real control valve. In *Proceedings of the 17th IFAC World Congress*, Seoul, Korea, July 6-11 2008.
- [71] Romano, R. A.; Garcia, C. Valve friction and nonlinear process model closed-loop identification. In *Proceedings of the 7th IFAC International Symposium on Advanced Control of Chemical Processes (ADCHEM)*, 2009.
- [72] Sotomayor, O. A. Z. Modelagem, simulação e controle de um processo de neutralização de pH. Dissertação de mestrado, Escola Politécnica da Universidade de São Paulo, São Paulo, SP, 1997.
- [73] Srinivasan, R.; Rengaswamy, R.; Narasimhan, S.; Miller, R. Control loop performance assessment. 2. Hammerstein model approach for stiction diagnosis. *Industrial & Engineering Chemistry Research*, 44(17):6719–6728, 2005.
- [74] Stenman, A.; Gustafsson, F.; Forsman, K. A segmentation-based method for detection of stiction in control valves. *International Journal of Adaptive Control and Signal Processing*, 17:625–634, 2003.

- [75] Swevers, J.; Al-Bender, F.; Ganseman, C. G.; Prajogo, T. An integrated friction model structure with improved presliding behavior for accurate friction compensation. *IEEE Transactions on Automatic Control*, 45(4):675–686, 2000.
- [76] Tjärnström, F. *Variance expressions and model reduction in system identification*. Licentiate thesis no. 730, Department of Electrical Engineering, Linköping University, Linköping, Sweden, Feb 2002.
- [77] Tulleken, H. J. A. F. Generalized binary noise test-signal concept for improved identification-experiment design. *Automatica*, 26(1):37–49, 1990.
- [78] Uehara, D. Detecção e quantificação de atrito em válvulas de controle. Dissertação de mestrado, Escola Politécnica da Universidade de São Paulo, São Paulo, SP, 2009.
- [79] Uehara, D.; Garcia, C.; Romano, R. Comparação e equivalência dos modelos de atrito de Kano e Karnopp aplicados a válvulas de controle. In *Anais do 17th Congresso Brasileiro de Automatica*, Juiz de Fora, Brasil, Set 14-17 2008.
- [80] Van den Bosch, P. P. J.; Van der Klauw, A. C. *Modeling, Identification and Simulation of Dynamical Systems*. CRC Press, 1994.
- [81] Vörös, J. Parameter identification of Wiener systems with discontinuous nonlinearities. *Systems & Control Letters*, 44:363–372, 2001.
- [82] Vörös, J. Identification of nonlinear dynamic systems using extended Hammerstein and Wiener models. *Control-Theory and Advanced Technology*, 10(4): 1203–1212, 1995.
- [83] Wahlberg, B. Model reduction of high order estimated models: the asymptotic ML approach. *International Journal of Control*, 49(1):169–192, 1989.
- [84] Walters, F. H.; Morgan, S. L.; Deming, S. N.; Parker, L. R. *Sequential simplex optimization: A Technique for Improving Quality and Productivity in Research, Development, and Manufacturing*. CRC Press, Boca Raton, 1991.
- [85] Westwick, D. T.; Verhaegen, M. Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing*, 52(2):235–258, 1996.

-
- [86] Wigren, T. Recursive prediction error identification using the nonlinear Wiener model. *Automatica*, 29(4):1011–1025, 1993.
 - [87] Yuan, H.; Westwick, D. T.; Ingenito, E. P.; Lutchen, K. R. Parametric and nonparametric nonlinear system identification of lung tissue strip mechanics. *Annals of Biomedical Engineering*, 27:548–562, 1999.
 - [88] Zhu, Y. Identification of Hammerstein models for control using ASYM. *International Journal of Control*, 73(18):1692–1702, 2000.
 - [89] Zhu, Y. *Multivariable System Identification for Process Control*. Elsevier Science, Oxford, 2001.
 - [90] Zhu, Y. Estimation of an N-L-N Hammerstein-Wiener model. *Automatica*, 38:1607–1614, 2002.
 - [91] Zhu, Y. Distillation column identification for control using Wiener model. In *Proceedings of the American Control Conference*, Hyatt Regency San Diego, California, USA, June 1999.
 - [92] Zhu, Y.; Van den Bosch, P. P. J. Optimal closed-loop identification test design for internal model control. *Automatica*, 36:1237–1241, 2000.

Apêndice A

Mínimos quadrados com restrições de igualdades

No capítulo 3 uma estimativa inicial dos parâmetros do modelo de Wiener foi calculada ao resolver o seguinte problema de otimização:

$$\hat{\xi} = \arg \min_{\xi} \|\Psi \cdot \xi\|_2^2 \quad (\text{A.1})$$

sujeito à restrição $\mathcal{R}\xi = \kappa$, em que:

$$\mathcal{R} = \left(\underbrace{0 \ \dots \ 0}_m \ \underbrace{1 \ \dots \ 1}_n \right), \quad (\text{A.2})$$

κ denota o ganho do bloco dinâmico linear de ordem n e m é a ordem da polinômio seccionado que parametriza o bloco estático não-linear.

De acordo com (3.18), a matriz de regressores é dada por:

$$\Psi^T = \begin{pmatrix} |y(n+1) - y_2|^3 & \cdots & |y(N-1) - y_2|^3 & |y(N) - y_2|^3 \\ \vdots & \vdots & \vdots & \vdots \\ |y(n+1) - y_{m-1}|^3 & \cdots & |y(N-1) - y_{m-1}|^3 & |y(N) - y_{m-1}|^3 \\ 1 & \cdots & 1 & 1 \\ y(n+1) & \cdots & y(N-1) & y(N) \\ -u(n) & \cdots & -u(N-2) & -u(N-1) \\ \vdots & \vdots & \vdots & \vdots \\ -u(1) & \cdots & -u(N-1-n) & -u(N-n) \end{pmatrix}. \quad (\text{A.3})$$

Neste apêndice, pretende-se derivar a expressão (3.23), usada para estimar ξ . Aplicando o método dos multiplicadores de Lagrange [53], a solução de (A.1), sujeita

a $\mathcal{R}\xi = \kappa$, é obtida ao resolver o sistema de equações:

$$\begin{cases} \frac{\partial}{\partial \xi} (\xi^T \Psi^T \Psi \xi - p(\mathcal{R}\xi - \kappa)) = 0 \\ \frac{\partial}{\partial p} (p(\mathcal{R}\xi - \kappa)) = 0 \end{cases} \Rightarrow \begin{cases} 2\xi^T \Psi^T \Psi + p\mathcal{R} = 0 \\ \mathcal{R}\xi - \kappa = 0. \end{cases} \quad (\text{A.4})$$

Ao isolar ξ na primeira equação do sistema (A.4) resulta:

$$\xi^T = -\frac{1}{2}p\mathcal{R}(\Psi^T \Psi)^{-1}. \quad (\text{A.5})$$

A segunda igualdade em (A.4) pode ser expressa por:

$$\xi^T \mathcal{R}^T = \kappa.$$

Com isso, de (A.5) segue que:

$$\kappa = -\frac{1}{2}p\mathcal{R}(\Psi^T \Psi)^{-1}\mathcal{R}^T \Rightarrow p = -2\kappa \left(\mathcal{R}(\Psi^T \Psi)^{-1}\mathcal{R}^T \right)^{-1}. \quad (\text{A.6})$$

Por fim, o multiplicador de Lagrange p de (A.6) é substituído em (A.5):

$$\begin{aligned} \xi^T &= \kappa \left(\mathcal{R}(\Psi^T \Psi)^{-1}\mathcal{R}^T \right)^{-1} \mathcal{R}(\Psi^T \Psi)^{-1} \\ \Rightarrow \xi &= \kappa (\Psi^T \Psi)^{-1} \mathcal{R}^T \left(\mathcal{R}^T (\Psi^T \Psi)^{-1} \mathcal{R} \right)^{-1}. \end{aligned} \quad (\text{A.7})$$

Embora haja uma expressão analítica para estimar o vetor ξ , pode haver mau condicionamento numérico ao inverter as matrizes $\Psi^T \Psi$ e $\mathcal{R}^T (\Psi^T \Psi)^{-1} \mathcal{R}$, devido à grande quantidade de parâmetros a estimar, pois $\xi \in \mathbb{R}^{n+m}$ e, em geral*, $n+m \geq 33$. Para evitar tais problemas, ao invés de implementar diretamente (A.7), aproveita-se o fato de as matrizes que precisam ser invertidas serem (por construção) simétricas e definidas positivas.

Sabe-se que uma matriz A simétrica e definida positiva pode ser fatorada em (fatoração de Cholesky [1]):

$$A = L \cdot L^T, \quad (\text{A.8})$$

*Conforme discutido na seção 3.2.3, uma *spline* cúbica é parametrizada por $n \geq 3$ e, normalmente, a ordem do modelo FIR é admitida em $n \geq 30$.

em que L é triangular inferior com as mesmas dimensões de A e os elementos positivos na diagonal. Portanto, a inversa A^{-1} é dada por:

$$A^{-1} = (L \cdot L^T)^{-1} = (L^{-1})^T L^{-1}. \quad (\text{A.9})$$

A inversão de uma matriz triangular é mais precisa e requer menos esforço computacional, pois o cálculo do determinante e dos cofatores é significativamente simplificado.

A seguir é mostrado como (A.9) pode ser implementada no Matlab[®]:

```
>> L = chol(A);  
>> invL = inv(L);  
>> invA = invL*invL';
```

A função `inv(·)` é capaz de se beneficiar da estrutura triangular de L para realizar a inversão matricial. Também deve ser mencionado que o último comando é ligeiramente diferente de (A.9) porque a função `chol(·)` retorna a transposta de L .

Apêndice B

Gráficos do limite superior de erro da plataforma HIL

Na seção 5.2 estudou-se a identificação de um processo de neutralização de pH. Os modelos do bloco dinâmico linear $\hat{G}(q)$ foram classificados de acordo com a relação entre o módulo da resposta em frequência $|G(e^{j\omega})|$ e a estimativa do limite superior do erro de modelagem $\Delta(\omega)$ e segundo o critério apresentado na Tabela 4.1. Entretanto, naquela oportunidade, os gráficos empregados na avaliação de $\hat{G}(q)$ foram omitidos para que os resultados pudessem ser discutidos de forma mais clara.

Neste apêndice são apresentados os comportamentos do módulo de $G(e^{j\omega})$ e do limite superior do erro de modelagem estimados quando se analisou:

- a degradação nos modelos devido à perturbações do processo (Figura B.1);
- o efeito da magnitude do sinal de teste (Figura B.2);
- a influência da sintonia do controlador da malha de pH (Figura B.3).

Figura B.1: Módulo de $\hat{G}(e^{j\omega})$ e limite superior do erro de modelagem dos modelos identificados durante a análise do efeito de perturbações.

Figura B.2: Comportamento de $|\hat{G}(e^{j\omega})|$ e de $\Delta(\omega)$ referente aos modelos estimados adotando sinais de teste com magnitudes distintas.

Figura B.3: Relação entre $|\hat{G}(e^{j\omega})|$ e $\Delta(\omega)$ dos modelos estimados ao variar a sintonia do controlador.