



UNIVERSIDAD NACIONAL DEL SUR

TESIS DE DOCTORA EN MATEMÁTICA

*Estructuras de Identificación basadas en  
Funciones Canónicas Lineales a Tramos*

Marcela P. Álvarez

BAHÍA BLANCA

ARGENTINA

2011



# Prefacio

Esta Tesis se presenta como parte de los requisitos para optar al grado Académico de Doctor en Matemática, de la Universidad Nacional del Sur y no ha sido presentada previamente para la obtención de otro título en esta Universidad u otra. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el ámbito del Departamento de Matemática durante el período comprendido entre octubre de 2009 y septiembre de 2011, bajo la dirección de la Dra. Liliana Raquel Castro.

Marcela Álvarez

DEPARTAMENTO DE MATEMÁTICA  
UNIVERSIDAD NACIONAL DEL SUR



UNIVERSIDAD NACIONAL DEL SUR  
Secretaría General de Posgrado y Educación Continua

La presente tesis ha sido aprobada el .../.../...,  
mereciendo la calificación de ... (.....).



*A mi hija Celeste, que supo entender “mis ausencias” de la realidad.*



## Agradecimientos

Deseo expresar mi profundo agradecimiento a la Dra. Liliana Castro por confiar en mi y haberme propuesto trabajar junto a ella. Su amistad y atención hicieron posible la realización de esta tesis.

Un reconocimiento especial al Dr. Osvaldo Agamennoni por el estímulo, los aportes y las sugerencias hechas durante el desarrollo del trabajo. Él y la Dra. Castro me invitaron a trabajar con el grupo de Control del Depto. de Ingeniería Eléctrica, que en todo momento me brindó su calidez y me hizo sentir parte del mismo.

Quiero agradecer también el apoyo brindado por las autoridades y personal administrativo del Departamento de Matemática de la UNS.

Por último, mi gratitud y todo mi amor a mi familia y a mis afectos.

Diciembre de 2011

Departamento de Matemática  
Universidad Nacional del Sur



## Resumen

Las técnicas de identificación permiten construir modelos matemáticos para sistemas dinámicos a partir de datos registrados de un experimento o del normal funcionamiento del sistema a modelar. El diseño de un modelo implica un compromiso entre su simplicidad y la necesidad de capturar los aspectos esenciales del sistema en estudio. Los modelos caja negra se diseñan enteramente a partir de los datos entrada/salida disponibles del sistema, sin tener en cuenta la interpretación de los parámetros que lo definen. Existen diferentes clases de modelos caja negra; considerando su mayor simplicidad, los primeros en desarrollarse fueron los modelos lineales. Posteriormente, dada la necesidad de modelar con mayor precisión, surgieron los modelos no lineales. Una de las principales clases de modelos no lineales de caja negra son los modelos tipo *Wiener*. Las estructuras que proponemos en esta tesis están dentro de esta familia de modelos.

Presentamos, en primer lugar, una estructura de modelo basada en funciones Canónicas Lineales a Tramos de Alto Nivel (CLATAN) y un algoritmo de identificación NOE (por sus siglas en inglés, Nonlinear Output Error). Exploramos además la capacidad de aproximación, de generalización así como también la estabilidad de este modelo. El algoritmo propuesto permite comenzar con una aproximación OE y aumentar fácilmente el orden hasta alcanzar la aproximación deseada, conservando la aproximación lograda hasta el orden inmediato anterior. Por otra parte, el algoritmo de aprendizaje para determinar los parámetros garantiza la BIBO estabilidad del modelo. Luego, proponemos dos esquemas de aproximación para los cuales probamos que permiten aproximar cualquier sistema dinámico discreto, no lineal, causal, invariante en el tiempo y con memoria evanescente. Estos modelos están compuestos por un conjunto finito de sistemas discretos de Laguerre o de Kautz, relacionados de manera no lineal mediante funciones CLATAN, cuyos parámetros ajustamos utilizando teoría de estimación con conjuntos de membresía (teoría SM). Con esta metodología, estimamos dichos parámetros asumiendo sólo que el ruido es desconocido pero acotado en alguna norma dada (ruido UBB), lo que constituye una hipótesis débil para el mismo. Por otra parte, mediante la teoría SM hallamos un conjunto que contiene todas las posibles soluciones del problema, lo que nos permite estimar las cotas de incertidumbre asociadas al problema de estimación. La metodología resultante es robusta, en el sentido que el conjunto de datos utilizado para la identificación del sistema en estudio puede ser reproducido por al menos uno de los modelos en el conjunto de parámetros identificados.



## Abstract

System identification deals with mathematical models for dynamical systems built from gathered data from experiments. The design of such models implies a trade off between simplicity and the need to capture the essential features of the system under study. Black box models are based entirely upon the available input/output data, regardless of any interpretation of the parameters involved. Due to its simplicity, linear black box models were first developed. Later, on the urge for more accurate models led to the development of non-linear ones. Wiener like-models constitute one of the most relevant classes of non-linear models. The models proposed in this Thesis belong to this class.

We first propose a model structure based on High Level Canonical Piecewise Linear (HLCPWL) functions and a Nonlinear Output Error (NOE) identification algorithm. We explore the approximation capabilities of this structure together with its generalization and stability properties. Starting from a linear Output Error (OE) approximation, this model family yields an identification algorithm such that the order of the model can be easily increased during the identification process, retaining the previously achieved approximation. The parameters of the HLCPWL functions are learned using a simple algorithm that guarantees BIBO stability of the model.

Next, we consider two approximation schemes for non linear, discrete, causal, time-invariant dynamical systems with fading memory. In these models, the dynamic linear part is represented by a finite set of Laguerre or Kautz basis functions, while the non-linear static part is realized by High Level Canonical Piecewise Linear basis functions. We estimate the parameters of the HLCPWL functions using set membership estimation theory. This theory allows to estimate the models parameters under mild conditions for the noise; in fact we only assume that the noise is unknown but bounded (UBB). We also provide a methodology for estimating the uncertainty bounds for the models and prove that this structure allows to uniformly approximate any nonlinear discrete, causal, time-invariant systems with fading memory. The proposed methodology is robust, in the sense that the data set used for the identification of the system under study can be reproduced by at least one of the models within the set of all identified parameters.



# Notación

## Conjuntos, vectores, matrices y escalares

$\mathbb{N}$	Conjunto de los números naturales.
$\mathbb{Z}$	Conjunto de los números enteros.
$\mathbb{R}$	Conjunto de los números reales.
$\mathbb{C}$	Conjunto de los números complejos.
$\mathbb{Z}^+$	Conjunto de los números enteros positivos.
$\mathbb{Z}^-$	Conjunto de los números enteros negativos.
$\mathbb{Z}_0^+$	Conjunto de los números enteros positivos o cero.
$\mathbb{Z}_0^-$	Conjunto de los números enteros negativos o cero.
$\mathbb{R}^+$	Conjunto de los números reales positivos.
$\mathbb{R}^-$	Conjunto de los números reales negativos.
$\mathbb{R}_0^+$	Conjunto de los números reales positivos o cero.
$\mathbb{R}_0^-$	Conjunto de los números reales negativos o cero.
$\mathbb{Z}^n$	Conjunto de las $n$ -uplas de números enteros.
$\mathbb{R}^n$	Conjunto de las $n$ -uplas de números reales.
$\bar{\gamma} \in \mathbb{C}$	Conjugado de un número $\gamma \in \mathbb{C}$ .
$\operatorname{Re}(\cdot)$	Parte real de un número complejo.
$\operatorname{Im}(\cdot)$	Parte imaginaria de un número complejo.
$\mathbf{u}$	Vector en $\mathbb{R}^n$ .
$\mathbf{u}(k)$	Componente $k$ -ésima de un vector $\mathbf{u}$ en $\mathbb{R}^n$ .
$(\cdot)^T$	Vector o matriz traspuesta.

## Espacios de funciones

$\ell^2(\mathcal{J})$	Conjunto de las sucesiones de cuadrado sumable, definidas en el conjunto $\mathcal{J}$ .
$u(k)$	Componente $k$ -ésima de una sucesión $u$ de $\ell^2(\mathcal{J})$ .
$C^p(A)$	Espacio de funciones continuamente diferenciables de hasta orden $p$ en el conjunto $A$ , con la norma del supremo. Si no se especifica $A$ , es $A = \mathbb{R}$ .

$C^\infty(A)$	Espacio de funciones continuamente diferenciables de cualquier orden en el conjunto $A$ , con la norma del supremo. Si no se especifica $A$ , es $A = \mathbb{R}$ .
$L^p(A), 1 \leq p < \infty$	Espacio de funciones medibles Lebesgue definidas en un conjunto $A$ tales que la potencia de orden $p$ de su módulo es integrable en $A$ .
$H^p(T)$	Espacio de Hardy de funciones complejas definidas en $T$ .

## Normas

$\ell^\infty, \ell^2, \ell^1$	Normas usuales en los espacios de sucesiones $\ell^\infty(\mathbb{N})$ , $\ell^2(\mathbb{N})$ y $\ell^1(\mathbb{N})$ , respectivamente.
$\infty, 2, 1$	Normas vectoriales en espacios de dimensión finita.

# Índice general

<b>Índice de Figuras</b>	<b>XIX</b>
<b>Índice de Tablas</b>	<b>XXIII</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Objetivos . . . . .	2
1.3. Breve Descripción de cada Capítulo . . . . .	3
<b>2. Modelado e Identificación de Sistemas Dinámicos</b>	<b>5</b>
2.1. Sistemas y Modelos . . . . .	6
2.1.1. Clasificación de los Sistemas y de los Modelos . . . . .	7
2.2. Proceso de Identificación . . . . .	10
2.3. Modelos Caja Negra . . . . .	11
2.3.1. Estructura General de los Modelos Caja Negra . . . . .	12
2.4. Identificación de Sistemas Lineales . . . . .	14
2.4.1. Modelo de Respuesta Finita al Impulso (FIR) . . . . .	16
2.4.2. Modelo de Error de Salida (OE) . . . . .	17
2.4.3. Modelo Autorregresivo con Entradas Exógenas (ARX) . . . . .	17
2.4.4. Modelo Autorregresivo de Media Móvil con Entradas Exógenas (AR- MAX) . . . . .	18
2.4.5. Representación con Bases de Funciones Ortonormales . . . . .	19
2.4.6. Algunas Bases Ortonormales en $H^2(T)$ . . . . .	22
2.5. Identificación de Sistemas no Lineales . . . . .	26

2.5.1.	Elección de los Regresores . . . . .	27
2.5.2.	Elección de la Transformación no Lineal . . . . .	28
2.6.	Modelos no Lineales Orientados a Bloques . . . . .	31
2.7.	Representaciones Canónicas Lineales a Tramos de Alto Nivel . . . . .	32
2.7.1.	Introducción . . . . .	32
2.7.2.	Aproximación de una Función Continua Conocida . . . . .	39
2.7.3.	Aproximación de Funciones a partir de un Conjunto de Datos . . . . .	40
2.7.4.	Aproximación Óptima utilizando Norma Infinito . . . . .	41
2.7.5.	Aproximación Óptima utilizando Norma Cuadrática . . . . .	42
2.7.6.	Aproximación Robusta utilizando Norma Infinito . . . . .	42
2.8.	Validación del Modelo . . . . .	44
2.9.	Conclusiones . . . . .	45
<b>3.</b>	<b>Identificación de un Modelo NOE usando Funciones CLATAN</b>	<b>47</b>
3.1.	Estructura del Modelo . . . . .	48
3.2.	Algoritmo de Identificación . . . . .	49
3.2.1.	Estabilidad BIBO del Modelo . . . . .	52
3.3.	Ejemplos de Aplicación . . . . .	53
3.3.1.	Ejemplo 1: Modelo Logístico NOE . . . . .	54
3.3.2.	Ejemplo 2 . . . . .	56
3.3.3.	Ejemplo 3 . . . . .	58
3.3.4.	Ejemplo 4 . . . . .	60
3.4.	Conclusiones . . . . .	60
<b>4.</b>	<b>Estructuras de Identificación tipo Wiener</b>	<b>65</b>
4.1.	Diferentes Modelos de Aproximación E/S . . . . .	65
4.1.1.	Métodos de las Series de Volterra y de Wiener . . . . .	66
4.1.2.	Modelos Generales Orientados a Bloques . . . . .	68
4.1.3.	Modelos Orientados a Bloques en Paralelo . . . . .	69
4.2.	Estructuras Laguerre-CLATAN y Kautz-CLATAN . . . . .	70
4.2.1.	Estructura Laguerre-CLATAN . . . . .	72

4.2.2.	Estructura Kautz-CLATAN . . . . .	72
4.2.3.	Metodología de Modelado . . . . .	73
4.2.4.	Identificación del Bloque Lineal Dinámico . . . . .	74
4.3.	Identificación de la Parte no Lineal Estática . . . . .	76
4.3.1.	Teoría de Estimación con Conjuntos de Membresía . . . . .	77
4.3.2.	Identificación del Bloque no Lineal utilizando Teoría SM . . . . .	87
4.4.	Propiedades de Aproximación de las Estructuras Laguerre-CLATAN y Kautz-CLATAN . . . . .	89
4.4.1.	Sistemas con Memoria Evanescente . . . . .	90
4.4.2.	Teorema de Aproximación . . . . .	96
4.5.	Ejemplos de Aplicación . . . . .	98
4.5.1.	Ejemplo 1: Sistema SISO no Lineal Discreto . . . . .	100
4.5.2.	Ejemplo 2: Sistema Mecánico con Saturación en la Entrada . . . . .	103
4.5.3.	Ejemplo 3: Modelo MIMO de una Unidad de Generación de Vapor . . . . .	106
4.6.	Conclusiones . . . . .	111
<b>5.</b>	<b>Identificación Robusta de los Modelos Laguerre-CLATAN y Kautz-CLATAN</b>	<b>113</b>
5.1.	Introducción . . . . .	113
5.2.	Formulación del Problema . . . . .	114
5.3.	Modelo de Incertidumbre . . . . .	117
5.3.1.	Caso MIMO . . . . .	118
5.4.	Ejemplos de Aplicación . . . . .	122
5.4.1.	Modelo de Incertidumbre del Sistema Masa-Resorte Amortiguado . . . . .	122
5.4.2.	Modelo de incertidumbre de un Sistema SISO no Lineal Discreto . . . . .	124
5.4.3.	Modelo de Incertidumbre de la Unidad de Generación de Vapor . . . . .	124
5.5.	Conclusiones . . . . .	125
	<b>Aportes y Trabajo Futuro</b>	<b>131</b>
	<b>Bibliografía</b>	<b>135</b>
<b>A.</b>	<b>Algoritmo de Actualización para el Modelo NOE-CLAT</b>	<b>143</b>

<b>B. Método de los Coeficientes de Lipschitz</b>	<b>145</b>
<b>C. Demostraciones Complementarias del Capítulo 4</b>	<b>149</b>

# Índice de figuras

2.1. Modelo FIR . . . . .	16
2.2. Modelo OE . . . . .	17
2.3. Modelo ARX . . . . .	18
2.4. Modelo ARMAX . . . . .	19
2.5. Modelo tipo Hammerstein . . . . .	31
2.6. Modelo tipo Wiener . . . . .	31
2.7. Función LAT con dominio $\mathbf{S} \subset \mathbb{R}^2$ . . . . .	35
2.8. Configuración de fronteras para el hipercubo $[-1, 1]^2$ . . . . .	37
2.9. Funciones con $n.a. = 1$ (a,b,c,d) y $n.a. = 2$ (e,f,g,h) en $\mathbb{R}^2$ . . . . .	38
3.1. Modelo NOE-CLATAN . . . . .	49
3.2. Características del error en el Ejemplo 3.3.1 . . . . .	55
3.3. $f_{lat}$ resultante del Ejemplo 3.3.1 con $ndiv = 2, 4$ y $8$ . . . . .	56
3.4. Performance del error en el Ejemplo 3.3.2. Errores de aproximación (muestras 1 a 1800) y validación (muestras 1801 a 2000) . . . . .	57
3.5. $f_{lat}$ resultante del Ejemplo 3.3.2 con $ndiv = 2, 4$ y $8$ . . . . .	58
3.6. Función $f_{lat}$ resultante para el Ejemplo 3.3.3 . . . . .	59
3.7. Características del error en el Ejemplo 3.3.4 para el caso con ruido . . . . .	61
3.8. $f_{lat}$ para el Ejemplo 3.3.4 . . . . .	62
4.1. Representación del esquema de Wiener discreto . . . . .	68
4.2. Modelo general de bloques en cascada . . . . .	69
4.3. Representación del esquema de Korenberg y Paarman . . . . .	69
4.4. Estructura Laguerre-CLATAN y Kautz-CLATAN . . . . .	71
4.5. Gráfico ilustrativo de un problema de estimación genérico . . . . .	80

4.6.	Aproximación exterior del conjunto $FSS_y$ por medio de ortotopos . . . . .	85
4.7.	Sucesiones $u$ , $u_e$ y $Pu_e$ . . . . .	91
4.8.	Sucesiones causal, extendida y sus respuestas asociadas . . . . .	92
4.9.	Reconstrucción de $f$ a partir de la funcional asociada $T$ . . . . .	92
4.10.	Sistema con memoria evanescente . . . . .	94
4.11.	Validación del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)	102
4.12.	Validación del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)	103
4.13.	Resultados sobre el conjunto de datos de prueba del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo) . . . . .	103
4.14.	Resultados sobre el conjunto de datos de prueba del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo) . . . . .	104
4.15.	Ejemplo 2. Sistema masa-resorte amortiguado . . . . .	105
4.16.	Validación del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo) .	107
4.17.	Validación del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo) .	107
4.18.	Resultados sobre el conjunto de datos de prueba del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo) . . . . .	108
4.19.	Resultados sobre el conjunto de datos de prueba del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo) . . . . .	108
4.20.	Salidas de la unidad de generación de vapor. $\mathbf{y}_1$ : presión, $\mathbf{y}_2$ : nivel . . . . .	109
4.21.	Modelo Laguerre-CLATAN. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo) . . . . .	110
4.22.	Modelo Laguerre-CLATAN. Sistema ( $\mathbf{y}_2$ : azul), modelo ( $\tilde{\mathbf{y}}_2$ : rojo) . . . . .	111
4.23.	Modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo) . . . . .	111
4.24.	Modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $\mathbf{y}_2$ : azul), modelo ( $\tilde{\mathbf{y}}_2$ : rojo) . . . . .	112
5.1.	Estructura tipo Wiener SISO. . . . .	119
5.2.	Cotas de incertidumbre para el modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde). . . . .	123
5.3.	Cotas de incertidumbre para el modelo Kautz-CLATAN sobre el conjunto de validación. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde). . . . .	123
5.4.	Cotas de incertidumbre para el modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde). . . . .	124

5.5.	Cotas de incertidumbre para el modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde). . . . .	125
5.6.	Cotas de incertidumbre para la salida $\mathbf{y}_1$ con ruido del modelo Laguerre-CLATAN. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo), cotas (verde). . . . .	126
5.7.	Cotas de incertidumbre para la salida $\mathbf{y}_2$ con ruido del modelo Laguerre-CLATAN. Sistema ( $\mathbf{y}_2$ : azul), modelo ( $\tilde{\mathbf{y}}_2$ : rojo), cotas (verde). . . . .	126
5.8.	Cotas de incertidumbre para la salida $\mathbf{y}_1$ del modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo), cotas (verde).	127
5.9.	Cotas de incertidumbre para la salida $\mathbf{y}_2$ del modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $\mathbf{y}_2$ : azul), modelo ( $\tilde{\mathbf{y}}_2$ : rojo), cotas (verde).	127
5.10.	Cotas de incertidumbre para el modelo Laguerre-CLATAN sobre un nuevo conjunto de datos. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo), cotas (verde). . . .	128
5.11.	Cotas de incertidumbre para el modelo Laguerre-CLATAN sobre un nuevo conjunto de datos. Sistema ( $\mathbf{y}_2$ : azul), modelo ( $\tilde{\mathbf{y}}_2$ : rojo), cotas (verde). . . .	128



# Índice de tablas

2.1. Funciones con $n.a.=1$ y $n.a.=2$ y su comportamiento en un dominio perteneciente a $\mathbb{R}^2$ . . . . .	37
4.1. Estructura del conjunto $FPS_y$ según la $Y$ -norma, caso lineal . . . . .	81
4.2. Error $RMSE$ y coeficiente de correlación $FIT$ para el Ejemplo 1 . . . . .	102
4.3. Error $RMSE$ y coeficiente de correlación $FIT$ para el Ejemplo 2 . . . . .	106
4.4. Variables de la UGVD . . . . .	109



# Capítulo 1

## Introducción

### 1.1. Motivación

La identificación de sistemas dinámicos consiste en describir el comportamiento de un proceso de la vida real mediante una formulación matemática. Esto posibilita, posteriormente, el análisis y estudio del mismo fuera de línea o el desarrollo de mecanismos que modifiquen su comportamiento original, a fin de satisfacer necesidades o especificaciones de operación concretas del proceso real.

Cuanto más ajustado a la realidad es un modelo, mejores son las acciones de control, análisis, predicción, etc., que se pueden llevar a cabo. Luego el estudio de técnicas de identificación de sistemas constituye una rama científica de suma importancia, especialmente en lo que se refiere a la búsqueda de nuevas formulaciones matemáticas que posibiliten una mejor descripción del sistema en estudio.

Si bien para un gran porcentaje de procesos reales es suficiente un enfoque de identificación lineal, en muchos otros casos, los modelos lineales son insuficientes para realizar un diseño que cumpla con las especificaciones requeridas. Estos casos pueden corresponder a sistemas con dinámicas altamente no lineales y grandes anchos de banda. Por ello, es necesario desarrollar modelos y metodologías de identificación no lineales de complejidad reducida, con la finalidad de capturar aquellos aspectos que con modelos lineales no es posible y para posibles aplicaciones en el control de los procesos. En el caso de modelos no lineales de naturaleza recursiva, una estructura de identificación NOE (por sus siglas en inglés, Nonlinear Output Error) permite evaluar modelos dinámicos aptos para su utilización como simuladores fuera de línea, ya que se minimiza el efecto del ruido en la estimación de los parámetros del mismo. Pero en este caso, la evaluación del orden se hace sumamente difícil debido a la complejidad computacional inherente a la naturaleza recursiva del modelo. Luego, es de interés desarrollar alguna estrategia de identificación sencilla de implementar para abordar el modelado mediante estructuras NOE.

Asimismo, es relevante determinar cómo se comportan los modelos dinámicos y definir qué método de estimación de parámetros de los mismos convendría emplear. En este sentido, surge la posibilidad de aplicar la teoría de estimación basada en la utilización de conjuntos de membresía (Set Membership Theory, SM en forma abreviada) en la identificación de sistemas, con el fin de ajustar los parámetros de los modelos. Es en esta línea, y con este punto de vista, que se desarrolla el trabajo de esta tesis.

## 1.2. Objetivos

Los modelos matemáticos de sistemas dinámicos son utilizados en distintas áreas para describir la evolución de procesos reales de distinta índole. Entre algunas de sus aplicaciones, podemos mencionar el estudio de las características más relevantes del comportamiento de un sistema, la predicción de su comportamiento futuro en áreas como economía y control de procesos, el diagnóstico de fallas en sistemas complejos, el diseño de sistemas de control que requieren un modelo del sistema a controlar sobre el cual trabajar o la simulación de sistemas y entrenamiento de operadores, que permite estudiar y analizar situaciones anómalas que no pueden ser tratadas sobre sistemas reales. En este contexto, el propósito de un modelo es capturar cualitativa y cuantitativamente las principales características de un sistema dinámico.

Cuando es posible construir un modelo a partir de los principios que gobiernan el funcionamiento del sistema, se obtienen modelos de caja blanca. Sin embargo, cuando no se cuenta con suficiente información de los principios fenomenológicos que gobiernan al sistema en estudio, o este contiene una gran cantidad de elementos que lo componen, la única alternativa posible es utilizar técnicas de caja negra.

En este trabajo los objetivos son:

- Desarrollar un algoritmo de identificación para el modelo realimentado de caja negra NOE (por sus siglas en inglés, Nonlinear Output Error), basado en el método de aproximación con funciones Canónicas Lineales a Tramos de Alto Nivel (CLATAN).
- El objetivo de este algoritmo es lograr la identificación y simulación de un modelo no lineal para aplicaciones que requieren capacidad de procesamiento rápido y gran cantidad de datos. Cabe destacar que ya se ha realizado la implementación en hardware de las bases de aproximación LAT [77].
- Obtener modelos de caja negra de sistemas no lineales con memoria evanescente, que pueden considerarse dentro de las estructuras de aproximación tipo Wiener, encontrando el mejor compromiso entre su simplicidad y la necesidad de capturar los aspectos esenciales del sistema en estudio.
- Analizar la convergencia y estabilidad de algoritmos de identificación para modelos dinámicos tipo NOE.

- Estimar los parámetros de los esquemas de aproximación tipo Wiener propuestos mediante el método de estimación basado en la utilización de conjuntos de membresía.

## 1.3. Breve Descripción de cada Capítulo

En este trabajo hemos incluido los conceptos necesarios para su lectura, de manera de hacerlo autocontenido. La tesis está organizada de la siguiente manera.

En el Capítulo 2 hacemos una breve revisión de los conceptos y consideraciones generales sobre sistemas y modelos. Describimos las estructuras de modelo tipo caja negra, que son las que utilizamos en esta tesis, y las bases de representación de funciones Lineales a Tramos (LAT) continuas. Este esquema de aproximación permite representar sistemas no lineales estáticos, y lo utilizamos en el Capítulo 4 como parte de la formulación de los modelos de aproximación de sistemas no lineales que proponemos. Presentamos luego los teoremas que aseguran su capacidad de aproximación de funciones no lineales estáticas, que son fundamentales para establecer el teorema de aproximación propuesto en esta tesis.

Los Capítulos 3 y 4 conforman la parte central de este trabajo de tesis. En el Capítulo 3 describimos la estructura de modelo NOE-CLATAN y el correspondiente algoritmo de identificación. Luego analizamos las capacidades de aproximación de esta nueva estructura, su grado de generalización y obtenemos condiciones suficientes que aseguran la estabilidad BIBO (de sus siglas en inglés, Bounded Input Bounded Output) del modelo propuesto. En la última sección presentamos ejemplos de aplicación de la metodología que muestran el potencial de nuestro enfoque. Finalizamos el capítulo con algunas conclusiones acerca de las ventajas que proporciona este algoritmo de identificación basado en el método de aproximación con funciones CLATAN.

En el Capítulo 4 definimos las estructuras de identificación Laguerre-CLATAN y Kautz-CLATAN, que pueden considerarse dentro de las estructuras de aproximación tipo Wiener y presentamos una síntesis de la teoría de estimación con conjuntos de membresía (SM), que es la que utilizamos para ajustar los parámetros de los modelos mencionados. Demostramos que estas estructuras de modelado son capaces de aproximar sistemas que pueden caracterizarse como no lineales, discretos, causales, invariantes en el tiempo y que poseen memoria evanescente. Concluimos el capítulo con tres ejemplos de aplicación, cada uno de ellos con diferentes no linealidades, que ilustran las capacidades de aproximación de los modelos propuestos y las ventajas de considerar sistemas de Laguerre o de Kautz, para la identificación de los sistemas considerados, de acuerdo a sus características. A través del último ejemplo, mostramos que es posible extender el método de identificación presentado para sistemas MIMO (por sus siglas en inglés, Multiple-Input, Multiple-Output), aún cuando resta formular una demostración formal de sus capacidades de aproximación.

En el Capítulo 5 desarrollamos una metodología para identificar la incertidumbre de los modelos Laguerre-CLATAN y Kautz-CLATAN. La estructura de modelo nominal está dada

por la representación paramétrica Laguerre-CLATAN o Kautz-CLATAN, ambas presentadas en el Capítulo 4. El método se basa en la identificación de cotas de los parámetros de dichos modelos utilizando teoría de estimación SM. Por lo tanto, describimos la incertidumbre del modelo como un conjunto de parámetros para el bloque no lineal estático, cuyos valores se obtienen resolviendo un problema de optimización. La metodología resultante es robusta dado que la totalidad de los datos utilizados para la identificación del sistema en estudio, puede ser reproducido por al menos uno de los modelos en el conjunto de parámetros identificados. Ilustramos el algoritmo de identificación desarrollado con tres ejemplos, utilizando los modelos nominales obtenidos en el Capítulo 4, Sección 4.5. Con el último ejemplo mostramos que es posible aplicar el método propuesto para la identificación robusta de modelos MIMO.

Finalmente, a continuación del Capítulo 5, damos las conclusiones y una propuesta de futuras actividades de investigación.

Se incluyen tres Apéndices que contienen aspectos complementarios necesarios para el mejor seguimiento de la tesis. En el Apéndice A incluimos el algoritmo de actualización para estimar los parámetros de los sucesivos modelos de la estructura NOE-CLATAN presentada en el Capítulo 3. En el Apéndice B describimos el algoritmo utilizado para seleccionar el orden del modelo lineal utilizando el coeficiente de Lipschitz. El Apéndice C contiene las demostraciones de los lemas necesarios para la demostración del teorema de aproximación que enunciamos y probamos en el Capítulo 4.

La bibliografía está organizada por orden alfabético y puede consultarse antes de los Apéndices.

## Capítulo 2

# Modelado e Identificación de Sistemas Dinámicos

Los modelos matemáticos de sistemas dinámicos son de gran importancia en el ámbito de la ciencia y la tecnología. Son utilizados en distintas áreas para describir la evolución de procesos de distinta índole como, por ejemplo, químicos, aeronáuticos, nucleares, térmicos, hidráulicos, biológicos y ecológicos. Entre algunas de sus aplicaciones, podemos mencionar el estudio de las características más relevantes del comportamiento de un sistema, la predicción de su comportamiento futuro en áreas como economía y control de procesos, el diagnóstico de fallas en sistemas complejos, el diseño de sistemas de control que requieren un modelo del sistema a controlar sobre el cual trabajar, la simulación de sistemas y entrenamiento de operadores, que permite estudiar y analizar situaciones anómalas que no pueden ser tratadas sobre sistemas reales.

La construcción de un modelo matemático que describa adecuadamente la evolución dinámica de un sistema, puede realizarse a partir de los principios o leyes que lo rigen, o mediante la utilización de datos entrada/salida obtenidos experimentalmente o de la propia evolución del mismo. Estas dos construcciones se conocen, respectivamente, como *modelado* e *identificación* de sistemas dinámicos. En general, en el contexto de la presente tesis asumiremos que en el caso de contar con datos de entrada/salida, los mismos provienen de la realización de un experimento donde tenemos control sobre la variable de entrada que se introduce.

En el modelado, los mecanismos internos de funcionamiento del sistema son conocidos y puede construirse un modelo a partir de las leyes que lo gobiernan. A diferencia de éste, en la identificación, las leyes no son conocidas y se debe construir el modelo matemático a partir de mediciones de las excitaciones que son utilizadas durante el funcionamiento del sistema y las respuestas correspondientes a ellas, es decir, basándonos en las entradas y salidas observadas del sistema. Estas dos construcciones se suelen emplear de forma conjunta ya que se complementan adecuadamente.

El término identificación de sistemas fue acuñado por Lofti Zadeh en 1962 [107] como:

*Identification is the determination on the basis of inputs and outputs, of a system within a specified class of systems, to which the system under test is equivalent.*<sup>1</sup>

Podemos decir que la identificación de sistemas quedó establecida como un campo de investigación reconocido dentro del área de control automático, a mediados de los sesenta en el tercer congreso de la International Federation of Automatic Control (IFAC por sus siglas en inglés) en Londres, en el que fue presentado un artículo de visión general sobre identificación de sistemas [31]. Un año después era organizado en Praga el primer Symposium IFAC sobre identificación de sistemas.

El inicio de las técnicas de identificación aplicadas a procesos con una entrada y una salida tienen su origen a principios de los años setenta [3]. Sin embargo, sólo a fines de la década de los noventa comienzan a aplicarse de manera intensiva a procesos industriales, siendo algunas de ellas útiles para el estudio de sistemas de múltiple entrada-múltiple salida.

Una manera adecuada de introducirse en el estudio sobre los distintos aspectos teóricos y prácticos de la identificación de sistemas es a través de los libros de Ljung [58] y de Söderström y Stoica [88], donde también se pueden encontrar numerosas referencias.

En este capítulo presentamos, en primer lugar, los conceptos y consideraciones generales sobre sistemas y modelos. Seguidamente describimos la clase de modelos que se utiliza en la identificación de sistemas dinámicos lineales y no lineales, poniendo énfasis en las estructuras de modelo tipo caja negra, que son las que utilizaremos en esta tesis.

## 2.1. Sistemas y Modelos

Un *sistema dinámico* es un sistema que presenta un cambio en su estado o evoluciona con el tiempo. Para describir los sistemas dinámicos debemos considerar las variables que intervienen en el sistema y los espacios sobre los cuales están definidas dichas variables. Llamamos *salidas* o *medidas disponibles* a las señales observables producidas por el sistema que son de nuestro interés. A las variables que pueden ser manipuladas o “controladas” para afectar al sistema las llamamos *entradas* o *excitaciones*.

Algunos ejemplos de sistemas dinámicos son los sistemas mecánicos como un vehículo espacial, una red eléctrica, un reactor atómico o una columna de destilación. Pero puede ser también la economía de un país o el cuerpo humano. En este último ejemplo, algunas de las variables de entrada al sistema podrían ser los alimentos, la luz, los ejercicios, los medicamentos y el ruido. Las variables de salida o respuestas del sistema pueden ser la temperatura corporal, la presión, el sueño, el dolor y la composición sanguínea. Si enfocamos el proceso de identificación sobre algún aspecto de la economía de un país, entonces podemos considerar como variables de entrada los precios, los salarios, los impuestos, las tasas de

---

<sup>1</sup>Identificación es la determinación, basada en entradas y salidas, de un sistema dentro de una clase específica de sistemas en la cual existe uno equivalente al sistema bajo testeo.

interés, las leyes de consumo e impuestos a la exportación; y como variables de salida la producción, los precios (si no son variables sujetas a control), el producto bruto interno o PBI, la balanza de pagos y el consumo.

Un *modelo* es una representación matemática de un sistema o proceso dinámico que permite analizar, describir, simular, controlar y predecir ese proceso.

Para obtener un modelo se hacen experimentos previos con el sistema, denominado *proceso de experimentación*, que ponen de manifiesto las relaciones causa-efecto del mismo. Es importante que la entrada al sistema durante la toma de datos excite todos los modos propios del sistema, para obtener una señal de salida con suficiente información que permita caracterizar su dinámica. Este proceso permite luego postular hipótesis del sistema real, de modo que lo que se quiere estudiar esté suficientemente plasmado en la representación. En los casos en los cuales no es posible experimentar con el sistema, registramos los valores de las variables de interés, basados en la evolución del mismo.

En general, el grado de exactitud de un modelo depende de su aplicación futura, pues ningún modelo es lo suficientemente general como para describir todos los aspectos del comportamiento del sistema bajo estudio. Los modelos obtenidos resultarán adecuados sólo para reproducir aceptablemente determinados comportamientos del sistema, bajo ciertas condiciones y dentro de un rango de operaciones dado. Es decir, el sistema idealizado dependerá no sólo del sistema real en sí, sino también del problema a resolver y del intervalo de validez que se pretenda tener para el modelo resultante. Luego, cada modelo elegido tendrá validez siempre que se respeten las condiciones bajo las cuales se realizó el registro de los datos, que son las que definen su *dominio de aplicación o validez*.

De acuerdo a las cualidades del sistema estudiado, a la exactitud requerida y a las herramientas de modelado utilizadas, los modelos tendrán diferentes características, pero todos buscan asociar lo observado con un patrón que lo explique.

### 2.1.1. Clasificación de los Sistemas y de los Modelos

Hay varias formas de clasificar los modelos matemáticos. De acuerdo al número de entradas y salidas del sistema se denominan:

*Modelos SISO (Single-Input, Single-Output por sus siglas en inglés)*. Son modelos con una sola entrada exógena y una sola salida.

*Modelos MIMO (Multiple-Input, Multiple-Output por sus siglas en inglés)*. Son modelos con múltiples entradas y múltiples salidas.

Las señales o variables de los sistemas dinámicos son, en general, función del tiempo. De acuerdo con esto, los modelos que los representan son:

*Modelos de sistemas continuos*. Operan con señales de entrada y salida de naturaleza continua, las cuales están definidas para todo instante de tiempo (temperatura, velocidad, espacio,

etc.). Los sistemas dinámicos de tiempo continuo se expresan por ecuaciones diferenciales ordinarias (EDO), ecuaciones diferenciales en derivadas parciales (EDP) y ecuaciones diferenciales con retardo (EDR).

*Modelos de sistemas discretos.* Se opera con señales disponibles únicamente en instantes de tiempo discretos. Los sistemas dinámicos discretos están descritos por ecuaciones en diferencias (ED) también conocidas como mapas iterados.

En este caso se deben distinguir las señales de naturaleza discreta (interés bancario, producción de una fábrica de zapatos), y las señales muestreadas, que son señales de naturaleza continua, pero que únicamente se dispone de ellas en instantes de tiempo discreto.

*Modelos invariantes o variantes en el tiempo.* Si el sistema es invariante en el tiempo TI (por sus siglas en inglés, Time Invariant), la relación de las variables con los parámetros es constante. Luego, la respuesta a una cierta entrada no depende del tiempo en que ésta se aplique. Los coeficientes del modelo matemático TI son constantes.

Si el sistema es variante en el tiempo, entonces la relación de las variables con los parámetros depende del tiempo y los coeficientes del modelo matemático son función del tiempo.

*Modelos entrada/salida (E/S).* En estos modelos sólo es de interés la relación existente entre las entradas y las salidas al sistema.

*Modelos entrada-estados-salida.* Son modelos que describen el estado completo del sistema. El estado de un sistema dinámico se define como la menor colección de variables reales cuyo valor en un determinado instante resume el pasado dinámico del sistema y es suficiente para predecir su evolución futura. La descripción interna establece entre las señales de entrada y de salida una relación indirecta, la cual se formaliza a través de las ecuaciones de estado y de salida. La forma más general de representación por variables de estado de un sistema continuo está dada por dos ecuaciones. La primera define los cambios de las variables de estado en función de estas mismas variables, las entradas y el tiempo; y la segunda define la salida en función de las variables de estado, las entradas y el tiempo. Es decir

$$\begin{aligned}\dot{\mathbf{x}}(t) &= f(t, \mathbf{x}(t), \mathbf{u}(t)) \\ \mathbf{y}(t) &= g(t, \mathbf{x}(t), \mathbf{u}(t)),\end{aligned}\tag{2.1}$$

donde  $\mathbf{u}(t) \in \mathbb{R}^p$  es el vector de entradas al sistema, siendo  $p = 1$  en el caso SISO;  $\mathbf{y}(t) \in \mathbb{R}^q$  es el vector de salidas del sistema, donde  $q = 1$  para el caso SISO;  $f$  es la función de transición que permite determinar la evolución del vector de estados en un determinado instante a partir del conocimiento del estado actual y de las entradas  $\mathbf{u}(t)$ ;  $\mathbf{x}(t) \in \mathbb{R}^n$  es el vector de estados con todas las señales relacionadas con la energía del sistema, como por ejemplo, posiciones y velocidades de masas, cargas de condensadores o temperaturas. Si el sistema es invariante en el tiempo, las funciones  $f$  y  $g$  no dependen explícitamente del tiempo.

El conocimiento del vector de estados en un determinado instante  $t_0$ , junto con las entradas aplicadas al sistema a partir de ese momento, permite conocer el estado y la salida del sistema en cualquier instante posterior  $t$ . El *espacio de estados* de un proceso se define como el conjunto que contiene a todos los posibles valores del vector de estados, y resulta ser un espacio vectorial  $n$ -dimensional.

De manera similar, podemos describir en forma genérica un sistema discreto por la siguiente representación en variables de estado

$$\begin{aligned}\mathbf{x}(k+1) &= f(k, \mathbf{x}(k), \mathbf{u}(k)) \\ \mathbf{y}(k) &= g(k, \mathbf{x}(k), \mathbf{u}(k)),\end{aligned}\tag{2.2}$$

donde  $\mathbf{u}(k) \in \mathbb{R}^p$  es el vector de entradas al sistema, siendo  $p = 1$  en el caso SISO;  $\mathbf{y}(k) \in \mathbb{R}^q$  es el vector de salidas del sistema, donde  $q = 1$  para el caso SISO;  $f$  es la función de transición que permite determinar la evolución del vector de estados en un determinado instante a partir del conocimiento del estado actual y de las entradas  $\mathbf{u}(k)$ ;  $\mathbf{x}(k) \in \mathbb{R}^n$  es el vector de estados con todas las señales relacionadas con la energía del sistema. Si el sistema es invariante en el tiempo, las funciones  $f$  y  $g$  no dependen explícitamente del tiempo.

*Modelos lineales y no lineales.* Los modelos lineales expresan relaciones lineales entre sus variables. En este caso se verifica la propiedad de superposición  $f(ax_1+bx_2) = af(x_1)+bf(x_2)$  de efectos.

En los modelos no lineales, la relación entre las variables es no lineal, no siendo válido el principio de superposición de efectos.

*Modelos de sistemas estáticos o dinámicos.* Si el sistema es estático, la salida actual del sistema sólo es función de la entrada actual,  $\mathbf{y}(t) = f(\mathbf{u}(t))$ . Los sistemas estáticos se representan mediante ecuaciones algebraicas lineales y/o no lineales.

Si el sistema es dinámico, la salida es función de la entrada actual y del estado del sistema. Este tipo de sistemas suele representarse a partir de ecuaciones diferenciales o en diferencias.

Según la posibilidad de predecir el comportamiento de un sistema, es decir su respuesta, se clasifican en:

*Modelos de sistemas determinísticos o estocásticos.* Se clasifican en función de la naturaleza de sus señales.

*Señales determinísticas.* Se pueden representar con una expresión matemática explícita; de esta forma los valores futuros son perfectamente predecibles. En este caso el sistema se denomina determinístico.

*Señales estocásticas.* Las señales implicadas son aleatorias; no es posible describirlas analíticamente con una expresión explícita como en el caso anterior. Sin embargo, cuando una señal estocástica se observa durante un largo período de tiempo puede verse cierta regularidad y puede ser descripta en términos de probabilidades y promedios estadísticos. Los

valores futuros de una señal aleatoria no se pueden predecir con exactitud, sólo se pueden basar en los promedios de conjuntos de señales con características similares. En este caso el sistema es estocástico.

*Modelos de sistemas de parámetros concentrados.* El sistema es descrito por un conjunto de ecuaciones diferenciales a derivadas totales. Sus parámetros se expresan mediante un conjunto finito o *concentrado* de valores.

*Modelos de sistemas de parámetros distribuidos.* El sistema es representado por un conjunto de ecuaciones diferenciales en derivadas parciales. Los parámetros de estas ecuaciones se expresan mediante un conjunto infinito o *distribuido* de valores.

*Modelos causales y no causales.* Los modelos causales responden a una relación de causa-efecto, *i.e.*, las salidas del modelo dependen solamente de su evolución previa y no de eventos futuros. En caso contrario, se denominan modelos no causales.

## 2.2. Proceso de Identificación

Si bien la identificación de un sistema dinámico no consiste en seguir sistemáticamente una serie de etapas, ellas se presentan de modo natural en la literatura [30, 43, 58, 88]. Las etapas en el proceso de identificación son las siguientes:

1. *Recopilación de información.* Esta primera etapa consiste en la reunión de información, basada generalmente en el conocimiento previo del sistema. De acuerdo a cuan acabado sea el conocimiento del sistema, la información conduce a modelos de caja negra, blanca o gris.
2. *Selección de una clase de señales.* Es importante considerar una clase de señales o variables de entrada para la adquisición de información experimental, usualmente a través del conocimiento *a posteriori* del sistema. La clase de señales a emplear debe ser lo suficientemente rica en componentes frecuenciales como para excitar adecuadamente al sistema y obtener una señal de salida con suficiente información para caracterizar la dinámica del mismo. El diseño de experimentos para el registro de estos datos involucra, entre otros, la selección de las señales a medir, tiempos de muestreo, el tipo de señales de entrada.
3. *Elección de la estructura del modelo.* A continuación, es necesario determinar la clase de modelos a emplear. Se debe tener en cuenta la complejidad del proceso real y la estimación de parámetros, así como el uso final que se desee dar al modelo.
4. *Estimación de parámetros.* Una vez que se tiene la estructura del modelo y los datos experimentales, el paso siguiente es encontrar los parámetros del modelo que dan la

respuesta más cercana a la experimental. Los métodos más comunes de estimación de parámetros están basados en un enfoque de optimización, donde el mejor conjunto de parámetros es aquél que hace que la respuesta del modelo sea la más cercana a la real según un criterio o función de costo.

5. *Validación del modelo.* Tras completarse los pasos anteriores, es necesario asegurar que el modelo identificado es suficientemente representativo del proceso estudiado, teniendo en cuenta los objetivos deseados.

La validación consiste en chequear continuamente el comportamiento del modelo contra las características más importantes del sistema. Un modelo será “aceptado” si dichos chequeos son favorables. Si un modelo es “rechazado”, se procede a repetir las etapas de selección efectuando cambios tanto en la estructura del modelo como en las condiciones experimentales y la estimación de parámetros. Al desacoplar las etapas de modelación de las etapas de estimación y validación, es posible establecer un concepto de identificación independiente de la aplicación particular. Este paso convierte el proceso de identificación en un procedimiento de carácter iterativo.

Asimismo, las técnicas de identificación pueden dividirse en dos grandes clases:

*Fuera de línea.* En este caso los datos son recogidos tomando medidas durante la experimentación y, una vez terminada ésta, se procesan para producir el modelo. Esta técnica es la que se utiliza a lo largo de la tesis.

*En línea.* En este caso se emplea un algoritmo o método de actualización de parámetros de tipo recursivo que procesa los datos tal como son producidos por el sistema real. Esta técnica se emplea, principalmente, en control adaptativo y en aplicaciones de tiempo real cuando la dinámica del proceso debe ser monitoreada en forma continua.

Las siguientes secciones están orientadas a la descripción de los modelos más utilizados en identificación y estimación de sistemas dinámicos. Si bien esta tesis se centra sobre sistemas SISO, la mayoría de las técnicas expuestas son válidas para sistemas MIMO. Por otra parte, dado que se propone el empleo de modelos basados en datos muestreados, la exposición se centrará en modelos de tiempo discreto.

## 2.3. Modelos Caja Negra

El diseño de un modelo implica un compromiso entre la simplicidad y la necesidad de capturar los aspectos esenciales del sistema en estudio. Los modelos caja negra se diseñan enteramente a partir de los datos entrada-salida disponibles del sistema, sin tener en cuenta la interpretación de los parámetros que lo definen. La base del modelo es un modelo general cuya estructura sea muy flexible y que ya ha dado resultados en aplicaciones pasadas. Los parámetros de estos modelos por lo general no tienen significado físico y se ajustan para

describir el proceso generador de los datos observados. El origen físico e interpretación de las variables medidas no es importante, y de hecho tal información puede perderse, por ejemplo, a través de la manipulación de los datos observados.

En general, se pueden distinguir tres niveles del conocimiento previo que se tiene de un proceso y que han dado lugar a tres estructuras de modelos que se clasifican por colores.

*Modelos caja blanca.* Son modelos que se construyen con información *a priori* sobre el sistema y con los principios de la física y/o la mecánica que gobiernan el funcionamiento del mismo. Los parámetros tienen una directa relación con el sistema.

*Modelos caja gris.* En este caso se tiene un conocimiento físico del proceso, pero hay algunos parámetros del modelo que deben ser estimados a partir de los datos de entrada-salida del sistema. Se pueden considerar dos subclases diferentes:

- (i) *Modelado físico.* La estructura del modelo puede establecerse en base a consideraciones físicas y/o mecánicas, debiendo estimar los parámetros con las observaciones disponibles del sistema.
- (ii) *Modelado semi-físico.* Para construir este tipo de modelos se utiliza el conocimiento que se tiene sobre el comportamiento del sistema, para sugerir alguna combinación no lineal de los datos medidos. Estas nuevas señales estarán luego sujetas a una estructura del tipo caja negra.

*Modelos caja negra.* En este caso no hay disponibilidad o no se utiliza ningún conocimiento físico del sistema, pero la estructura del modelo seleccionado pertenece a la familia que previamente se identifica como apropiada pues tiene gran flexibilidad y se ha utilizado en casos similares. Los parámetros a ajustar no tienen relación alguna con el sistema.

Existen diferentes clases de modelos dentro de esta última categoría. Considerando su mayor simplicidad, los primeros en desarrollarse fueron los modelos lineales. Posteriormente, dada la necesidad de modelar con mayor precisión, surgieron los modelos no lineales. Las principales clases de modelos no lineales de caja negra son los modelos de *Hammerstein* y de *Wiener*, los modelos neuronales basados en redes neuronales recurrentes, los modelos difusos tipo Mamdani o relacionales, tipo Takagi-Sugeno y los modelos neuro-difusos.

Las estructuras que proponemos en esta tesis están dentro de esta familia de modelos ya que sólo tendremos disponibilidad de los datos de entrada-salida del sistema que deseamos modelar.

### 2.3.1. Estructura General de los Modelos Caja Negra

Podemos plantear matemáticamente el problema de identificación de un modelo caja negra de la siguiente manera.

Sea  $u(k)$  la entrada al sistema que deseamos modelar en el instante  $k$  y notemos

$$u^k = [u(1) \ u(2) \ \dots \ u(k)], \quad (2.3)$$

al vector que contiene el valor de las primeras  $k$  entradas exógenas al sistema.

Si  $y(k)$  es la salida respectiva observada e  $\tilde{y}(k)$  es la salida estimada para cada valor de  $k$ , escribimos

$$y^k = [y(1) \ y(2) \ \dots \ y(k)], \quad \tilde{y}^k = [\tilde{y}(1) \ \tilde{y}(2) \ \dots \ \tilde{y}(k)]. \quad (2.4)$$

Es decir,  $y^k$  es el vector que contiene el valor de las salidas del proceso en el instante  $k$  y anteriores e  $\tilde{y}^k$  es el vector que contiene los valores hasta el instante  $k$  de las salidas estimadas, que dependen de los parámetros del modelo.

Definimos el *error de estimación* del modelo en el instante  $k$  como  $e(k) = y(k) - \tilde{y}(k)$  y notamos

$$e^k = [e(1) \ e(2) \ \dots \ e(k)], \quad (2.5)$$

al vector de valores hasta el instante  $k$  de los errores de estimación, que dependen de los parámetros del modelo.

Entonces, en este tipo de modelos se desea hallar una relación entre observaciones pasadas  $[u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}]^2$  y observaciones futuras que estará dada por

$$y(k) = f(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}) + \nu(k), \quad (2.6)$$

donde el término  $\nu(k)$  es una perturbación aleatoria independiente que llamamos *error en las mediciones* y expresa que  $y(k)$  no es una función exacta de los datos pasados.

El objetivo es que  $\nu(k)$  sea lo más pequeño posible de tal manera que nuestro modelo  $f(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1})$  resulte una buena predicción de la verdadera salida  $y(k)$  a partir de los datos del pasado.

Podemos decir entonces que la ecuación (2.6) modela, en general, sistemas dinámicos discretizados. Luego, el problema es hallar la función  $f$ . En general, se busca una función  $f$  que sea parametrizable, es decir que tenga un número finito de parámetros. A estos parámetros se los representa con un vector  $\theta$  y a toda la familia de funciones candidatas se las denomina *estructura del modelo*. Obtenemos entonces una formulación en forma paramétrica dada por

$$f(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}, \theta). \quad (2.7)$$

Podemos medir la calidad de la parametrización  $\theta$  teniendo en cuenta  $N$  observaciones, por medio del error cuadrático entre el modelo y los datos verdaderos, *i.e.* minimizando

$$\sum_{k=1}^N \|y(k) - f(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}, \theta)\|^2, \quad (2.8)$$

---

<sup>2</sup>Notemos que estamos considerando  $u^k$  y no  $u^{k-1}$  pues, en general, debemos tener en cuenta la entrada actual.

respecto de  $\theta$ .

La norma utilizada y el método de minimización de este error pueden diferir de un esquema a otro, pero la mayoría de los esquemas de identificación siguen esta propuesta.

Como la familia de estructuras de modelo (2.7) es muy general, muchas veces es útil escribir a la función  $f$  como composición de dos transformaciones, una que tiene en cuenta el número creciente de observaciones pasadas  $[u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}]$  y la transforma en un vector de dimensión fija al que llamamos  $\varphi(k)$  y otra que toma este vector y lo lleva al espacio de salidas, es decir

$$f(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}, \theta) = f(\varphi(k), \theta), \quad (2.9)$$

donde  $\varphi(k) = \varphi(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1})$ .

Al vector  $\varphi(k)$  se lo denomina comúnmente *vector de regresión* y a sus componentes se las llama *regresores*.

Luego, la estructura general de un modelo caja negra puede escribirse como

$$\begin{aligned} y(k) &= f(\varphi(k), \theta) + \nu(k), \\ \varphi(k) &= \varphi(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}), \\ \nu(k) &\sim N(0, \sigma^2), \end{aligned} \quad (2.10)$$

donde  $f$  es una función lineal o no lineal,  $\varphi$  es una transformación de las entradas  $u$  en el instante  $k$  y anteriores, de las salidas reales y salidas estimadas pasadas  $y$  e  $\tilde{y}$ , respectivamente, y de los errores de estimación  $e$  en instantes anteriores. El valor  $\nu(k)$  es el error o perturbación adicionado a la salida en cada instante  $k$ , denominado *ruido blanco estacionario*. Generalmente se asume que los términos de error  $\nu(k)$  son variables aleatorias independientes idénticamente distribuidas, tomadas de una muestra con distribución normal de media nula y varianza constante  $\sigma^2$ :  $\nu(k) \sim N(0, \sigma^2)$ .

Llamamos *estructura de identificación* al modelo matemático definido por ecuaciones lineales o no lineales como en (2.9), que se utiliza para representar la conducta de entradas-salidas del proceso dado. Los parámetros del modelo se determinan fundamentalmente minimizando el error entre las salidas del proceso y las salidas de la estructura analítica que pretende representar al proceso, *i.e.* las salidas del modelo. A estos parámetros, que es posible obtener mediante un algoritmo de optimización, se los denomina *parámetros nominales* y al modelo obtenido utilizando este conjunto de parámetros se lo llama *modelo nominal*.

## 2.4. Identificación de Sistemas Lineales

La teoría de identificación para sistemas lineales es una teoría bien conocida, que suministra una serie de modelos y métodos de ajuste de parámetros para el modelado de procesos dinámicos lineales. Los modelos más frecuentes son los modelos lineales e invariantes en el tiempo, o modelos LTI (por sus siglas en inglés, Linear Time Invariant).

Consideremos un sistema que admite ser modelado como lineal, discreto e invariante en el tiempo, con una entrada  $u$  y una salida  $y$ , ambas escalares. Resultados análogos son válidos para sistemas MIMO con las debidas consideraciones. Una descripción en variables de estado para dicho sistema es

$$\begin{aligned}\mathbf{x}(k+1) &= A\mathbf{x}(k) + Bu(k) \\ y(k) &= C\mathbf{x}(k),\end{aligned}\tag{2.11}$$

donde  $k \in \mathbb{Z}$ ,  $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]$  representa a los  $n$  estados del sistema,  $\mathbf{x}_0 = \mathbf{x}(0)$  es el estado inicial y  $A$ ,  $B$  y  $C$  son matrices de dimensiones  $n \times n$ ,  $n \times 1$  y  $1 \times n$ , respectivamente.

Empleando el operador  $q^{-1}$  de retardo unitario en el tiempo definido como  $q^{-1}[u](k) = u(k-1)$ , un sistema lineal SISO, sujeto a perturbaciones aleatorias de tipo aditivo, queda representado por una ecuación en diferencias dada por

$$A(q)y(k) = \frac{B(q)}{F(q)}u(k) + \frac{C(q)}{D(q)}\nu(k),\tag{2.12}$$

siendo  $A(q)$ ,  $B(q)$ ,  $C(q)$ ,  $D(q)$  y  $F(q)$  polinomios en  $q^{-1}$  de  $n_a$ ,  $n_b$ ,  $n_c$ ,  $n_d$  y  $n_f$  coeficientes, respectivamente, de la forma

$$\begin{aligned}A(q) &= 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \\ B(q) &= b_0 + b_1q^{-1} + \dots + b_{n_b}q^{-n_b}, \\ C(q) &= 1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c}, \\ D(q) &= 1 + d_1q^{-1} + \dots + d_{n_d}q^{-n_d}, \\ F(q) &= 1 + f_1q^{-1} + \dots + f_{n_f}q^{-n_f}.\end{aligned}$$

La característica común de los estimadores lineales que tratan de modelar este tipo de procesos, consiste en que la salida estimada  $\tilde{y}(k)$  en el instante  $k$ , se obtiene multiplicando el vector de parámetros  $\theta$  del modelo, siendo  $\theta$  un vector con coeficientes constantes, por el vector de entradas disponibles en el instante  $k$  o vector de regresores  $\varphi(k)$

$$\tilde{y}(k) = \theta^T \varphi(k),\tag{2.13}$$

que son modelos lineales en los parámetros.

El vector de regresión  $\varphi$  puede constar de:

(RL1) Valores presentes y pasados de señales externas que influyen en la salida:  $u(k)$ ,  $u(k-1), \dots, u(k-n_b)$  (asociados con el polinomio  $B$ ).

(RL2) Valores pasados de las salidas reales del proceso:  $y(k-1), y(k-2), \dots, y(k-n_a)$  (asociados con el polinomio  $A$ ).

- (RL3) Valores pasados de las salidas estimadas sólo a partir de muestras pasadas de  $u$ , que dependen de los parámetros del modelo:  $\tilde{y}_u(k-1 | \theta), \tilde{y}_u(k-2 | \theta), \dots, \tilde{y}_u(k-n_f | \theta)$  (asociados con el polinomio  $F$ ).
- (RL4) Valores pasados de los *errores de estimación o predicción*  $e(k) = y(k) - \tilde{y}(k | \theta)$ , que dependen de los parámetros del modelo:  $e(k-1), e(k-2), \dots, e(k-n_c)$  (asociados con el polinomio  $C$ ).
- (RL5) Valores pasados de los *errores de simulación*  $e_u(k) = y(k) - \tilde{y}_u(k | \theta)$ :  $e(k-1), e(k-2), \dots, e(k-n_d)$  (asociados con el polinomio  $D$ ).

A continuación daremos una breve descripción de las estructuras de estos modelos y luego haremos su extensión al caso no lineal.

### 2.4.1. Modelo de Respuesta Finita al Impulso (FIR)

El sistema lineal más sencillo queda descrito por el modelo de respuesta finita al impulso o modelo FIR (por sus siglas en inglés, Finite Impulse Response)

$$y(k) = B(q)u(k) + \nu(k). \quad (2.14)$$

El estimador viene dado por

$$\tilde{y}(k) = b_0 u(k) + \dots + b_{n_b} u(k - n_b), \quad (2.15)$$

siendo el vector de parámetros

$$\theta = [b_0, \dots, b_{n_b}]^T$$

y el regresor o vector de entradas disponibles

$$\varphi(k) = [u(k), \dots, u(k - n_b)]^T.$$

Claramente, este es un modelo de memoria finita. Al aumentar el orden  $n_u$ , la estructura FIR

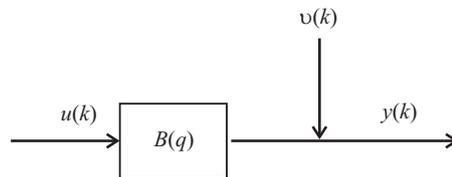


Figura 2.1: Modelo FIR

permite ajustar la mayoría de los procesos lineales habituales, lo cual implica incrementar

la cantidad de parámetros a identificar. Sin embargo, esta clase de modelos es de muy fácil identificación, se obtienen estimaciones que son mínimos globales, los algoritmos son estables y son fáciles de utilizar en simulación. Una desventaja es que no permite modelar las características del ruido.

### 2.4.2. Modelo de Error de Salida (OE)

El modelo de error de salida OE (por sus siglas en inglés, Output Error) es una variante del modelo FIR, en el que se incluyen como entradas al sistema valores pasados de la salida estimada

$$y(k) = \frac{B(q)}{F(q)}u(k) + \nu(k). \quad (2.16)$$

El estimador correspondiente es

$$\tilde{y}(k) = b_0u(k) + \dots + b_{n_b}u(k - n_b) - f_1\tilde{y}(k - 1) - \dots - f_{n_f}\tilde{y}(k - n_f), \quad (2.17)$$

siendo el vector de parámetros

$$\theta = [b_0, \dots, b_{n_b}, f_1, \dots, f_{n_f}]^T$$

y el regresor

$$\varphi(k) = [u(k), \dots, u(k - n_b), -\tilde{y}(k - 1), \dots, -\tilde{y}(k - n_f)]^T.$$

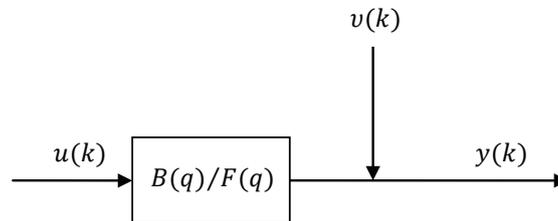


Figura 2.2: Modelo OE

Al igual que en el caso de los modelos FIR, esta estructura permite modelar la mayoría de los procesos lineales, pero no modela la característica del ruido. La ventaja con respecto al modelo FIR es que requiere menos parámetros, pero tiene la desventaja que es un modelo realimentado, por lo que el ajuste del vector de parámetros es más complejo.

### 2.4.3. Modelo Autorregresivo con Entradas Exógenas (ARX)

El modelo autorregresivo con entradas exógenas ARX (por sus siglas en inglés, Auto Regressive with eXogenous inputs) se diferencia del modelo OE en que se incluyen como

entradas valores pasados de las salidas reales del proceso, en lugar de las salidas estimadas. El modelo viene dado por

$$A(q)y(k) = B(q)u(k) + \nu(k). \quad (2.18)$$

El estimador es

$$\tilde{y}(k) = b_0u(k) + \dots + b_{n_b}u(k - n_b) - a_1y(k - 1) - \dots - a_{n_a}y(k - n_a), \quad (2.19)$$

el vector de parámetros es

$$\theta = [b_0, \dots, b_{n_b}, a_1, \dots, a_{n_a}]^T$$

y el vector de entradas está dado por

$$\varphi(k) = [u(k), \dots, u(k - n_b), -y(k - 1), \dots, -y(k - n_a)]^T.$$

Las componentes de este vector de regresión asociadas a valores pasados de las salidas se

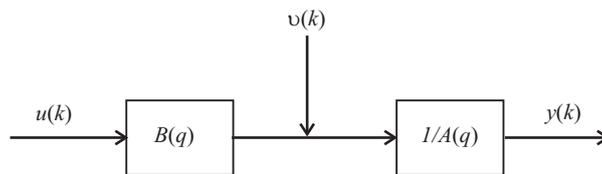


Figura 2.3: Modelo ARX

denominan *autorregresivas*. Este modelo permite ajustar todo proceso lineal corrompido con ruido aditivo, aumentando los órdenes  $n_y$  y  $n_u$ . Su ventaja frente al modelo FIR es que requiere menor cantidad de parámetros y, al ser un modelo no recurrente, la estimación de los mismos es más sencilla que en el caso de modelos OE.

#### 2.4.4. Modelo Autorregresivo de Media Móvil con Entradas Exógenas (ARMAX)

En predicción de series temporales es frecuente que el análisis de correlaciones residuales muestre un alto contenido de información en la serie del error de predicción. Esta información, enmascarada como ruido coloreado, puede ser realimentada al estimador de modo que su utilización mejore la calidad de la estimación, sin aumentar desproporcionadamente el número de parámetros a estimar.

La forma general de un proceso ARMAX (por sus siglas en inglés, Auto Regressive Moving Average with eXogenous inputs) viene dada por

$$A(q)y(k) = B(q)u(k) + C(q)\nu(k). \quad (2.20)$$

El estimador utilizado en este caso es

$$\begin{aligned} \tilde{y}(k) = & b_0 u(k) + \dots + b_{n_b} u(k - n_b) - a_1 y(k - 1) - \dots - a_{n_a} y(k - n_a) \\ & - c_1 e(k - 1) - \dots - c_{n_c} e(k - n_c), \end{aligned} \quad (2.21)$$

donde  $e(k)$  es el error de estimación en el instante  $k$ . Como en el caso del modelo ARX, las

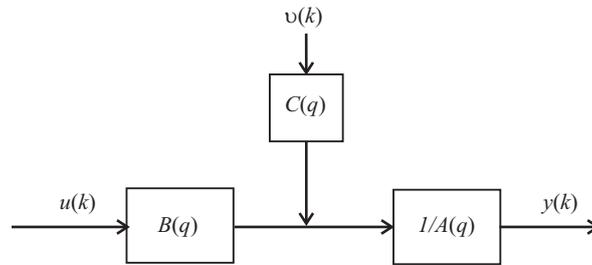


Figura 2.4: Modelo ARMAX

componentes del vector de regresión asociadas a valores pasados de las salidas se denominan autorregresivas, mientras que las asociadas a errores de estimación son llamadas de *media móvil*. El vector de parámetros es

$$\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}, c_1, \dots, c_{n_c}]^T$$

y el vector de regresión está dado por

$$\varphi(k) = [y(k - 1), \dots, y(k - n_a), u(k), \dots, u(k - n_b), e(k - 1), \dots, e(k - n_c)]^T.$$

Esta estructura permite aproximar procesos lineales corrompidos con ruido aditivo, con un modelo más compacto que el modelo ARX para el caso de ruido coloreado. Como en el caso de modelos OE, el problema que presentan es que, al ser modelos recursivos, la estimación de los parámetros es mucho más compleja.

### 2.4.5. Representación con Bases de Funciones Ortonormales

Existen otras clases de regresores que no son sólo funciones lineales de las salidas del modelo sino que también están basados en el conocimiento físico del sistema. Desde un punto de vista práctico, es suficiente considerar a la entrada  $u$  y a la salida  $y$  como transformadas de los datos medidos, de tal manera que se utilice lo que se conoce acerca del sistema. A este tipo de regresores se los denomina *regresores semi-físicos*. Un ejemplo clásico es la utilización de regresores que son versiones filtradas de la señal de entrada  $u$ , del tipo

$$L_j(q) u(k), \quad j = 1, \dots, d, \quad (2.22)$$

donde los filtros  $L_j$  son diferentes según la aplicación. Los filtros de *Laguerre* y de *Kautz* han sido utilizados con este fin como puede verse en [96, 97, 84]. En Van den Hof *et al.* [91], se pueden encontrar diferentes generalizaciones para la elección de este tipo de regresores.

En esta subsección analizamos en detalle las bases de funciones ortonormales, poniendo especial énfasis en las de *Laguerre* y de *Kautz*, debido a que las utilizaremos en esta tesis en la definición de una estructura de modelado de sistemas dinámicos no lineales.

El uso de bases de funciones ortonormales en la descomposición de un sistema dinámico permite disminuir el orden del modelo, manteniendo el grado de aproximación del mismo, y cuantificar los errores causados por dinámica no modelada o debido a perturbaciones. Los modelos con bases ortonormales son aptos para estimaciones utilizando cuadrados mínimos ya que los algoritmos son robustos y se tiene una reducida carga computacional. Además, es posible extender los resultados a sistemas multivariables.

Consideremos espacios funcionales con estructura de espacio de Hilbert, *i.e.* espacios con producto interno que son completos con respecto a la métrica inducida por el producto interno, asociados a las funciones de transferencia racionales de sistemas discretos, causales y estables.

Sea  $L^2(T)$  el espacio de funciones de cuadrado integrable según Lebesgue en la circunferencia unitaria en el plano complejo  $T = \{z \in \mathbb{C} : \|z\| = 1\}$ . Definimos el siguiente producto interno en  $L^2(T)$

$$\langle F, G \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(e^{j\omega}) \overline{G(e^{j\omega})} d\omega \quad (2.23)$$

o, equivalentemente,

$$\langle F, G \rangle = \frac{1}{2\pi j} \oint_T F(z) \overline{G(1/z)} \frac{dz}{z}. \quad (2.24)$$

Entonces el espacio  $L^2(T)$  con este producto interno es un espacio de Hilbert.

Sea  $H^2(T)$  el espacio de Hardy de las funciones de cuadrado integrable según Lebesgue en la circunferencia unitaria, que son analíticas en el exterior del disco unitario  $D = \{z \in \mathbb{C} : \|z\| < 1\}$ .

$H^2(T)$  es un subespacio propio de  $L^2(T)$  completo; por lo tanto es un espacio de Hilbert con el producto interno heredado de  $L^2(T)$ . Este espacio puede pensarse como el espacio de las funciones de transferencia racionales discretas, causales y estables.

**Definición 2.4.1**  $\{B_k(q)\}_{k=0}^{\infty}$  es una base ortonormal del espacio de Hilbert  $H^2(T)$  si es un conjunto ortonormal y completo, *i.e.* si verifica

$$\langle B_k, B_l \rangle = \delta_{kl},$$

donde  $\delta_{kl}$  es la delta de Kronecker.

Esto implica que cualquier función  $G(q) \in \mathbb{H}^2(T)$  tiene una representación única de la forma

$$G(q) = \sum_{k=0}^{\infty} \theta_k B_k(q), \quad (2.25)$$

donde  $\theta_k = \langle G, B_k \rangle$ .

La estructura de modelo es  $y(k) = G(q)u(k) + \nu(k)$ , donde  $\nu(k)$  es una perturbación que se considerará un proceso aleatorio. La función  $G(q)$  es representada como una expansión en bases ortonormales de la forma (2.25). Obviamente, como  $G(q)$  es desconocida, los coeficientes de la expansión no pueden calcularse mediante la igualdad  $\theta_k = \langle G, B_k \rangle$ . El objetivo es entonces estimar los parámetros de un modelo de dimensión finita

$$G(q) = \sum_{k=0}^{n-1} \theta_k B_k(q), \quad (2.26)$$

de modo que una estimación de  $G(q)$  es

$$G(q, \hat{\theta}) = \sum_{k=0}^{n-1} \hat{\theta}_k B_k(q), \quad (2.27)$$

con  $\hat{\theta}_k$  una estimación de  $\theta_k$ .

La estructura de modelo (2.26),(2.27) puede escribirse en la forma de regresor lineal

$$y(k) = \varphi^T(k)\theta + \nu(k), \quad (2.28)$$

definiendo

$$\begin{aligned} \varphi^T(k) &= [B_0(q)u(k), B_1(q)u(k), \dots, B_{n-1}(q)u(k)] \\ &= [z_1(k), \dots, z_m(k)]^T, \end{aligned} \quad (2.29)$$

$$\theta = [\theta_0 \ \theta_1 \ \dots \ \theta_{n-1}]^T. \quad (2.30)$$

La estimación del vector de parámetros  $\theta$  puede hacerse usando el método de cuadrados mínimos. Debemos minimizar la función de costo

$$J(\theta) = \frac{1}{N} \sum_{k=1}^N (y(k) - \varphi^T(k)\theta)^2$$

que en este caso tiene una solución analítica y es

$$\hat{\theta}_N \in \arg \left( \min_{\theta} \{V_N(\theta)\} \right) = \left[ \sum_{k=1}^N \varphi(k)\varphi^T(k) \right]^{-1} \left[ \sum_{k=1}^N \varphi(k)y(k) \right], \quad (2.31)$$

donde la función objetivo  $V_N(\theta)$  es un criterio cuadrático en los errores de predicción  $y(k) - \varphi^T(k)\theta$

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^n \text{Tr}\{[y(k) - \varphi(k)^T\theta][y(k) - \varphi^T(k)\theta]^T\}. \quad (2.32)$$

Esta metodología para estimar  $\theta$  se denomina *método de identificación del error de predicción* o PEM (por sus siglas en inglés, Prediction Error Method).

Es claro que la precisión del modelo, en términos de la mínima desviación posible entre sistema y modelo en alguna norma, dependerá de la elección de las funciones de la base  $B_k$ .

Por esta razón, el desarrollo de una base de funciones apropiada es un tópico que tiene considerable interés. El problema aquí es elegir una base de funciones que refleje la dinámica dominante del sistema a ser modelado.

### 2.4.6. Algunas Bases Ortonormales en $H^2(T)$

#### Modelo FIR

Se puede considerar a los modelos FIR desde el punto de vista de aproximaciones funcionales como

$$y(t) = \sum_{k=0}^{n_u} b_k q^{-k} u(t), \quad (2.33)$$

donde  $q^{-k}u(t) = u(t - k)$  es el operador retardo de  $k$  muestras en el tiempo. En este caso es

$$B_k(q) = q^{-k},$$

es decir, estas funciones tienen polos en el origen del plano complejo. Si el sistema tiene una respuesta al impulso de larga duración, se necesita un gran número de términos de la expansión en serie para representarlo adecuadamente.

Podemos ver entonces que el sistema es expandido en una base ortonormal formada por retardos unitarios. El operador de retardo unitario  $q^{-1}$  tiene una memoria de sólo un valor, lo que explica la gran cantidad de términos que hay que considerar para aproximar algunos sistemas.

Vale el siguiente teorema

**Teorema 2.4.2** *Sea un sistema estable  $G(z) \in H^2(T)$  descrito por*

$$G(z) = \sum_{k=0}^{\infty} g_k z^{-k},$$

donde  $\{g_k\}$  es la sucesión de respuestas al impulso. Entonces existe una única expansión

$$G(z) = \sum_{k=0}^{\infty} \theta_k B_k(z) = \sum_{k=0}^{\infty} \theta_k z^{-k}, \quad (2.34)$$

donde  $\theta_k \in \mathbb{R}, \forall k$ , son los coeficientes de la expansión.

### Sistemas de Laguerre

El conjunto de sistemas de Laguerre puede ser expresado en el campo complejo en términos de su transformada  $\mathcal{Z}$  por

$$L_j(z, a) = z^{-nd} \frac{\sqrt{(1-a^2)T}}{z-a} \left( \frac{1-az}{z-a} \right)^{j-1}, \quad j = 1, 2, \dots \quad (2.35)$$

donde el parámetro real  $a$  es el polo generador denominado coeficiente de Laguerre y  $|a| < 1$ , lo que asegura la estabilidad del filtro,  $nd$  es el retardo puro del sistema de  $nd$  muestras y  $T$  es el tiempo de muestreo. Cuando el sistema es discreto,  $T = 1$ .

Este conjunto de sistemas es una generalización de los presentados en [96] dado que incluyen un retardo de  $nd$  muestras. En [96] los sistemas son estrictamente propios, luego los sistemas definidos en (2.35) coinciden con los allí presentados si  $nd = 1$ .

La ecuación (2.35) puede expresarse en base al operador de retardo unitario  $q^{-1}$  en la forma

$$L_j(q^{-1}, a) = q^{-nd} \frac{\sqrt{(1-a^2)T}}{1-aq^{-1}} \left( \frac{q^{-1}-a}{1-aq^{-1}} \right)^{j-1}, \quad j = 1, 2, \dots \quad (2.36)$$

Se puede ver que estos sistemas se generan a partir de

$$L_0(q^{-1}, a) = \frac{\sqrt{(1-a^2)T}}{1-aq^{-1}}, \quad (2.37)$$

que representa un filtro pasabajos básico, al que se le van agregando secciones pasabajo en cascada de la forma  $\left( \frac{q^{-1}-a}{1-aq^{-1}} \right)$ . Entonces, podemos expresar a cada nuevo sistema en base al anterior de una manera recursiva simple

$$L_j(q^{-1}, a) = L_{j-1}(q^{-1}, a) \frac{\sqrt{(1-a^2)T}}{1-aq^{-1}}. \quad (2.38)$$

Utilizando estos sistemas como una base, un sistema lineal estrictamente propio puede representarse por medio de la ecuación (2.36) con  $nd = 1$  como:

$$y(k) = \sum_{j=0}^{n_L} b_j L_j[u(k)]. \quad (2.39)$$

Vale el siguiente lema cuya demostración se puede consultar en [96].

**Lema 2.4.3** *Sea  $G(z)$  una función estrictamente propia ( $G(\infty) = 0$ ), analítica en el conjunto  $\{z \in \mathbb{C} : \|z\| > 1\}$  y continua en  $\{z \in \mathbb{C} : \|z\| \geq 1\}$  y sea  $|a| < 1$ . Entonces existe una sucesión  $\{\theta_k\}$  tal que*

$$G(z) = \sum_{k=0}^{\infty} \theta_k L_k(z, a), \quad \|z\| \geq 1, \quad (2.40)$$

donde las funciones  $L_k(z, a)$  son las transformadas  $\mathcal{Z}$  de los sistemas de Laguerre.

Es conocido que la velocidad de aproximación de una función por filtros de Laguerre depende en gran medida de la elección del polo. Un ejemplo típico consiste en aproximar una función  $h(t)$  por una serie de Laguerre truncada con un número fijo de términos. Si

$$h_n(t) = \sum_{i=0}^n c_i(a) l_i(t, a), \quad (2.41)$$

donde  $l_i(t, a)$  es la respuesta en tiempo de  $L_i(z, a)$  y  $c_i(a)$  son los coeficientes óptimos de la combinación lineal. Es decir,

$$c_i(a) \in \arg \left( \begin{array}{c} \min \\ c_i \in \mathbb{R} \\ 0 \leq i \leq n \end{array} \|h(t) - h_n(t)\|_2 \right), \quad (2.42)$$

con  $h_n(t)$  dado por (2.41).

Supongamos que queremos evaluar el mínimo error en la estimación en norma  $\ell^2$ , esto es

$$E_n(a) = \|h(t) - h_n(t)\|_2, \quad (2.43)$$

donde el parámetro  $a$  está indicado para enfatizar la dependencia del error en la aproximación del polo.

Podemos reescribir el error como

$$E_n(a) \leq h(t), \quad h(t) > - \sum_{i=0}^n c_i^2(a). \quad (2.44)$$

El objetivo es hallar la posición óptima del polo  $a$ , que es el valor de  $a$  que minimiza  $E_n(a)$ . Está probado que en cada punto estacionario de  $E_n(a)$  se verifica la siguiente condición [39, 53, 99]

$$c_n(a)c_{n+1}(a) = 0.$$

Claramente, la determinación de los puntos estacionarios de  $E_n^2(a)$  requiere la solución de las ecuaciones  $c_i(\cdot) = 0$ , para  $i = n$  y para  $i = n+1$ . Dado que, en general, este procedimiento es muy dificultoso, en los últimos años han sido propuestos algunos procedimientos para resolver este problema; ver, por ejemplo, [11, 29, 39, 53, 59, 79, 99]. Otra propuesta que conduce a la elección óptima del polo para una clase de sistemas que satisfacen ciertas medidas ha sido dada en [37, 72].

Con una correcta selección del polo  $a$ , se pueden lograr aproximaciones de orden menor que las que se obtienen con una aproximación FIR. En [96], Wahlberg sugiere que esto se debe a que el operador de los sistemas Laguerre tiene más memoria que el operador retardo unitario utilizado en un modelo FIR. Las ventajas a tener en cuenta son apreciables desde el punto de vista de la aproximación y de la estabilidad. En efecto, si los utilizamos en un esquema de identificación adaptativo, podemos asegurar la estabilidad de dicho esquema controlando que  $|a| < 1$ . Por otro lado, es posible utilizar estos modelos en simulación para horizontes de tiempo arbitrariamente largos.

Al aproximar la representación externa de un sistema lineal, discreto, utilizando sistemas de Laguerre es posible mejorar la aproximación obtenida simplemente aumentando el orden del modelo.

### Sistemas de Kautz

Aunque el uso de filtros de Laguerre en identificación de sistemas es muy común, una desventaja de esta clase de funciones es que, en general, los sistemas subamortiguados son dificultosos de aproximar con un número razonable de términos. Por esta razón, otros sistemas de funciones ortonormales propuestos son los filtros de Kautz con dos parámetros. Estos filtros permiten aproximar más eficientemente señales con fuerte comportamiento oscilatorio [97, 98].

Las funciones de Kautz discretas pueden escribirse de varias formas. Una forma posible es definir las a través de su transformada  $\mathcal{Z}$  como

$$K_{2j}(z, b, c) = \frac{\sqrt{(1-c^2)(1-b^2)}}{z^2 + b(c-1)z - c} \left( \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{j-1}$$

$$K_{2j-1}(z, b, c) = \frac{(z-b)\sqrt{(1-c^2)}}{z^2 + b(c-1)z - c} \left( \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{j-1}, \quad (2.45)$$

donde  $b = (\beta + \bar{\beta})/(1 + \beta\bar{\beta})$ ,  $c = -\beta\bar{\beta}$  son constantes reales tales que

$$|b| < 1, \quad |c| < 1, \quad j = 1, 2, \dots$$

y  $\beta, \bar{\beta} \in \mathbb{C}$  son los polos de Kautz en la región  $\|z\| < 1$ , de modo que se puede asegurar estabilidad, ver [97].

Luego, la función  $G(z)$  tiene una expansión en serie de Kautz dada por

$$G(z) = \sum_{j=0}^{\infty} \theta_j K_j(z, a), \quad \|z\| \geq 1. \quad (2.46)$$

Comparado con los filtros de Laguerre, ahora hay dos parámetros,  $b$  y  $c$ , que permiten asignar un par de polos complejos conjugados a las funciones de transferencia  $K_j(z)$ .

Notemos que las funciones de transferencia de Laguerre discretizadas son un caso particular de las funciones de transferencia de Kautz [97].

Tal como sucede con los filtros de Laguerre, para un número grande de términos en la expansión, la elección de los parámetros  $b$  y  $c$  no es crucial, pero para una cantidad finita de términos la elección de los parámetros es de gran importancia para obtener una buena aproximación de una función dada.

Es importante hacer una comparación entre el modelo FIR y el modelo obtenido utilizando bases de Kautz. La diferencia esencial es que la memoria del elemento de la base de Kautz es infinita y la del retardo unitario es uno. De esta manera, la representación por medio de filtros de Kautz soluciona el problema de la falta de memoria a través de los elementos de la base reduciendo la cantidad de términos involucrados en la aproximación.

La estructura de modelo lineal para identificar un sistema utilizando bases de funciones ortonormales, genera un regresor lineal que sólo depende de las entradas al sistema. Entonces, la estimación cuadrática tiene una solución cerrada correspondiente a un mínimo global. Con esto se evitan costosos métodos de iteración en la estimación de parámetros y la aparición de mínimos locales. El uso de bases ortonormales también permite incorporar al proceso de identificación toda la información *a priori* disponible de la dinámica del sistema.

## 2.5. Identificación de Sistemas no Lineales

Debido a la complejidad de los sistemas dinámicos no lineales, debemos restringir el campo de estudio a una subclase de ellos. En esta tesis nos interesan los sistemas no lineales, discretos que poseen *memoria evanescente*.

La representación interna o representación entradas-estados-salidas de un sistema SISO no lineal discreto, invariante en el tiempo es

$$\begin{aligned} \mathbf{x}(k+1) &= f(\mathbf{x}(k), u(k)) \\ y(k) &= g(\mathbf{x}(k)), \end{aligned} \quad (2.47)$$

con  $k \in \mathbb{Z}$ , donde  $u(k)$ ,  $\mathbf{x}(k)$  e  $y(k)$  son los elementos de las sucesiones de entrada, estados y salidas, respectivamente, en el instante de tiempo  $k$ . Sea  $\mathbf{x}(0) = \mathbf{x}_0$  el estado inicial.

De acuerdo con (2.10), el modelo E/S es

$$\begin{aligned} y(k) &= f(\varphi(k), \theta) + \nu(k), \\ \varphi(k) &= \varphi(u^k, y^{k-1}, \tilde{y}^{k-1}, e^{k-1}), \\ \nu(k) &\sim N(0, \sigma^2), \end{aligned} \quad (2.48)$$

donde ahora  $f$  es una función no lineal,  $\varphi$  es una transformación de las entradas en el instante  $k$  y anteriores, y de las salidas observadas, salidas estimadas y errores de estimación pasados;  $\nu(k)$  es el error o ruido blanco estacionario adicionado a la salida en cada instante  $k$ , de media nula y varianza constante  $\sigma^2$ .

Para modelar estos sistemas, ampliamos los modelos lineales vistos en la sección anterior, extendiéndolos al caso no lineal mediante la sustitución de la transformación lineal por un aproximador no lineal representado por una función  $f$  de modo que obtendremos una estimación de la salida del sistema de la forma

$$\tilde{y}(k) = f(\varphi(k), \theta).$$

Entonces, la salida estimada  $\tilde{y}(k)$  en el instante  $k$ , se obtiene a partir de toda la información disponible en dicho instante “guardada” en el vector de regresión o vector de entradas  $\varphi(k)$ . Este vector está formado por valores de las señales de entrada  $u(k)$  en los instantes  $k$  y anteriores, por los valores de las salidas reales  $y(k)$  del sistema, de las salidas estimadas  $\tilde{y}(k)$  y de los errores de estimación  $e$  en instantes anteriores.

Para ajustar el vector de parámetros mediante la minimización de una función de costo basada en el error de estimación, como por ejemplo el error cuadrático medio, debemos calcular las derivadas de las salidas del modelo respecto a cada uno de sus parámetros  $dy/d\theta$ .

Luego, el proceso de elección de  $f$  consta de dos pasos:

Paso 1. Cómo elegir el vector de regresión  $\varphi(k)$ , *i.e.* cómo elegimos los regresores y la forma de combinarlos.

Paso 2. Cómo elegir la transformación no lineal  $f$ .

### 2.5.1. Elección de los Regresores

Llamemos, por comodidad,

$$\tilde{y}(k | \theta) = f(\varphi(k), \theta). \quad (2.49)$$

Los regresores están, en general, dados por

(RN1) Valores presentes y pasados de las entradas:  $u(k), u(k-1), \dots, u(k-n_u)$ .

(RN2) Valores pasados de las salidas reales del proceso:  $y(k-1), y(k-2), \dots, y(k-n_y)$ .

(RN3)  $\tilde{y}_u(k-1 | \theta), \tilde{y}_u(k-2), \dots, \tilde{y}_u(k-n_{\tilde{y}})$ , que representan las salidas estimadas sólo a partir de muestras pasadas de  $u$ .

(RN4)  $e(k-t | \theta) = y(k-t) - \tilde{y}(k-t | \theta)$ , *i.e.* el error de estimación o de predicción.

(RN5)  $e_u(k-t | \theta) = y(k-t) - \tilde{y}_u(k-t | \theta)$ , *i.e.* el error de simulación.

Observemos que, en los dos primeros casos, las componentes de los regresores son simplemente las medidas presentes o pasadas que se tienen. En cambio, los otros tipos de regresores están basados en salidas previas del modelo caja negra y, por lo tanto, debemos notar el vector de regresión como  $\varphi(k, \theta)$  en lugar de hacerlo simplemente con  $\varphi(k)$ . Por otra parte, en estos casos, las salidas simuladas  $\tilde{y}_u(k-t | \theta)$  son iguales a las salidas del modelo (2.49) si todas las salidas medidas  $y(k-t)$  son reemplazadas, en los regresores, por las últimas salidas calculadas  $\tilde{y}_u(k-t | \theta)$  y, por lo tanto, los errores de predicción y de simulación coinciden.

### 2.5.2. Elección de la Transformación no Lineal

Una vez elegidos los regresores, debemos determinar la función no lineal  $f(\varphi, \theta)$  o, mejor dicho, una estimación o parametrización de la misma. En este punto no interesa cómo ha sido elegido el vector de regresión  $\varphi$  sino que simplemente lo tomamos como un vector dado. Es natural pensar a la familia de funciones parametrizadas como superposición de funciones, de la siguiente manera

$$f(\varphi, \theta) = \sum \alpha_i f_i(\varphi). \quad (2.50)$$

A las funciones  $f_i$  se las llama *funciones base* por el papel que juegan, similar al de las funciones de una base en un espacio funcional. Más aún, a veces ellas realmente forman una base del espacio que consideramos. Luego, la pregunta natural que surge en este momento es cómo elegir las funciones base.

Una de las funciones base más conocidas es  $f_i(x) = \cos(ix)$ . Entonces (2.50) expresa la expansión en serie de Fourier de la función  $f$ .

Los llamados *modelos difusos* pertenecen a la clase de modelos con la estructura dada por (2.50). Las funciones base  $f_i$  se construyen utilizando miembros de un conjunto de funciones difusas y reglas de inferencia [86].

En nuestro caso, la función no lineal  $f$  se obtendrá como aproximación de la salida verdadera  $y$ , por medio de la expansión en una base de funciones lineales a tramos canónicas de alto orden, que definimos en la Sección 2.7.

Siguiendo los trabajos de Chen *et al.* [23] y de Chen y Billings [22], de acuerdo a qué regresores se eligen, se tienen los siguientes modelos.

### Modelo NFIR

El sistema no lineal más sencillo es la versión no lineal del sistema de respuesta finita al impulso NFIR (por sus siglas del inglés, Nonlinear Finite Impulse Response). Estos modelos utilizan sólo la información  $u(k-t)$ .

El estimador viene dado por

$$\tilde{y}(k) = f(u(k), \dots, u(k - n_u), \theta).$$

Al no existir conexiones recurrentes, las derivadas de las salidas estimadas respecto de los parámetros del modelo estarán dadas por las salidas del aproximador funcional respecto de sus parámetros

$$\frac{\partial \tilde{y}_{i(k)}}{\partial \theta} = \frac{\partial f_i}{\partial \theta}.$$

### Modelo NOE

El modelo no lineal de error de salida NOE (por sus siglas del inglés, Nonlinear Output Error) utiliza  $u(k-t)$  e  $\tilde{y}_u(k-t | \theta)$  como entradas al sistema.

La salida del modelo es

$$\tilde{y}(k | \theta) = f(\tilde{y}_u(k-1), \dots, \tilde{y}_u(k - n_{\tilde{y}}), u(k), \dots, u(k - n_u), \theta). \quad (2.51)$$

Al tratarse de un modelo realimentado la expresión analítica de las derivadas  $dy/d\theta$  ya no depende sólo de las derivadas del aproximador funcional  $f$  respecto de sus parámetros.

Se calculan las derivadas suponiendo constantes las entradas del aproximador (lazo abierto), y posteriormente se incluye el efecto de la realimentación calculando la derivada de cada salida respecto de cada entrada recurrente y multiplicando estos valores por las derivadas de las entradas recurrentes respecto de los pesos.

Formalmente,

$$\frac{\partial \tilde{y}_{i(k)}}{\partial \theta} = \frac{\partial \tilde{y}_{i(k)}}{\partial \theta} \Big|_{u(k), \tilde{y}(k-1)} + \sum_{j=1}^l \sum_{r=1}^{n_{\tilde{y}_j}} \frac{\partial \tilde{y}_{i(k)}}{\partial \tilde{y}_{i(k-\tau)}} \Big|_{u(k), \theta} \frac{\partial \tilde{y}_{i(k-\tau)}}{\partial \theta}. \quad (2.52)$$

Los términos  $\frac{\partial \tilde{y}_{i(k)}}{\partial \theta} \Big|_{u(k), \tilde{y}(k-1)} = \frac{\partial f_i}{\partial \theta}$  son las derivadas de las salidas del aproximador funcional respecto de sus parámetros, manteniendo constantes las entradas exógenas y los parámetros.

La ecuación (2.52) permite diseñar un algoritmo recursivo para la evaluación de las derivadas del modelo NOE. Los primeros términos de las derivadas  $\frac{\partial \tilde{y}_{i(k)}}{\partial \theta}$ , se inicializan en cero para  $k = -1, \dots, -n_{\tilde{y}_j}$  para que el cálculo del gradiente parta siempre del mismo punto inicial.

Las estructuras de identificación NOE tienen una reducida sensibilidad al ruido.

### Modelo NARX

Los modelos no lineales autorregresivos con entrada exógena NARX (Nonlinear Auto Regressive with eXogenous inputs) utilizan  $u(k-t)$  e  $y(k-t)$  como variables de entrada al sistema. El estimador en este caso es

$$\tilde{y}(k) = f(y(k-1), \dots, y(k-n_y), u(k), \dots, u(k-n_u), \theta),$$

donde, a diferencia del modelo NOE, las entradas son ahora los valores medidos de las salidas. De esta forma se evitan las recursividades, simplificando el cálculo de las derivadas de las salidas estimadas respecto de los parámetros del aproximador

$$\frac{\partial \tilde{y}_{i(k)}}{\partial \theta} = \frac{\partial f_i}{\partial \theta}.$$

### Modelo NARMAX

Un sistema no lineal autorregresivo de media móvil con entrada exógena NARMAX (por sus siglas en inglés, Nonlinear Auto Regressive Moving Average with eXogenous inputs) admite el estimador

$$\tilde{y}(k) = f(y(k-1), \dots, y(k-n_y), u(k), \dots, u(k-n_u), e(k-1), \dots, e(k-n_e), \theta).$$

Como es un modelo recursivo, la expresión analítica de las derivadas  $d\tilde{y}/d\theta$  de las salidas estimadas con respecto a los parámetros, ya no depende sólo de las derivadas del aproximador funcional  $f$ . Podemos obtener la expresión analítica de estas derivadas de manera similar a como se hizo para el modelo NOE.

### Modelo NBJ

Los modelos no lineales Box Jenkins o NBJ (por sus siglas en inglés Nonlinear Box Jenkins) utilizan  $u(k-t)$ ,  $\tilde{y}(k-t)$ ,  $e(k-t | \theta)$  y  $e_u(k-t | \theta)$  como datos de entrada al sistema dinámico. En este caso, la salida simulada  $\tilde{y}_u$  se obtiene como la salida de (2.49) utilizando la misma estructura, pero reemplazando  $e$  y  $e_u$  por ceros en el vector de regresión  $\varphi(k, \theta)$ .

Narendra y Parthasarathy [68] utilizan una notación diferente para los mismos modelos cuando se utilizan conjuntamente con redes neuronales. El modelo NARX se denomina *serie-paralelo* y al NOE se lo llama *modelo paralelo*. Las estructuras NOE, NBJ y NARMAX corresponden a estructuras recursivas dado que parte del vector de regresión consiste de salidas pasadas del modelo.

## 2.6. Modelos no Lineales Orientados a Bloques

La mayoría de los sistemas dinámicos tienen un comportamiento no lineal, excepto en un determinado rango de operación donde pueden ser considerados lineales. Por lo tanto, si utilizamos una estructura lineal para modelar un sistema no lineal, al variar el punto de operación del sistema real las características predictivas del modelo se ven deterioradas. Esto hace que para la identificación de sistemas con estas características se recurra a técnicas y modelos no lineales.

Muchos de los sistemas dinámicos no lineales pueden ser representados por la conexión en cascada o en paralelo de sistemas o *bloques* lineales dinámicos y *bloques* no lineales estáticos. Estos modelos se denominan *orientados a bloques*.

Las no linealidades estáticas aparecen, por ejemplo, debido a saturación de actuadores, sensores con características no lineales, etc. Entre los modelos orientados a bloques, los que han sido más estudiados son los modelos de Hammerstein y de Wiener.

Un modelo de Hammerstein consiste básicamente en un modelo compuesto de un bloque no lineal estático, seguido por otro bloque lineal dinámico. Similarmente, un modelo tipo

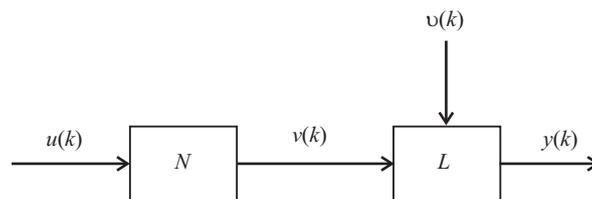


Figura 2.5: Modelo tipo Hammerstein

Wiener consiste en un bloque lineal dinámico seguido por un bloque no lineal estático. Podemos incluir el modelo tradicional de Wiener dentro de los modelos tipo caja negra con

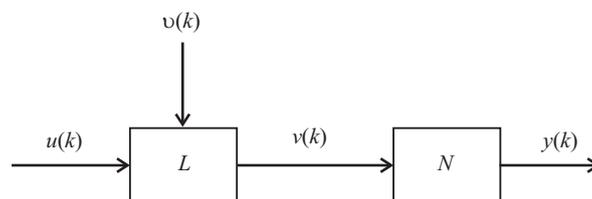


Figura 2.6: Modelo tipo Wiener

regresores semi-físicos ya que se describe matemáticamente mediante la ecuación

$$\tilde{y}(k+1) = h(v_0[k], v_1[k], \dots), \quad (2.53)$$

donde  $h: \ell^\infty \rightarrow \mathbb{R}$  es la parametrización de la transformación no lineal utilizando como base los polinomios de Hermite y  $v_j[k] = L_j[u(k)]$  es la salida del  $j$ -ésimo filtro de Laguerre. De la ecuación (2.53) se ve inmediatamente la relación con los tipos de modelos antes descritos. En el Capítulo 4 nos ocuparemos nuevamente de este tipo de modelos.

Notaremos a los bloques lineal y no lineal de un modelo orientado a bloques por las letras  $L$  y  $N$ , respectivamente. De modo similar, denominaremos a los bloques en paralelo mediante la letra  $P$ , seguido de un índice de multiplicidad que indica el número de modelos en bloque que utiliza el esquema de identificación. Así, los modelos Hammerstein son llamados “modelos  $NL$ ”, y los modelos Wiener “modelos  $LN$ ”. Otros modelos en bloques son los modelos  $LNL$ , conocidos también como modelos “tipo sándwich” y los modelos de *Uryson* que consisten de  $m$  modelos de Hammerstein en paralelo y se designan como  $PNLm$ .

En el Capítulo 4 definiremos los esquemas de identificación *tipo Wiener* que consideramos dentro de la clase de modelos orientados a bloques. Seguidamente introduciremos los modelos de identificación tipo Wiener propuestos en esta tesis.

## 2.7. Representaciones Canónicas Lineales a Tramos de Alto Nivel

En esta sección describimos las bases de representación de funciones lineales a tramos (LAT) continuas, definidas sobre una partición simplicial de un dominio  $\mathbf{S} \subset \mathbb{R}^n$ . La base está compuesta por funciones con varios niveles de anidamiento de valores absolutos, que se obtienen a partir de una única función generadora  $\gamma$ . Comenzamos con una descripción de estas funciones y sus principales características. Esta estructura de aproximación permite representar sistemas no lineales estáticos, y la utilizaremos en el Capítulo 4 como parte de la formulación de los modelos de aproximación de sistemas dinámicos no lineales. Presentamos luego los teoremas que aseguran su capacidad de aproximación de funciones no lineales estáticas. Estos resultados son fundamentales para establecer luego el teorema de aproximación propuesto en esta tesis.

### 2.7.1. Introducción

La primera representación canónica de funciones lineales a tramos (CLAT) fue introducida por Chua y Kang [27, 52] en el contexto de teoría de circuitos no lineales. La expresión canónica propuesta en los artículos antes citados da una formulación eficiente para representar cualquier función arbitraria LAT con dominio en  $\mathbb{R}$ , y fue extendida a dominios en

$\mathbb{R}^2$  (ver [26]). Más tarde, se demostró en [57] la existencia de una representación canónica para funciones LAT con dominio en  $\mathbb{R}^n$ . Sin embargo, este resultado es meramente teórico y no provee una metodología para hallar las funciones y coeficientes necesarios para sintetizar una cierta función.

La representación canónica lineal a tramos de alto nivel (CLATAN) para la aproximación de funciones continuas propuesta en [48], es la primera expresión lineal a tramos capaz de representar mapeos en dominios de dimensión arbitraria, y es luego utilizada en esta tesis para describir el mapeo estático no lineal de las estructuras tipo Wiener propuestas para la aproximación de un sistema dinámico no lineal. Esta representación permite caracterizar numéricamente una función LAT de manera muy eficiente, ya que utiliza el número mínimo y exacto de parámetros necesario para representar una función LAT continua arbitraria, en comparación con otras representaciones LAT.

En consecuencia, a continuación damos una descripción de esta representación y sus características más importantes. Para más detalles, el lector puede consultar [48].

Una función LAT  $f : \mathbf{S} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ , donde  $\mathbf{S}$  es un conjunto compacto, se define como sigue.

**Definición 2.7.1** (Función LAT) Una función  $f : \mathbf{S} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ , siendo  $\mathbf{S}$  un conjunto compacto, se dice que es lineal a tramos (LAT) si y sólo si satisface

- (i) El dominio  $\mathbf{S}$  está dividido en un número finito de regiones poliédricas  $R^{(1)}, R^{(2)}, \dots, R^{(N)}$  tales que  $\mathbf{S} = \cup_{i=1}^N \overline{R}^{(i)}$ , por un conjunto (finito) de fronteras

$$H = \{H_i \subset \mathbf{S}, i = 1, 2, \dots, h\}, \quad (2.54)$$

tal que cada frontera es un hiperplano  $(n - 1)$ -dimensional (o un subconjunto del mismo)

$$H_i = \{\mathbf{x} \in \mathbb{R}^n : \pi_i(\mathbf{x}) = \boldsymbol{\alpha}_i^T \mathbf{x} - \beta_i = 0\}, \quad (2.55)$$

donde  $\boldsymbol{\alpha}_i \in \mathbb{R}^n$  y  $\beta_i \in \mathbb{R}$  para  $i = 1, 2, \dots, h$  y no puede ser cubierta<sup>3</sup> por un hiperplano de dimensión  $n - 2$ .

- (ii)  $f$  se expresa mediante la representación lineal afín

$$f^{(i)}(\mathbf{x}) = J^{(i)}\mathbf{x} + \mathbf{w}^{(i)}, \quad (2.56)$$

para cualquier  $\mathbf{x} \in R^{(i)}$ , donde  $J^{(i)} \in \mathbb{R}^{m \times n}$  es llamada la matriz Jacobiana de la región  $R^{(i)}$  y  $\mathbf{w}^{(i)} \in \mathbb{R}^m$ .

- (iii)  $f$  es continua en cualquier frontera de dos regiones vecinas, es decir,

$$J^{(p)}\mathbf{x} + \mathbf{w}^{(p)} = J^{(q)}\mathbf{x} + \mathbf{w}^{(q)}, \quad (2.57)$$

para cualquier  $\mathbf{x} \in \overline{R}^{(p)} \cap \overline{R}^{(q)}$ .

---

<sup>3</sup>Una frontera se dice *cubierta* por un hiperplano  $H$  si y sólo si  $B \subset H$

**Definición 2.7.2** (Símplice) Sean  $\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n$ ,  $n + 1$  puntos en  $\mathbb{R}^n$ . Un símlice  $\Delta(\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n)$  se define como

$$\Delta(\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^n) = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \sum_{i=0}^n \mu_i \mathbf{x}^i \right\}, \quad (2.58)$$

donde  $0 \leq \mu_i \leq 1, i \in \{1, \dots, n\}$  y  $\sum_{i=0}^n \mu_i = 1$ . Un símlice se dice *propio* si y sólo si no está contenido en un hiperplano de dimensión  $(n - 1)$ .

Una *partición simplicial* de un dominio compacto  $\mathbf{S} \subset \mathbb{R}^n$  es un conjunto de hiperplanos que subdividen el dominio en símlices.

Consideremos un dominio compacto en  $\mathbb{R}^n$  de la forma:

$$\mathbf{S} \doteq \{ \mathbf{x} \in \mathbb{R}^n : 0 \leq x_i \leq \delta \, n \text{div}_i, i = 1, 2, \dots, n \}, \quad (2.59)$$

donde  $\delta$  y  $n \text{div}_i \in \mathbb{Z}_+$  son el tamaño de la grilla y el número de subdivisiones, respectivamente, asociadas con el eje  $x_i$ .

Cada componente dimensional  $x^i$  del dominio compacto  $\mathbf{S}$  puede subdividirse en  $n \text{div}_i$  subintervalos de amplitud  $\delta$ . En consecuencia,  $\mathbf{S}$  queda particionado en  $\prod_{i=1}^n n \text{div}_i$  hiperrectángulos y contiene  $\prod_{i=1}^n n \text{div}_i + 1$  vértices.

Podemos subdividir el dominio compacto  $\mathbf{S}$  en símlices utilizando una partición simplicial con *configuración de fronteras*  $H$ , la cual está definida por el conjunto de hiperplanos

$$\begin{aligned} & \{ \mathbf{x} \in \mathbb{R}^n : x_q - k_q \delta = 0 \}, \\ & \{ \mathbf{x} \in \mathbb{R}^n : x_i - (x_j - k_j \delta) = 0 \}, \\ & \{ \mathbf{x} \in \mathbb{R}^n : (x_i - k_i \delta) - x_j = 0 \}, \end{aligned} \quad (2.60)$$

y la frontera  $\partial \mathbf{S}$ , donde  $q = 1, 2, \dots, n$ ,  $k_q = 1, 2, \dots, m_q - 1$ ,  $i \in \{1, \dots, n - 1\}$ ,  $j \in \{i + 1, \dots, n\}$ ,  $k_i = 1, \dots, m_i - 1$  y  $k_j = 0, 1, \dots, m_j - 1$ .

Este tipo de partición subdivide al dominio compacto  $\mathbf{S} \subset \mathbb{R}^n$  en un conjunto de símlices propios  $S^{(i)}$ ,  $i = 1, 2, \dots, q$ , de  $n + 1$  vértices de manera tal que  $\mathbf{S} = \cup_{i=1}^q \overline{S^{(i)}}$ .

Llamamos  $PWL_H[\mathbf{S}]$  al conjunto de todas las funciones LAT continuas definidas sobre una partición simplicial con configuración de fronteras  $H$  del dominio  $\mathbf{S}$ .

Definimos la suma de funciones pertenecientes a  $PWL_H[\mathbf{S}]$  y el producto de una función perteneciente a  $PWL_H[\mathbf{S}]$  por un escalar  $\alpha \in \mathbb{R}$  como:

$$\begin{aligned} (f + g)(\mathbf{x}) &= f(\mathbf{x}) + g(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{S} \\ (\alpha f)(\mathbf{x}) &= \alpha f(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbf{S}, \end{aligned} \quad (2.61)$$

entonces  $PWL_H[\mathbf{S}]$  es un espacio vectorial.

El siguiente ejemplo ilustra como son las funciones del espacio vectorial  $PWL_H[\mathbf{S}]$ . Supongamos que el dominio  $\mathbf{S} \subset \mathbb{R}^2$ . Definimos una configuración de fronteras  $H$  tal que el

dominio  $\mathbf{S}$  quede particionado en simplices propios. Luego, si asociamos un valor de funcion a cada vertice, es posible determinar una unica funcion lineal (local) para cada simplice de modo tal que la coleccion de todas las funciones determina una funcion LAT continua. Ver Figura 2.7.

Observemos que cualquier funcion LAT arbitraria  $f : \mathbf{S} \rightarrow \mathbb{R}$  queda univocamente determinada por los valores que toma la funcion sobre los vertices de la particion simplicial de  $\mathbf{S}$ . Este procedimiento resultara de suma utilidad para la aproximacion de funciones no lineales continuas.

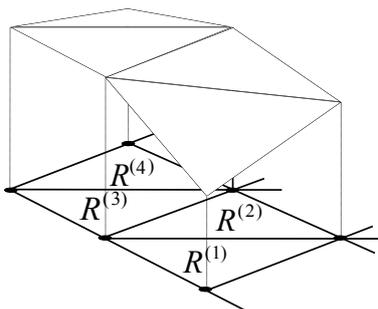


Figura 2.7: Funcion LAT con dominio  $\mathbf{S} \subset \mathbb{R}^2$

De acuerdo con [48], podemos construir una base para el espacio vectorial  $PWL_H[\mathbf{S}]$  a partir de sucesivas composiciones de una unica funcion generadora.

Sea  $C$  el espacio de las funciones continuas que mapean  $\mathbf{S} \subset \mathbb{R}^n$  en  $\mathbb{R}$  y sea  $\mathcal{F}$  el conjunto de funciones

$$\mathcal{F} = \{f_i : f_i \in C, 1 \leq i \leq p\}.$$

La funcion generadora  $\gamma : \mathcal{F} \times \mathcal{F} \rightarrow C$  se define como

$$\gamma(f_i, f_j) \doteq 1/4\{|-f_i + f_j| - |-f_i + |f_j|| + |-f_i| + |f_j| - |-f_i + f_j|\}. \quad (2.62)$$

La base esta compuesta por una funcion lineal  $\gamma^0 : \mathcal{F} \rightarrow C$ , y un conjunto de funciones no lineales  $\gamma^k : \mathcal{F} \times \dots \times \mathcal{F} \rightarrow C$ , con  $k \geq 1$ , definidas a partir de sucesivas composiciones de la funcion generadora  $\gamma$ :

$$\gamma^0(f_r) \doteq f_r, \gamma^1(f_r) \doteq \gamma(f_r, f_r), \gamma^2(f_{r_1}, f_{r_2}) \doteq \gamma(f_{r_1}, f_{r_2})$$

y, en general,

$$\gamma^k(f_{r_1}, f_{r_2}, \dots, f_{r_k}) = \gamma(f_{r_1}, \gamma^{k-1}(f_{r_2}, \dots, f_{r_k})). \quad (2.63)$$

Si la dimension del dominio es igual a  $n$ , solamente deben considerarse las funciones  $\gamma^k(\cdot, \dots, \cdot)$  con  $k = 0, 1, \dots, n$ .

Una elección apropiada de  $\mathcal{F}$  para el caso de funciones LAT es

$$\mathcal{F} = \{1, \pi_i \in C, 1 \leq i \leq h\},$$

donde  $\pi_i$  son las funciones lineales de la forma (2.55). Notemos que los hiperplanos asociados con esas funciones lineales pertenecen a la partición simplicial de  $\mathbf{S}$ .

Entonces, el primer elemento de la base es la función constante  $\gamma^0(1)$  y los elementos restantes están formados por la composición de la función  $\gamma^k(\cdot, \dots, \cdot)$  con las funciones lineales  $\pi_i$ :

$$\gamma^k(\pi_{r_1}, \pi_{r_2}, \dots, \pi_{r_k}) = \gamma(\pi_{r_1}, \gamma^{k-1}(\pi_{r_2}, \dots, \pi_{r_k})). \quad (2.64)$$

Una característica distintiva de (2.64) es que posee  $k$  anidamientos de valores absolutos de funciones y de acuerdo a esto, se dice que tiene *nivel de anidamiento* (*n.a.*) igual a  $k$ .

Las funciones de la base pueden escribirse en forma vectorial como

$$\Lambda_{\mathbf{S}} = [\Lambda^{0^T}, \Lambda^{1^T}, \dots, \Lambda^{n^T}]^T, \quad (2.65)$$

ordenadas de acuerdo a su nivel de anidamiento, donde cada  $\Lambda^i$  es el vector que contiene todas las funciones con *n.a.* =  $i$  y  $T$  simboliza la operación de trasponer un vector o una matriz. Entonces se tiene el siguiente resultado fundamental.

**Teorema 2.7.3** *Las funciones LAT que componen la matriz  $\Lambda_{\mathbf{S}}$  son una base del espacio vectorial  $PWL_H[\mathbf{S}]$ .*

**Demostración:** La demostración puede consultarse en [49, 51]. □

Como consecuencia del Teorema (2.7.3), sigue que cualquier  $f \in PWL_H[\mathbf{S}]$  puede escribirse como:

$$f(\mathbf{x}) = \mathbf{c}^T \Lambda_{\mathbf{S}}(\mathbf{x}), \quad (2.66)$$

donde  $\mathbf{c} = [\mathbf{c}^{0^T}, \mathbf{c}^{1^T}, \dots, \mathbf{c}^{n^T}]^T$  y cada vector  $\mathbf{c}^i$  es un vector de parámetros asociado a la función vectorial con nivel de anidamiento  $i$ ,  $\Lambda^i$ . Como cualquier función LAT queda determinada unívocamente por sus valores en los vértices (ver [48, 51]), las filas de la matriz  $\Lambda_{\mathbf{S}}$  se hallan calculando las funciones  $\Lambda^i$  en los vértices de  $\mathbf{S}$  con configuración simplicial de fronteras  $H$ .

Es importante notar que en virtud del Teorema 2.7.3, la expresión CLATAN (2.66) posee el *mínimo y exacto* número de parámetros necesarios para representar una función LAT genérica  $f \in PWL_H[\mathbf{S}]$ , obteniéndose una representación canónica con funciones LAT. Desde el punto de vista computacional, las expresiones canónicas halladas mediante la metodología propuesta en [48, 51] da un procedimiento concreto para obtener el número de funciones necesarias para la representación y los coeficientes asociados. Otra propiedad relevante del método es que la matriz asociada al sistema resulta una matriz triangular inferior.

Veamos un ejemplo en  $\mathbb{R}^2$  que ilustra como es una base de estas funciones.

**Ejemplo 2.7.4** Sea  $\mathbf{S} = [-1, 1] \subset \mathbb{R}^2$  y sea  $\delta = 1$ . Los hiperplanos

$$\begin{aligned} &\{\mathbf{x} \in \mathbb{R}^2 : \pi_1 = x_1 = 0\}, \\ &\{\mathbf{x} \in \mathbb{R}^2 : \pi_2 = x_2 = 0\}, \\ &\{\mathbf{x} \in \mathbb{R}^2 : \pi_{1,2}^+ = x_1 + x_2 = 0\}, \\ &\{\mathbf{x} \in \mathbb{R}^2 : \pi_{1,2}^- = x_1 - x_2 = 0\} \end{aligned} \tag{2.67}$$

y la frontera  $\partial[-1, 1]^2$  determinan la configuración de fronteras  $H$ . Los ocho símlices propios en que queda dividido  $\mathbf{S}$  se pueden ver en la siguiente Figura 2.8. Las funciones  $\gamma^1$  y  $\gamma^2$  con

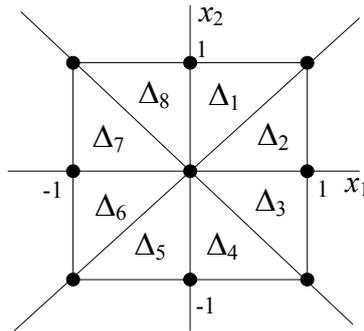


Figura 2.8: Configuración de fronteras para el hipercubo  $[-1, 1]^2$

$n.a. = 1$  y  $n.a. = 2$  respectivamente, se pueden ver en la Figura 2.9 y su comportamiento se describe en la Tabla 2.1.

	$n.a.$	Región donde la función es no nula
$\gamma^1(x_1)$	1	$\{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0\}$
$\gamma^1(-x_1)$	1	$\{\mathbf{x} \in \mathbb{R}^2 : x_1 \leq 0\}$
$\gamma^1(x_2)$	1	$\{\mathbf{x} \in \mathbb{R}^2 : x_2 \geq 0\}$
$\gamma^1(-x_2)$	1	$\{\mathbf{x} \in \mathbb{R}^2 : x_2 \leq 0\}$
$\gamma^1(x_1, x_2)$	2	$\{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\}$
$\gamma^1(-x_1, x_2)$	2	$\{\mathbf{x} \in \mathbb{R}^2 : x_1 \leq 0, x_2 \geq 0\}$
$\gamma^1(x_1, -x_2)$	2	$\{\mathbf{x} \in \mathbb{R}^2 : x_1 \geq 0, x_2 \leq 0\}$
$\gamma^1(-x_1, -x_2)$	2	$\{\mathbf{x} \in \mathbb{R}^2 : x_1 \leq 0, x_2 \leq 0\}$

Tabla 2.1: Funciones con  $n.a.=1$  y  $n.a.=2$  y su comportamiento en un dominio perteneciente a  $\mathbb{R}^2$

Luego,

$$\begin{aligned} \Lambda^0 &= \{\gamma^0(1)\}, \\ \Lambda^1 &= \{\gamma^1(x_1), \gamma^1(-x_1), \gamma^1(x_2), \gamma^1(-x_2)\}, \\ \Lambda^2 &= \{\gamma^1(x_1, x_2), \gamma^1(-x_1, x_2), \gamma^1(x_1, -x_2), \gamma^1(-x_1, -x_2)\}. \end{aligned} \tag{2.68}$$

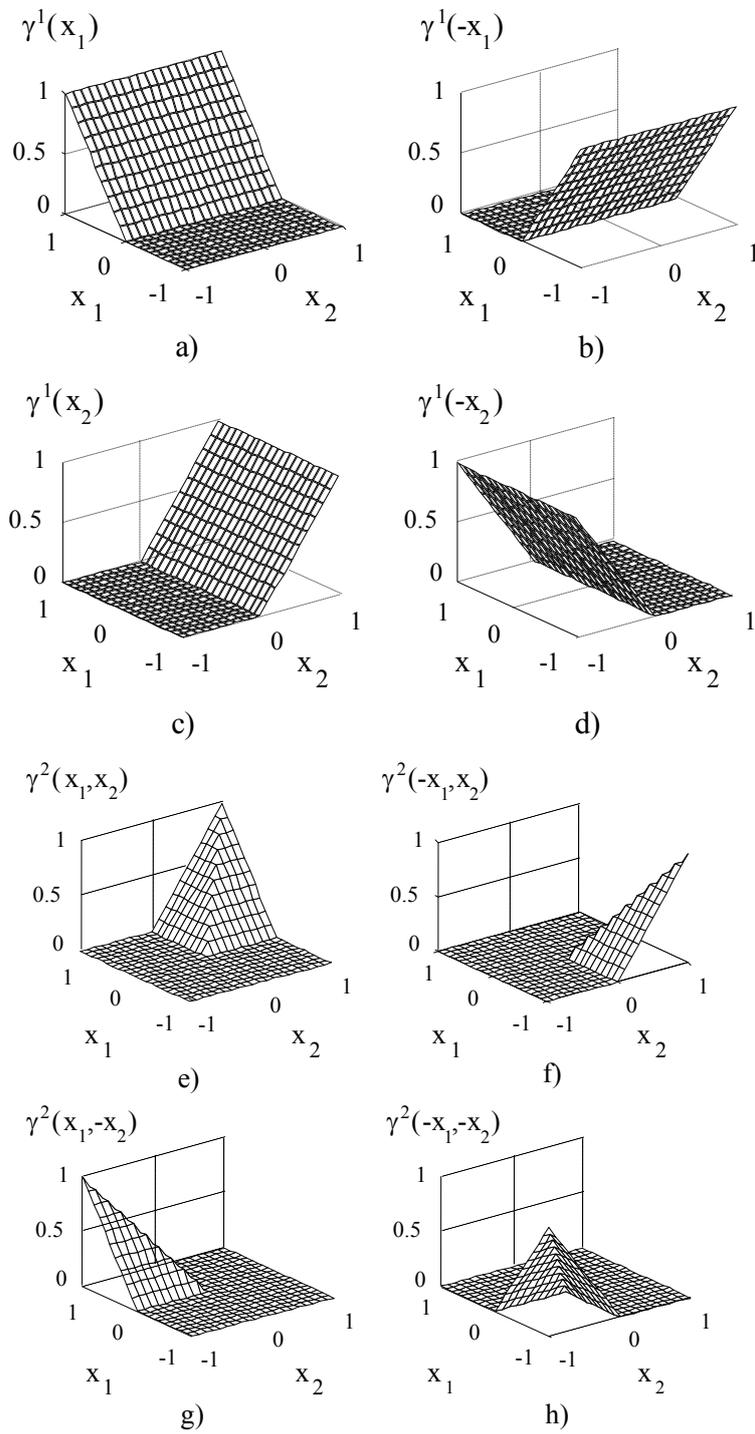


Figura 2.9: Funciones con  $n.a. = 1$  (a,b,c,d) y  $n.a. = 2$  (e,f,g,h) en  $\mathbb{R}^2$

Finalmente, la subdivisión producida por los hiperplanos (2.67) sobre  $\mathbf{S}$  determina un conjunto  $\mathbf{x}$  de  $\prod_{i=1}^2 (ndiv_i + 1) = 9$  vértices de la forma

$$\mathbf{x} = \{(0, 0), (1, 0), (-1, 0), (0, 1), (0, -1), (1, 1), (1, -1), (-1, 1), (-1, -1)\}.$$

Las filas de la matriz  $\Lambda_{\mathbf{S}}(\mathbf{x})$  resultan de evaluar cada vector de funciones en el conjunto  $\mathbf{x}$ , obteniendo así:

$$\Lambda_{\mathbf{S}}(\mathbf{x}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.69)$$

Una característica relevante de las funciones CLATAN es que pueden aproximar uniformemente cualquier función Lipschitz continua  $g : \mathbf{S} \mapsto \mathbb{R}$  [47]. De hecho, si la aproximación LAT de la función no lineal  $g$  se define como la función  $f_{lat} \in PWL_H[\mathbf{S}]$  que satisface

$$f_{lat}(\mathbf{v}^i) = g(\mathbf{v}^i),$$

donde  $\mathbf{v}^i$  son los vértices <sup>4</sup> de la partición simplicial con configuración de fronteras  $H$ , entonces el error de aproximación definido como

$$\varepsilon \doteq \sup_{\mathbf{x} \in \mathbf{S}} |f_p(\mathbf{x}) - g(\mathbf{x})| \quad (2.70)$$

satisface  $\varepsilon \leq \delta \cdot L$ , siendo  $L$  la constante Lipschitz de  $g(\cdot)$ .

### 2.7.2. Aproximación de una Función Continua Conocida

Consideramos funciones no lineales continuas  $f : \mathbf{S} \mapsto \mathbb{R}$ , donde  $\mathbf{S}$  es un dominio compacto definido en (2.59) particionado mediante una configuración simplicial de fronteras  $H_S$  cuyo tamaño de grilla es  $\delta$ .

Asumimos que conocemos la expresión analítica de la función  $f$ , aunque sólo basta conocer los valores de  $f$  sobre los vértices  $\mathbf{v}^i \in V_S$ , dado que esta es toda la información que necesitamos para definir una función perteneciente al espacio vectorial  $PWL_H[\mathbf{S}]$ .

---

<sup>4</sup>Los vértices del dominio (2.59) son los puntos de la forma  $\{(x_1, x_2, \dots, x_n) \in \mathbf{S} : x_i = k_i \cdot \delta\}$ , con  $k_i \in Z$ ,  $0 \leq k_i \leq m_i$  for  $i = 1, 2, \dots, n$ .

La metodología se basa en la aproximación LAT de una función, definida en [24]. En consecuencia, la aproximación CLATAN de  $f$  se define como la función  $f_p \in PWL_H[\mathbf{S}]$  que reproduce los valores de  $f$  sobre los vértices de la partición,

$$f_p(\mathbf{p}) = f(\mathbf{p}), \quad \forall \mathbf{p} \in V_S.$$

El siguiente resultado debido a Chien [24] permite cuantificar la precisión de la aproximación cuando se asume que la función es continua.

**Lema 2.7.5** *Si  $f$  es continua en  $\mathbf{S}$ , que es la unión de símplexes no superpuestos, entonces*

$$\|f_p(\mathbf{x}) - f(\mathbf{x})\| \leq \varepsilon, \quad \forall \mathbf{x} \in \mathbf{S},$$

donde

$$\varepsilon = \max_{\Delta \in \mathbf{S}} \left( \max_{\mathbf{x}_0, \mathbf{x}_1 \in \Delta} \|f(\mathbf{x}_1) - f(\mathbf{x}_0)\| \right). \quad (2.71)$$

Si asumimos que la función  $f$  es Lipschitz continua en  $\mathbf{S}$ , entonces es posible obtener condiciones útiles que relacionan el error de aproximación con el tamaño de la grilla  $\delta$ .

**Lema 2.7.6** *Sea  $f : \mathbf{S} \mapsto \mathbb{R}$  que satisfice*

$$\|f(\mathbf{x}_1) - f(\mathbf{x}_0)\| \leq L \|\mathbf{x}_1 - \mathbf{x}_0\|, \quad (2.72)$$

$$\forall \mathbf{x}_0, \mathbf{x}_1 \in \mathbf{S}.$$

Luego la aproximación CLATAN  $f_p$  de  $f$  satisfice

$$\|f_p(\mathbf{x}) - f(\mathbf{x})\| \leq \delta L \left\| \sum_{i=0}^n \mathbf{e}_i \right\|, \quad \forall \mathbf{x} \in \mathbf{S}. \quad (2.73)$$

**Demostración:** La demostración puede consultarse en [47]. □

### 2.7.3. Aproximación de Funciones a partir de un Conjunto de Datos

Supongamos que la única información disponible para describir el comportamiento de un sistema dinámico es un conjunto finito de mediciones de entrada-salida. Si deseamos hallar una función que sintetice los datos medidos, debemos seleccionar una estructura de modelado adecuada y luego debemos formular un criterio de error para hallar la aproximación de acuerdo a ese criterio. Veremos diferentes tipos de aproximaciones que se pueden realizar mediante funciones CLAT cuando se tienen sólo datos de entrada-salida de una función desconocida que representa un sistema dinámico.

Supongamos entonces que se dispone de un conjunto finito de datos de entrada:

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}, \quad (2.74)$$

donde  $\mathbf{x}_i \in \mathbf{S} \subset \mathbb{R}^n$  para  $i = 1, 2, \dots, m$  y que, asociado a este conjunto, existe un conjunto de valores de función

$$F = \{f_1, f_2, \dots, f_m\}, \quad (2.75)$$

los cuales pueden pensarse como los valores medidos de una función desconocida  $f : \mathbf{S} \rightarrow \mathbb{R}$ , de manera tal que

$$f_i = f(\mathbf{x}_i), \quad i = 1, \dots, m. \quad (2.76)$$

Luego, para aproximar el conjunto de valores (2.75) se propone una función CLATAN  $f_{lat} \in PWL_H[\mathbf{S}]$  y se plantea un criterio de error a través de la elección de una función objetivo adecuada. En la Sección 2.7.4, se utiliza un criterio de error basado en la norma  $\|\cdot\|_\infty$ , y en la Sección 2.7.5 el criterio de error utilizado se basa en la norma  $\|\cdot\|_2$ . Finalmente, en la Sección 2.7.6 se describe una extensión del enfoque de la Sección 2.7.4 que permite obtener dos funciones LAT  $f_{p1}, f_{p2} \in PWL_H[\mathbf{S}]$  que acotan en forma óptima los valores medidos (2.75).

#### 2.7.4. Aproximación Óptima utilizando Norma Infinito

En este caso, la función objetivo es el máximo valor absoluto del error entre  $f_{lat}$  y  $f$  sobre todos los datos del conjunto  $F$ , es decir

$$E_\infty = \max_{\mathbf{x}_i \in X} (|f_i - f_{lat}(\mathbf{x}_i)|). \quad (2.77)$$

Luego, este problema de optimización es un problema del tipo min-max. Si recordamos que  $f_{lat}(\mathbf{x}) = \mathbf{c}^T \Lambda(\mathbf{x})$ ,  $\mathbf{c} \in \mathbb{R}^{q_N}$ , donde  $q_N$  es el número de vértices de  $\mathbf{S}$ , entonces (2.77) puede escribirse de la siguiente forma equivalente

$$\min_{\mathbf{c} \in \mathbb{R}^{q_N}} \left( \max_{\mathbf{x}_i \in X} \{|f_i - \mathbf{c}^T \Lambda(\mathbf{x}_i)|\} \right). \quad (2.78)$$

El problema es no diferenciable y una técnica para su resolución es transformarlo en uno de optimización lineal [100].

**Lema 2.7.7** Sean  $X, F, H$  y  $\mathbf{S}$  en las condiciones antes mencionadas. Entonces, el problema (2.78) puede plantearse como el siguiente problema de programación lineal:

$$\begin{aligned} \min_{s.a} \quad & \lambda \\ & \mathbf{c}^T \Lambda(\mathbf{x}_i) + \lambda \geq f_i, \forall \mathbf{x}_i \in X, \\ & \mathbf{c}^T \Lambda(\mathbf{x}_i) - \lambda \leq f_i, \forall \mathbf{x}_i \in X, \\ & \lambda \geq 0, \end{aligned}$$

en los parámetros  $\mathbf{c}$  y  $\lambda$ .

**Demostración:** Consultar [100]. □

### 2.7.5. Aproximación Óptima utilizando Norma Cuadrática

En este caso, la función objetivo es la siguiente:

$$E_2 = \sum_{i=1}^n (f_i - f_{lat}(\mathbf{x}_i))^2. \quad (2.79)$$

Nuevamente, si se tiene en cuenta que la expresión de una función arbitraria  $f_{lat} \in PWL_H[\mathbf{S}]$  está dada por  $f_{lat}(\mathbf{x}) = \mathbf{c}^T \Lambda(\mathbf{x})$ , entonces (2.79) puede escribirse como

$$E_2 = \sum_{i=1}^n (f_i - \mathbf{c}^T \Lambda(\mathbf{x}_i))^2, \quad (2.80)$$

o, en forma vectorial, como

$$E_2 = \left( \mathbf{y} - \Lambda(X)^T \mathbf{c} \right)^T \left( \mathbf{y} - \Lambda(X)^T \mathbf{c} \right), \quad (2.81)$$

donde  $\mathbf{y} \in \mathbb{R}^d$  se define como  $\mathbf{y} = [f_1, f_2, \dots, f_m]^T$ .

La minimización de (2.80) implica la resolución de la ecuación

$$\frac{\partial E_2}{\partial \mathbf{c}} = -\mathbf{y} \Lambda(X) + \mathbf{c}^T \Lambda(X) \Lambda(X)^T = 0,$$

que puede obtenerse directamente como

$$\mathbf{c} = \left( \Lambda(X) \Lambda(X)^T \right)^{-1} \Lambda(X) \mathbf{y}$$

En este caso, la solución del problema requiere la inversión de una matriz simétrica de  $q_N \times q_N$  elementos.

### 2.7.6. Aproximación Robusta utilizando Norma Infinito

En este caso, se supone que la función medida tiene errores de medición. En el caso considerado, diremos que la función medida es incierta si se puede caracterizar de la siguiente manera

$$\Xi = \{f : \mathbf{S} \rightarrow \mathbb{R} : f(\mathbf{x}) = f_N(\mathbf{x}) + \Delta(\mathbf{x})\}, \quad (2.82)$$

donde  $f$  y  $\Delta$  son funciones continuas,  $f_N$  es una función nominal y  $\Delta$  satisface  $\sup_{\mathbf{x} \in \mathbf{S}} \|\Delta(\mathbf{x})\| \leq K$ . Además, como en los casos anteriores, suponemos que se dispone de un conjunto de valores medidos  $F = \{f_1, f_2, \dots, f_n\}$  tales que  $f_i = f(\mathbf{x}_i)$ , con  $f \in \Xi$ ,  $\mathbf{x}_i \in \mathbf{S}$ . En este caso, es necesario modelar un *conjunto de funciones* y por lo tanto un planteo natural es hallar dos funciones  $f_{p_1}$  y  $f_{p_2}$  en  $PWL_H[\mathbf{S}]$  que satisfacen

$$f_{lat_1}(\mathbf{x}_i) \leq f(\mathbf{x}_i) \leq f_{lat_2}(\mathbf{x}_i), \quad \forall \mathbf{x}_i \in \mathbf{S},$$

de modo que  $f$  se pueda representar como

$$f(\mathbf{x}_i) = \alpha f_{lat_1}(\mathbf{x}_i) + (1 - \alpha) f_{lat_2}(\mathbf{x}_i), \quad 0 \leq \alpha \leq 1,$$

para todo  $\mathbf{x}_i \in X$  y para toda  $f \in \Xi$ . De esta manera es posible obtener una caracterización de  $f$  en términos de  $f_{lat_1}$  y  $f_{lat_2}$ .

Además, para lograr una caracterización óptima, se pretende que la banda definida por estas dos funciones sea lo más angosta posible. Es decir, que las funciones  $f_{lat_1}$  y  $f_{lat_2}$  resuelvan los siguientes problemas de optimización:

**Problema 1:**

$$\min_{f_{lat_1} \in PWL_H[\mathbf{S}]} \left( \max_{\mathbf{x}_i \in X} \{|f_i - f_{lat_1}(\mathbf{x}_i)|\} \right) \quad (2.83)$$

s.a

$$f_i - f_{lat_1}(\mathbf{x}_i) \geq 0, \quad \forall \mathbf{x}_i \in X.$$

**Problema 2:**

$$\min_{f_{lat_2} \in PWL_H[\mathbf{S}]} \left( \max_{\mathbf{x}_i \in X} \{|f_i - f_{lat_2}(\mathbf{x}_i)|\} \right) \quad (2.84)$$

s.a

$$f_{lat_2}(\mathbf{x}_i) - f_i \geq 0, \quad \forall \mathbf{x}_i \in X.$$

En forma análoga a la Sección 2.7.4, estos dos problemas del área de optimización no diferenciable pueden expresarse como problemas de programación lineal, asumiendo que las funciones  $f_{lat_1}$  y  $f_{lat_2}$  tienen la forma  $f_{lat_1}(\mathbf{x}) = \mathbf{c}_1^T \Lambda(\mathbf{x})$  y  $f_{lat_2}(\mathbf{x}) = \mathbf{c}_2^T \Lambda(\mathbf{x})$ , respectivamente.

**Lema 2.7.8** Sean  $X, F, H$  y  $\mathbf{S}$  en las condiciones antes mencionadas. Entonces, los problemas (2.83) y (2.84) puede formularse como los siguientes problemas de programación lineal:

**Problema 1:**

$$\begin{aligned} \min_{s.a} \quad & \lambda_1 \\ & -\mathbf{c}_1^T \Lambda(\mathbf{x}_i) - \lambda_1 \leq -f_i, \forall \mathbf{x}_i \in X, \\ & -\mathbf{c}_1^T \Lambda(\mathbf{x}_i) \geq -f_i, \forall \mathbf{x}_i \in X, \\ & \lambda_1 \geq 0, \end{aligned} \quad (2.85)$$

**Problema 2:**

$$\begin{aligned}
& \underset{s.a}{\text{mín}} \quad \lambda_2 \\
& \mathbf{c}_2^T \Lambda(\mathbf{x}_i) - \lambda_2 \leq f_i, \forall \mathbf{x}_i \in X, \\
& \mathbf{c}_2^T \Lambda(\mathbf{x}_i) \geq f_i, \forall \mathbf{x}_i \in X, \\
& \lambda_2 \geq 0,
\end{aligned} \tag{2.86}$$

en los parámetros  $\mathbf{c}_1, \mathbf{c}_2, \lambda_1$  y  $\lambda_2$ .

**Demostración:** Consultar [100]. □

## 2.8. Validación del Modelo

Una vez propuesta una estructura de modelo determinada para identificar un proceso real, debemos seleccionar el mejor modelo dentro de esta estructura, es decir, ajustar el conjunto de parámetros de la estructura que minimizan el error de aproximación.

La validación del modelo ajustado trata de determinar si el modelo obtenido es lo suficientemente bueno.

Existen dos formas de validar un modelo. La primera de ellas es la *validación cruzada*, que analiza el comportamiento del modelo sobre una base de datos no utilizados en el ajuste de los parámetros denominado *conjunto de validación*. Si el modelo cumple con los requisitos preestablecidos sobre este conjunto de datos, entonces el modelo se considera “bueno”.

La segunda está basada en el análisis de los residuos. Al ajustar un modelo lineal a un proceso descrito por la ecuación (2.12), se supuso que  $\nu(k)$  es una serie de variables aleatorias independientes de media cero y varianza constante (ruido estacionario), distribuida según una función de densidad de probabilidad determinada  $p_\nu$ , típicamente normal. La validación del modelo consiste en comprobar que la serie de errores de estimación

$$e(k) = y(k) - \tilde{y}(k)$$

cumple las hipótesis establecidas, es decir, que es una realización de una serie de variables aleatorias independientes e idénticamente distribuidas según  $p_\nu$ , de media nula y varianza constante.

## 2.9. Conclusiones

En este capítulo hemos presentado una clasificación de modelos dentro de la cual podremos incluir los propuestos en este trabajo. Introducimos, además, las bases de funciones lineales a tramos que nos permitirán definir los modelos en forma rigurosa.

En los Capítulos 3 y 4 describimos nuestros modelos y su relación con los modelos NOE y tipo Wiener, respectivamente. Establecemos los alcances de los esquemas propuestos en el sentido del tipo de sistemas que es posible aproximar por medio de los mismos y desarrollamos varios ejemplos de aplicación de nuestras metodologías.

Los modelos que proponemos son tipo caja negra pues nos basamos sólo en los datos de entrada-salida para hallar los parámetros de los mismos. Los regresores propuestos son los que denominamos semi-físicos.



## Capítulo 3

# Identificación de un Modelo NOE usando Funciones CLATAN

Uno de los principales problemas en la identificación de sistemas es hallar un modelo que sea representativo de las características dinámicas del proceso real que se está estudiando. Este problema es más difícil en estructuras de identificación que admiten desde un modelo lineal simple hasta uno no lineal más complicado, debido a que el conjunto de modelos no lineales es mucho más rico que el conjunto de los lineales [85]. Entonces, elegir el mejor modelo dentro de una estructura no lineal, *i.e.* evaluar el orden del modelo no es una tarea sencilla. En la literatura existen herramientas muy conocidas para abordar eficazmente el problema de la evaluación del orden del modelo para determinadas estructuras. Por ejemplo, la teoría de regularización [73] se puede aplicar a los modelos no lineales de respuesta finita al impulso NFIR, el método de agregación se puede utilizar para la selección del modelo en estructuras tipo Wiener como en el algoritmo de Korenberg [55]. Si se utilizan redes neuronales, los métodos de crecimiento y poda son utilizados para ajustar el tamaño de la red durante el entrenamiento [41]. Por otra parte, los modelos afines lineales a tramos han sido muy usados en la identificación de sistemas no lineales [9, 32, 89] como una clase interesante de modelos NARMAX parametrizados [21, 40, 46, 86].

Una estructura de identificación NOE permite evaluar modelos dinámicos aptos para su utilización como simuladores fuera de línea, ya que se minimiza el efecto del ruido en la estimación de los parámetros del mismo. Pero en este caso, la evaluación del orden se hace sumamente difícil debido a la complejidad computacional inherente a la naturaleza recursiva del modelo [69]. Por otro lado, variar el orden del modelo en los algoritmos de identificación de esta estructura implica reiniciar el proceso de identificación, lo que se traduce en la evaluación de un nuevo conjunto de parámetros a partir de uno previo, lo cual puede ser computacionalmente muy costoso.

En esta tesis consideramos una estructura de modelo NOE basada en funciones CLATAN [13, 18] e implementamos un algoritmo de identificación recursivo que ofrece un mecanismo sencillo para aumentar el orden del modelo, conservando la aproximación lograda hasta el

orden inmediato anterior. Para ello identificamos los parámetros de la estructura NOE basada en funciones CLATAN de orden creciente, partiendo de una aproximación lineal OE; de esta manera la estructura puede ir aumentando los grados de libertad de la función CLATAN, conservando la aproximación obtenida previamente hasta que se logra un error deseado entre el modelo y el sistema. El algoritmo está basado en un método de optimización de máximo descenso con una tasa de aprendizaje adaptativo que permite el control de la estabilidad BIBO (por sus siglas en inglés, Bounded-Input, Bounded-Output) del modelo. Un sistema dinámico es BIBO estable si cualquier entrada acotada produce una salida acotada. En otras palabras, si ante entradas de valor finito la respuesta (en valor absoluto) no tiende a infinito.

En este capítulo describimos en primer lugar la estructura de modelo y el correspondiente algoritmo de identificación NOE basado en funciones CLATAN. Luego exploramos las capacidades de aproximación de esta nueva estructura y su grado de generalización y obtenemos condiciones suficientes que aseguran la estabilidad BIBO del algoritmo de identificación. En la última sección presentamos algunos ejemplos de aplicación de la metodología propuesta para demostrar el potencial de nuestro enfoque.

### 3.1. Estructura del Modelo

Sean  $\mathbf{u}, \mathbf{y}$  los vectores E/S de longitud fija  $L$  que corresponden a un sistema SISO, Lipschitz continuo e  $I = [\underline{y}, \bar{y}] \subset \mathbb{R}$ , con  $\underline{y} = \min \mathbf{y}, \bar{y} = \max \mathbf{y}$ . Si  $\tilde{\mathbf{y}}$  es el valor estimado correspondiente a la entrada  $\mathbf{u}$ , notamos

$$\begin{aligned} u^{k,M+1} &= [u(k), \dots, u(k-M)] \\ \tilde{\mathbf{y}}^{k-1,N} &= [\tilde{y}(k-1), \dots, \tilde{y}(k-N)], \end{aligned} \quad (3.1)$$

donde  $u^{k,M+1} \in U \subset \mathbb{R}^{M+1}$ ,  $\tilde{\mathbf{y}}^{k-1,N} \in O \subset \mathbb{R}^N$ ,  $U$  y  $O$  conjuntos compactos dados.

La estructura de identificación propuesta está dada por

$$\begin{aligned} \tilde{y}(k) &= f_{lat}(u(k), \dots, u(k-M), \tilde{y}(k-1), \dots, \tilde{y}(k-N)) \\ &= \mathbf{c}\Lambda(u^{k,M+1}, \tilde{\mathbf{y}}^{k-1,N}), \end{aligned} \quad (3.2)$$

donde  $f_{lat}(\mathbf{x}) = \mathbf{c}\Lambda(\mathbf{x})$  es la función CLATAN como se definió en el Capítulo 2 y los órdenes del modelo  $M, N$  son dados. De la ecuación (3.2), resulta que la estructura de identificación propuesta es de tipo caja negra, siendo  $\mathbf{z}^k = [u^{k,M+1}, \tilde{\mathbf{y}}^{k-1,N}]$  el vector de regresión asociado a la misma. En la Figura 3.1 se puede observar el gráfico asociado a (3.2).

Como se definió en el Capítulo 2, Sección 2.7, el dominio de la función  $f_{lat}$  es el conjunto compacto  $\mathbf{S} \subset \mathbb{R}^m$ ,  $m = M + 1 + N$ , definido como

$$\mathbf{S} = \{\mathbf{x} \in \mathbb{R}^m : a_i \leq x_i \leq a_i + \delta \text{ndiv}, i = 1, 2, \dots, m\}, \quad (3.3)$$

donde  $\delta$  es la medida de la grilla y  $\text{ndiv} \in \mathbb{Z}_+$  es el número de subintervalos de igual longitud  $\delta$  asociados con el eje  $x_i$ . Cuando la medida  $\delta$  de la grilla decrece, el número de divisiones  $\text{ndiv}$

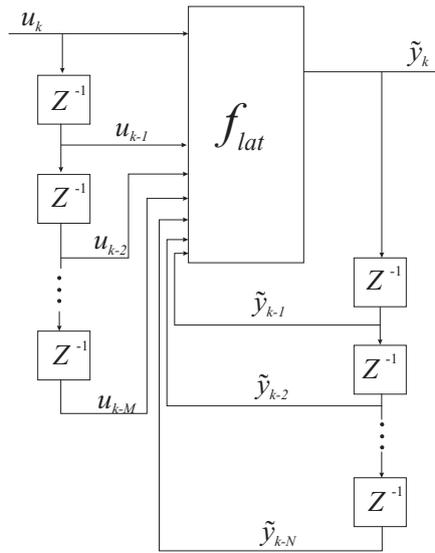


Figura 3.1: Modelo NOE-CLATAN

aumenta en cada dirección. Luego, cuando  $ndiv$  cambia, el modelo pasa de lineal ( $ndiv = 1$ ) a no lineal ( $ndiv > 1$ ) con una partición creciente más fina de  $\mathbf{S}$ . Las ventajas de esta clase de modelo se enumeran en [85].

El conjunto  $\mathbf{S}$  queda así particionado en regiones poliédricas usando una configuración de fronteras simplicial y la función  $f_{lat}$  se construye utilizando la metodología descrita en el Capítulo 2. La función resultante  $f_{lat}$  es lineal en cada símplice y continua en cada frontera adyacente de los símplices.

## 3.2. Algoritmo de Identificación

En [17] desarrollamos un algoritmo de identificación no lineal para esta estructura de modelo CLATAN que se inicializa mediante una aproximación lineal, que es un caso especial de función CLATAN. Posteriormente, el número de divisiones  $ndiv$  de la partición simplicial del dominio  $\mathbf{S}$  aumenta y el nuevo conjunto de parámetros se evalúa fácilmente utilizando un algoritmo simple (véase el Apéndice A). Por último, el vector de parámetros se actualiza utilizando un algoritmo de optimización adecuado, hasta que se cumple un criterio de parada. Este proceso consta entonces de tres pasos que pueden repetirse hasta lograr el error de aproximación deseado:

1. subdivisión de la grilla,
2. actualización de los parámetros, de acuerdo con la subdivisión de la nueva grilla,
3. optimización de los parámetros dados.

Sean  $M$ ,  $N$  dados los órdenes del modelo y  $\mathbf{S}$  el dominio compacto que contiene al conjunto completo de datos  $(u(k), y(k))$ ,  $1 \leq k \leq L$ . A continuación damos la notación que usaremos en el resto del capítulo.

### Notación

$ndiv = 2^d, d \geq 0$ : número de divisiones de la región  $\mathbf{S}$ . Asumimos el mismo número de divisiones en cada dirección.

$V^d$ : el conjunto de vértices de la partición simplicial  $H$  del conjunto  $\mathbf{S}$  donde  $ndiv$  es el número de divisiones.

$\Lambda^d$ : las bases CLATAN definidas sobre  $\mathbf{S}$  con vértices que pertenecen a  $V^d$ .

$\mathbf{c}^{d,*}$ : el vector de parámetros asociado con la mejor aproximación CLATAN obtenida usando las bases  $\Lambda^d$ . De [48, 50] puede verse que el número de parámetros es  $(ndiv + 1)^{M+1+N}$ .

$(A)_i$ : la  $i$ -ésima fila de la matriz  $A$ .

$Niter \in \mathbb{N}$ : máximo número de iteraciones del algoritmo de optimización.

$Maxerror$ : máximo error de aproximación permitido.

$lr^r$ : tasa de aprendizaje correspondiente a la iteración  $r$ ;  $lr_i^r > 0, \forall i$  (generalmente,  $lr_i = 0,0001$ ).

$mom$ : momento,  $mom > 0$  (generalmente,  $mom = 0,9$ ).

$lr_{crec}$ : tasa de aprendizaje del incremento,  $lr_{crec} > 1$  (generalmente,  $lr_{crec} = 1,05$ ).

$lr_{dec}$ : tasa de aprendizaje del decremento,  $0 < lr_{dec} < 1$  (generalmente,  $lr_{dec} = 0,9$ ).

### Algoritmo

*Paso 1.  $d = 0$ : Aproximación Lineal.*

Evaluar un modelo de aproximación lineal OE.

Calcular los parámetros  $\mathbf{c}^{d,*}$  de la representación CLATAN del hiperplano definido por los parámetros del modelo OE original. Esto es sencillo porque un hiperplano es un caso particularmente simple de una función CLATAN [48, 50].

*Paso 2.  $d \leftarrow d + 1$ : Evaluar  $\mathbf{c}^{d,*}$  a partir de  $\mathbf{c}^{d-1,*}$ .*

Evaluar  $\mathbf{c}^{d,*}$  de  $\mathbf{c}^{d-1,*}$  de acuerdo al algoritmo descrito en el Apéndice A, tomando inicialmente

$$r = 0, \quad \mathbf{c}^{d,r} = \mathbf{c}^{d,*}, \quad \Delta \mathbf{c}^{d,r} = [0, \dots, 0].$$

*Paso 3.  $r \leftarrow r + 1$ : Evaluación del error y del gradiente.*

Teniendo en cuenta las ecuaciones (3.1) y (3.2), la expresión del error en la iteración  $r$  es

$$E^r = \frac{1}{2} \sum_{i=1}^L [y(i) - \tilde{y}(i)]^2 = \frac{1}{2} \sum_{i=1}^L [y(i) - \mathbf{c}^{d,r-1} \Lambda^d (u^{i,M+1}, \tilde{y}^{i-1,N})]^2, \quad (3.4)$$

$$\begin{aligned} (\nabla E^r)_j &= \frac{\partial E^r}{\partial (\mathbf{c}_j^{d,r-1})} \\ &= - \sum_{i=1}^L [y(i) - \mathbf{c}^{d,r-1} \Lambda^d (u^{i,M+1}, \tilde{y}^{i-1,N})] (\Lambda^d (u^{i,M+1}, \tilde{y}^{i-1,N}))_j. \end{aligned} \quad (3.5)$$

*Paso 4. Actualización de los parámetros.*

Si  $E^r \leq Maxerror$  entonces FIN, en otro caso,

$$\Delta \mathbf{c}_j^{d,r} = \eta \left( -\nabla E_j^r l r_j^r + \Delta \mathbf{c}_j^{d,r-1} mom \right), \quad (3.6)$$

$$\mathbf{c}^{d,r} = \mathbf{c}^{d,r-1} + \Delta \mathbf{c}^{d,r}, \quad (3.7)$$

donde la constante  $\eta$  pertenece al intervalo definido por las ecuaciones (3.12) y (3.13) dadas en la Sección 3.2.1 y las componentes del vector tasa de aprendizaje  $\mathbf{lr}^r$  son modificadas como sigue

$$l r_j^r = \begin{cases} l r_j^{r-1} \times l r_{crec} & \text{si } \text{signo}(\nabla E_j^r) = \text{signo}(\nabla E_j^{r-1}) \\ l r_j^{r-1} \times l r_{dec} & \text{si } \text{signo}(\nabla E_j^r) \neq \text{signo}(\nabla E_j^{r-1}). \end{cases}$$

Si  $r < Niter$ , ir al Paso 3;

sino  $\mathbf{c}^{d,*} = \mathbf{c}^{d,r}$  e ir al Paso 2.

**Observación:** En el Paso 4, puede aplicarse cualquiera de las condiciones conocidas de terminación, basadas en la evolución del error.

El algoritmo propuesto garantiza la convergencia del método a un mínimo local. Al igual que en la optimización de muchos procesos, el mínimo alcanzado puede no ser un mínimo global, sino uno local; sin embargo, esta estructura de identificación ofrece numerosas ventajas

1. El cálculo del gradiente durante el proceso de optimización es sencillo y lineal en los parámetros.
2. La aproximación canónica CLATAN usa el mínimo número de parámetros.
3. El número de divisiones crece progresivamente, manteniendo la aproximación alcanzada. Además, introduce un grado de libertad en el proceso de identificación, con el objetivo de mejorar la calidad de la aproximación.
4. Las estructuras de identificación NOE son bien conocidas por su reducida sensibilidad al ruido.

### 3.2.1. Estabilidad BIBO del Modelo

Consideremos los conjuntos  $I$ ,  $U$  y  $O$  definidos en la Sección 3.1 y sea  $Q$  un subconjunto compacto de  $U \times O$ .

**Definición 3.2.1** Decimos que el modelo definido por (3.2) es BIBO estable si  $f_{lat}(Q) \subset I$ .

Esta definición establece que la salida del modelo permanece entre los valores de salida para cualquier señal de entrada  $\mathbf{u} \in U$ .

La expresión de la ecuación (3.2) define una transformación  $f_{lat} : Q \rightarrow I$ . Como  $Q$  es un conjunto compacto y  $f_{lat}$  es continua sobre  $Q$ , entonces alcanza sus valores máximo y mínimo sobre  $Q$ . Más aún, como  $f_{lat}$  es lineal en cada simplex, los valores extremos se obtienen sobre  $V_Q$ , el conjunto de vértices de  $Q$ . Entonces la estructura de identificación NOE dado por la ecuación (3.2) será BIBO estable si, para todo  $d$  fijo, las siguientes condiciones se verifican simultáneamente

$$\begin{cases} \min_{\mathbf{v} \in V^d} (\mathbf{c}^d \Lambda(\mathbf{v})) \geq \underline{y}, \\ \max_{\mathbf{v} \in V^d} (\mathbf{c}^d \Lambda(\mathbf{v})) \leq \bar{y}, \end{cases} \quad (3.8)$$

donde  $V^d$  se definió en la Sección 3.2.

Sea  $\mathbf{c}^d$  con  $d$  fijo, un vector de parámetros tal que el modelo es BIBO estable. Entonces, debemos garantizar que para cualquier  $r$ ,  $\tilde{y}(k) = \mathbf{c}^{d,r} \Lambda(\mathbf{z}^k) \in I$ ,  $1 \leq k \leq L$ , donde  $\mathbf{c}^{d,r}$  se obtiene del Paso 4 del algoritmo de identificación descrito en Sección 3.2.

La siguiente proposición nos da una condición suficiente para la estabilidad BIBO.

**Proposición 3.2.2** *Suponemos que para  $d$  y  $r-1$  fijos, el modelo es BIBO estable. Entonces el modelo será BIBO estable para  $d$  y  $r$  siempre que la siguiente condición se satisfaga*

$$\underline{y} - \min_{\mathbf{v} \in V^d} (\mathbf{c}^{d,r-1} \Lambda(\mathbf{v})) \leq \Delta \mathbf{c}^{d,r} \Lambda(\mathbf{v}) \leq \bar{y} - \max_{\mathbf{v} \in V^d} (\mathbf{c}^{d,r-1} \Lambda(\mathbf{v})), \quad (3.9)$$

donde  $\Delta \mathbf{c}^{d,r}$  está dada por la ecuación (3.6) y  $\mathbf{v} \in V^d$ .

**Demostración:** Teniendo en cuenta el modelo y la ecuación (3.7) obtenemos

$$\begin{aligned} \tilde{y}(k) &= \mathbf{c}^{d,r} \Lambda(\mathbf{z}^k) \\ &= \mathbf{c}^{d,r-1} \Lambda(\mathbf{z}^k) + \Delta \mathbf{c}^{d,r-1} \Lambda(\mathbf{z}^k). \end{aligned} \quad (3.10)$$

Hemos supuesto que  $\mathbf{c}^{d,r-1} \Lambda(\mathbf{z}^k) \in I$  para todo  $k$ . Luego, como el modelo alcanza sus valores extremos sobre  $V^d$  tenemos

$$\min_{\mathbf{v} \in V^d} \mathbf{c}^{d,r-1} \Lambda(\mathbf{v}) \leq \mathbf{c}^{d,r-1} \Lambda(\mathbf{z}^k) \leq \max_{\mathbf{v} \in V^d} \mathbf{c}^{d,r-1} \Lambda(\mathbf{v}). \quad (3.11)$$

Entonces, de (3.10) y (3.11),  $\tilde{y}(k)$  pertenece a  $I$  para cualquier  $k$ ,  $1 \leq k \leq L$  si la condición (3.9) se satisface.  $\square$

**Corolario 3.2.3** *Asumimos las mismas hipótesis que en la Proposición 3.2.2, y definimos*

$$\begin{aligned} a &= \underline{y} - \min_{\mathbf{v} \in V^d} (\mathbf{c}^{d,r-1} \Lambda(\mathbf{v})), \\ b &= \bar{y} - \max_{\mathbf{v} \in V^d} (\mathbf{c}^{d,r-1} \Lambda(\mathbf{v})). \end{aligned}$$

*Entonces las siguientes condiciones*

$$\eta \geq \frac{a}{\min_{\mathbf{v} \in V^d} [(-\nabla E^r \mathbf{l}^r + \Delta \mathbf{c}^{d,r-1} mom) \Lambda(\mathbf{v})]}, \quad (3.12)$$

$$\eta \leq \frac{b}{\max_{\mathbf{v} \in V^d} [(-\nabla E^r \mathbf{l}^r + \Delta \mathbf{c}^{d,r-1} mom) \Lambda(\mathbf{v})]}, \quad (3.13)$$

*se verifican simultáneamente.*

**Demostración:** La demostración sigue inmediatamente de la proposición anterior y del hecho que  $\Delta \mathbf{c}^{d,r-1} \Lambda(\mathbf{z}^k) \Lambda(\mathbf{z}) = \eta (-\nabla E^r) \Lambda(\mathbf{z})$  alcanza su valor extremo sobre  $V^d$ .  $\square$

**Observaciones:** De las cotas dadas por (3.12) y (3.13) para  $\eta$ , la única que tiene interés práctico es la menor positiva y es la utilizada en el Paso 4 del algoritmo de identificación.

El algoritmo garantiza la BIBO estabilidad del modelo siempre que el modelo lineal inicial sea BIBO estable [19].

### 3.3. Ejemplos de Aplicación

En esta sección presentamos cuatro ejemplos diferentes. Los dos primeros ilustran la transición de un modelo lineal a uno no lineal para sistemas con y sin ruido y la capacidad del modelo para extrapolar más allá de los datos, *i.e.* muestran que el modelo logra reproducir la dinámica de E/S fuera del dominio de la entrada. Los otros dos ejemplos muestran que la estructura posee atractivas propiedades de generalización, *i.e.* que hay poca discrepancia entre el sistema y los resultados del modelo en el conjunto de validación. Ambas características se deben, en parte, al aumento gradual del orden del modelo. Los ejemplos también muestran el rechazo al ruido, característico del algoritmo de identificación propuesto.

Las siguientes consideraciones se aplican a todos los ejemplos.

1. En el primer ejemplo utilizamos 50% de los datos para identificación y el resto para validación, mientras que en el segundo ejemplo consideramos el 90% de las muestras para identificar el modelo y el 10% para validar. En los dos últimos ejemplos, tomamos 70% de los datos disponibles para el proceso de identificación y el resto para el de validación.

2. Con el fin de obtener una representación gráfica clara, definimos el vector de regresión como  $\mathbf{z}^k = [u(k), \tilde{y}(k-1)]$ , *i.e.* tomamos  $M = 0$  y  $N = 1$ , de acuerdo con la metodología propuesta.
3. Definimos la entrada  $\mathbf{u}$  como una señal aleatoria con distribución uniforme.
4. Para sistemas con ruido, éste fue adicionado a las muestras usadas para la identificación, pudiendo así considerarse como una medición del ruido. Para mostrar con claridad la aproximación al sistema real, en la validación sólo se utilizaron muestras sin ruido.
5. De acuerdo al paso 1, evaluamos en primer lugar un modelo lineal OE para generar los parámetros iniciales; a continuación comenzamos con el algoritmo descrito en la Sección 3.1 con  $ndiv = 2$ . Los parámetros del modelo inicial de la estructura con  $ndiv = 2$  fueron evaluados con un modelo OE lineal con el algoritmo de actualización dado en el Apéndice A y fueron optimizados usando un número fijo de iteraciones. Luego,  $ndiv$  fue duplicado y los parámetros del nuevo modelo resultante fueron estimados a partir del previo que logra el mejor ajuste en el conjunto de validación, usando el algoritmo de actualización. Este procedimiento fue repetido hasta  $ndiv = 8$  con la excepción del Ejemplo 3 en el cual se repite hasta  $ndiv = 16$ . Luego, el número de parámetros  $(ndiv + 1)^2$  es 9, 25, 81 y 289, respectivamente.

### 3.3.1. Ejemplo 1: Modelo Logístico NOE

En este ejemplo consideramos el sistema dinámico logístico no lineal, discreto, con una entrada exógena altamente no lineal dada por

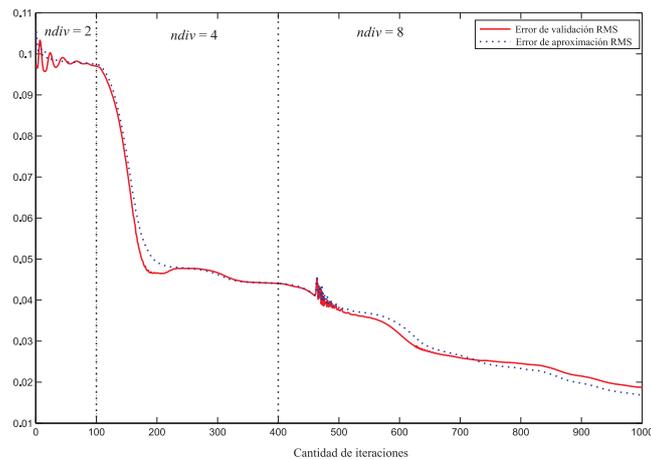
$$y(k+1) = 0,4 y(k) (1 - y(k)) + u(k)^7. \quad (3.14)$$

Con el propósito de resaltar las capacidades de modelado del algoritmo propuesto, optamos por elegir una entrada altamente no lineal, a pesar de que podría no estar asociada al sistema real.

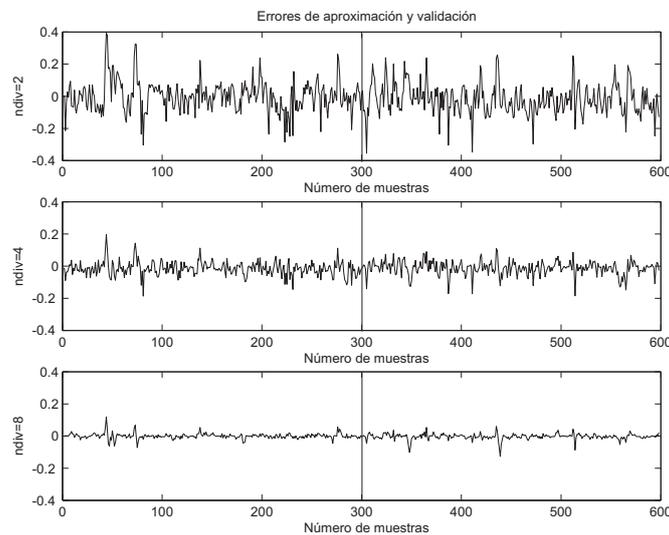
La entrada  $\mathbf{u}$  está definida como una señal aleatoria uniformemente distribuida entre -0.9 y 0.9 de 600 muestras. Utilizamos las primeras 300 para el proceso de identificación y las 300 restantes para validación.

La Figura 3.2 (a) muestra la aproximación y la validación del error RMS como una función del número de iteraciones del algoritmo de identificación. En las primeras 100 iteraciones el número de divisiones del espacio de regresores  $\mathbf{S}$  fue  $ndiv = 2$ ; luego  $ndiv$  se aumentó a 4. Los parámetros del nuevo modelo resultante fueron evaluados utilizando el algoritmo de actualización dado en el Apéndice A y esta estructura de modelo fue optimizada hasta la iteración 400. Finalmente, el mismo procedimiento fue repetido desde la iteración 401 hasta la iteración 1000 con  $ndiv = 8$ . Ambos errores RMS muestran un rápido descenso cuando el número de subdivisiones  $ndiv$  de  $\mathbf{S}$  cambia de 2 a 4, que es una característica

importante del enfoque propuesto. Al aumentar el grado de libertad de la estructura del modelo, el algoritmo de identificación mejora fácilmente la aproximación así como el error de validación, pues la nueva estructura del modelo que tiene más grados de libertad, comienza en la mejor aproximación inmediata anterior lograda con el algoritmo. Un efecto similar, con una leve reducción del error, puede apreciarse cuando el número de divisiones crece de 4 a 8. La Figura 3.2 (b) muestra los gráficos de la aproximación (muestras 1 a 300) y errores de validación (muestras 301 a 600) para los mejores modelos obtenidos con  $ndiv = 2, 4$  y 8, pueden apreciarse claramente las mejoras en el comportamiento del modelo cuando  $ndiv$  crece.



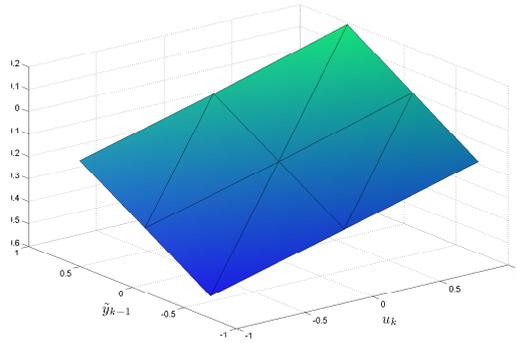
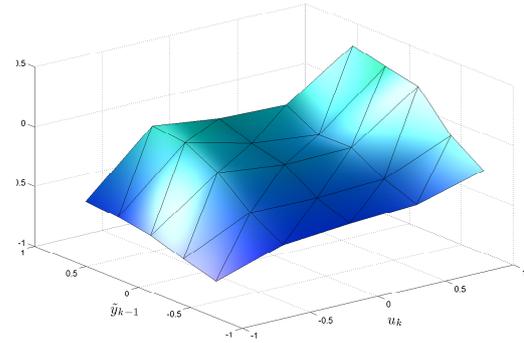
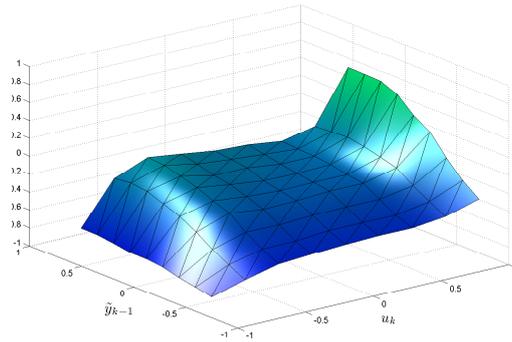
(a) Errores RMS de aproximación y validación para las funciones CLATAN usando  $ndiv = 2, 4$  y 8.



(b) Errores de aproximación (muestras 1 a 300) y de validación (muestras 301 a 600).

Figura 3.2: Características del error en el Ejemplo 3.3.1

La superficie del modelo  $f_{lat}$  resultante puede verse en la Figura 3.3. Es importante remarcar que dicha superficie con  $ndiv = 2$  se aproxima a una superficie plana, *i.e.* el modelo es casi un sistema lineal. La superficie  $f_{lat}$  mejora la aproximación a la superficie real cuando  $ndiv$  crece de 2 a 4 y de 4 a 8. La superficie real puede obtenerse de la ecuación (3.14), *i.e.* la forma es fuertemente no lineal en la dirección de  $u(k)$  y casi cuadrática en la dirección de  $\tilde{y}(k-1)$ .

(a)  $ndiv = 2$ .(b)  $ndiv = 4$ .(c)  $ndiv = 8$ .Figura 3.3:  $f_{lat}$  resultante del Ejemplo 3.3.1 con  $ndiv = 2, 4$  y  $8$ 

### 3.3.2. Ejemplo 2

Este ejemplo, debido a Yazdizdeh y Khorasani [104], está dado por las siguientes ecuaciones

$$\begin{aligned} y(k+1) &= \frac{y(k) + u(k)}{1 + y(k)^2} + \epsilon(k), \\ \epsilon(k+1) &= 0,25 (\epsilon(k) + \zeta); \end{aligned}$$

donde la entrada  $\mathbf{u}$  es una señal aleatoria de 2000 muestras con distribución uniforme entre -1 y 1,  $\zeta$  es un número uniformemente generado aleatoriamente entre -1 y 1; entonces  $\epsilon(k)$  puede considerarse ruido coloreado adicionado a la salida del sistema. Este ruido fue sumado a las primeras 1800 muestras utilizadas en el proceso de identificación; las restantes 200 muestras sin ruido fueron usadas para validar el modelo. En la Figura 3.4 los errores de aproximación y de validación también muestran mejoras en la aproximación cuando el número de divisiones  $ndiv$  crece. Se debe tener en cuenta que al conjunto de validación no se le adiciona ruido de medición para poder mostrar el comportamiento del algoritmo de identificación. Este ejemplo ilustra claramente la robustez del algoritmo NOE con respecto al ruido, asociado con la rápida convergencia a una estructura de modelo adecuada proporcionada por el algoritmo. Se pueden obtener mejores soluciones si aumenta la cantidad de datos disponibles o el número de iteraciones.

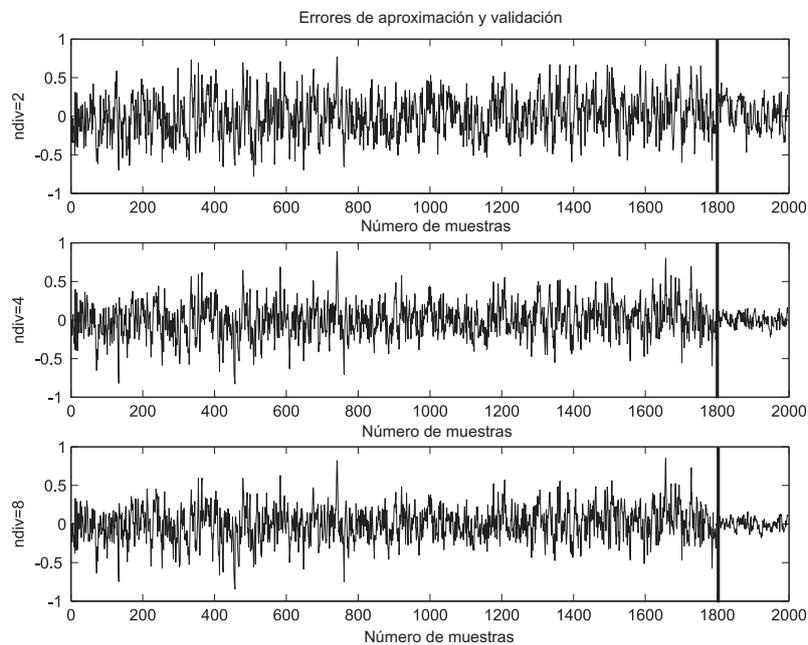


Figura 3.4: Performance del error en el Ejemplo 3.3.2. Errores de aproximación (muestras 1 a 1800) y validación (muestras 1801 a 2000)

Las superficies aproximantes  $f_{lat}$  para  $ndiv$  igual a 2, 4 y 8 puede observarse en la Figura 3.5. Como antes, estas superficies fueron evaluadas usando el conjunto de parámetros que mostró el mejor desempeño durante el proceso de aproximación, en el conjunto de validación. Podemos observar nuevamente que la aproximación mejora cuando  $ndiv$  aumenta. Al igual que en el ejemplo previo, la superficie con  $ndiv = 2$  se aproxima a una superficie plana.

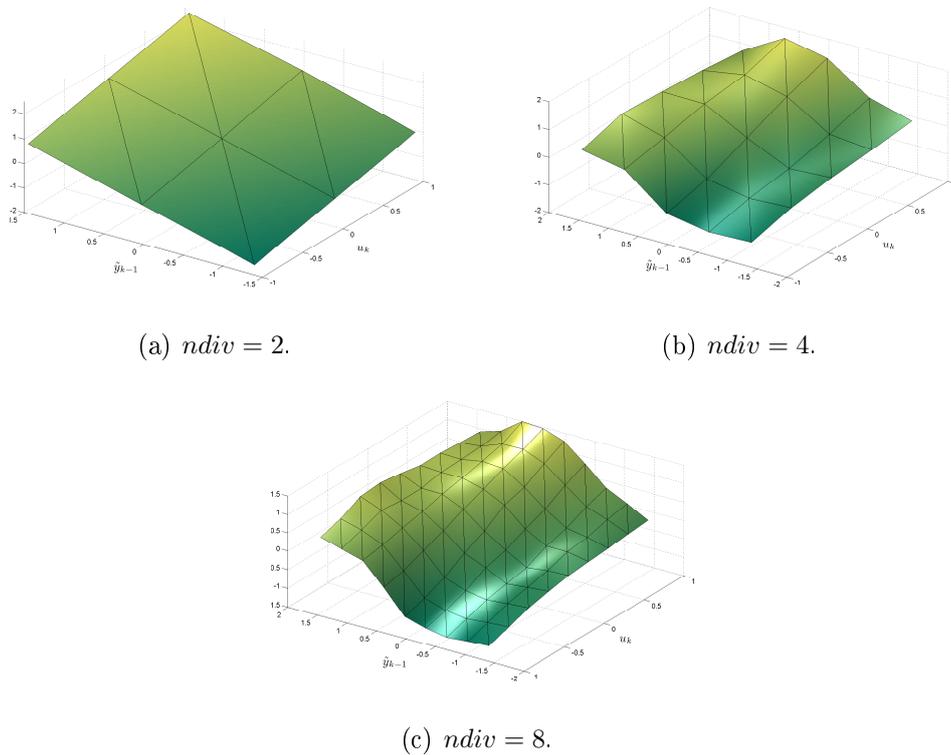


Figura 3.5:  $f_{lat}$  resultante del Ejemplo 3.3.2 con  $ndiv = 2, 4$  y  $8$

### 3.3.3. Ejemplo 3

Para este ejemplo consideramos el siguiente sistema

$$\begin{cases} y(k) = -0,99 y(k-1) + u(k)^3 + \gamma v(k), & \text{si } y(k) < 0 \\ y(k) = -0,01 y(k-1) + u(k)^3 + \gamma v(k), & \text{si } y(k) \geq 0 \end{cases}, \quad (3.15)$$

siendo  $\gamma$  la amplitud del ruido y

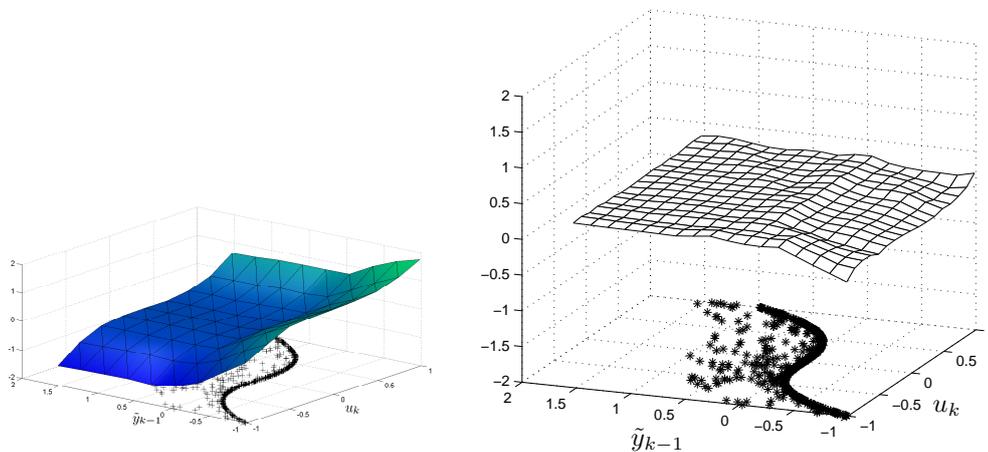
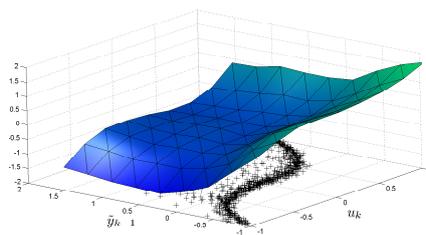
$$v(k) = 0,5 (v(k-1) + \epsilon), \quad (3.16)$$

donde  $\epsilon$  es un número aleatorio de amplitud unitaria. Entonces, el vector de ruido  $\mathbf{v}$  con componentes dadas por la ecuación (3.16) puede considerarse como ruido coloreado adicionado a la salida del sistema.

Hemos tomado  $\gamma = 0$  y  $\gamma = 0,2$  para los casos sin ruido y con ruido, respectivamente. La señal de entrada  $\mathbf{u}$  oscila entre  $-1$  y  $1$  y tiene una longitud de  $600$  muestras para el primer caso y  $1000$  para el segundo caso.

La superficie de validación  $f_{lat}$  correspondiente a  $ndiv = 8$  sobre  $(u(k), \tilde{y}(k-1))$ , junto con este conjunto de datos, puede observarse en las Figuras 3.6 (a) y (c) para los sistemas sin ruido y con ruido, respectivamente.

Observemos que en las Figuras 3.6 (a) y (c), el conjunto de datos utilizados para evaluar el modelo no está uniformemente distribuido en todo el espacio de entrada definido por  $\mathbf{S}$ . A pesar de que algunas regiones del espacio de entrada no tienen datos, la superficie resultante tiene una forma similar a la esperada como resultado de la ecuación (3.15) en el dominio de entrada, *i.e.*, dos formas lineales diferentes en la dirección de  $\tilde{y}(k-1)$  y una forma cúbica en la dirección de  $u(k)$ . Esta aseveración está confirmada por la superficie del error representada en la Figura 3.6 (b). Finalmente, la Figura 3.6 (c) muestra que es posible arribar a la misma conclusión cuando los datos medidos con ruido son usados en la identificación. Esta es una propiedad muy importante del algoritmo de identificación propuesto que debemos enfatizar. Es sabido que, a pesar de que un modelo de orden reducido puede no ajustar bien los datos, puede generalizar mejor que un modelo de orden alto. Este ejemplo sugiere que la aproximación paso a paso propuesta, permite generalizar propiedades de modelos de orden reducido que se mantienen en modelos de orden superior.

(a) Superficie sin ruido con  $ndiv = 8$ .(b) Superficie del error sin ruido con  $ndiv = 16$ .(c) Superficie con ruido para  $ndiv = 8$ . El conjunto de datos usados en la identificación es notado con + en las figuras (a) y (c), y con \* en la figura (b).Figura 3.6: Función  $f_{lat}$  resultante para el Ejemplo 3.3.3

### 3.3.4. Ejemplo 4

Consideremos el siguiente sistema

$$\begin{cases} y(k) = 0,9y(k-1)^3 + 5u(k)^3 + \gamma v(k) & \text{if } y(k) \leq 0 \\ y(k) = -0,9y(k-1)^3 + 5u(k)^3 + \gamma v(k) & \text{if } y(k) > 0 \end{cases}, \quad (3.17)$$

donde  $\gamma$  es la amplitud del ruido. Las componentes del vector de ruido coloreado  $\mathbf{v}$  están dadas por la ecuación (3.16).

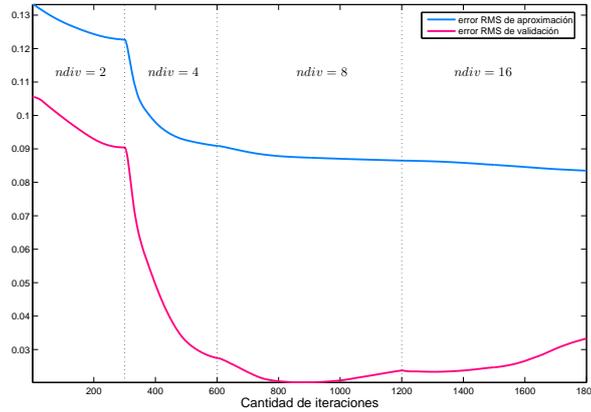
En este ejemplo elegimos  $\gamma = 0$  y  $\gamma = 0,25$ . La amplitud de  $\mathbf{u}$  oscila entre  $-0,53$  y  $0,53$ , mientras que la cantidad de muestras es de 500 para el caso sin ruido y de 1000 para el caso con ruido.

Como en el Ejemplo 3.3.3, en ambos casos la aproximación mejora a medida que aumenta el número de divisiones. Esto puede observarse en la Figura 3.7 (a) y (b) para el caso con ruido. En la Figura 3.7 (a) se muestran los errores RMS de aproximación y validación *versus* el número de iteraciones para cada número de divisiones. La tasa de crecimiento del error disminuye a medida que el número de divisiones aumenta. El algoritmo se detiene luego de 300, 300, 600 y 600 iteraciones para  $ndiv = 2$ ,  $ndiv = 4$ ,  $ndiv = 8$  y  $ndiv = 16$ , respectivamente. Por otro lado, en la Figura 3.7 (b) mostramos los errores de aproximación y validación para los modelos NOE-CLATAN con  $ndiv = 2, 4, 8$  y  $16$ . Puede apreciarse una reducción significativa de ambos errores cuando el número de divisiones aumenta a  $ndiv = 8$ . Sin embargo, con  $ndiv = 16$  el sobreajuste deteriora la propiedad de generalización del modelo, *i.e.* el error de validación crece con respecto a  $ndiv = 8$ .

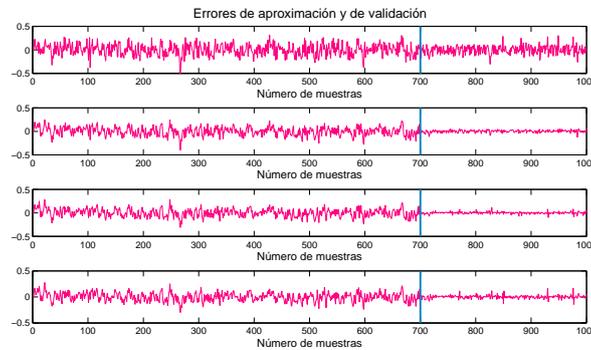
La superficie de aproximación  $f_{lat}$  correspondiente a la validación del modelo y con  $ndiv = 8$ , junto con el conjunto de datos, puede verse en las Figuras 3.8 (a) y (b) para los sistemas sin ruido y con ruido, respectivamente. Como en el ejemplo previo, la forma de la superficie evaluada se corresponde con la forma real, *i.e.* una forma cúbica con cambio de signo en la dirección de  $\tilde{y}(k-1)$  y una forma cúbica en la dirección de  $u(k)$ . Debemos remarcar que esta correspondencia ocurre sobre todo el espacio de entrada, a pesar de que el conjunto de datos usados para la aproximación está distribuido sobre un subconjunto limitado del dominio de entradas.

## 3.4. Conclusiones

En este capítulo desarrollamos un algoritmo de identificación NOE basado en el método de aproximación con funciones CLATAN. La metodología propuesta permite una evaluación gradual de un modelo NOE, a partir de un modelo lineal OE. La característica principal es su capacidad para aumentar el orden del modelo de una manera directa y sencilla, manteniendo la aproximación obtenida con el modelo anterior. Es importante remarcar que la estabilidad BIBO del algoritmo está garantizada si la aproximación lineal inicial OE es BIBO estable.



(a) Errores RMS de aproximación y validación para las funciones de aproximación CLATAN usando  $ndiv = 2, 4, 8, 16$ .



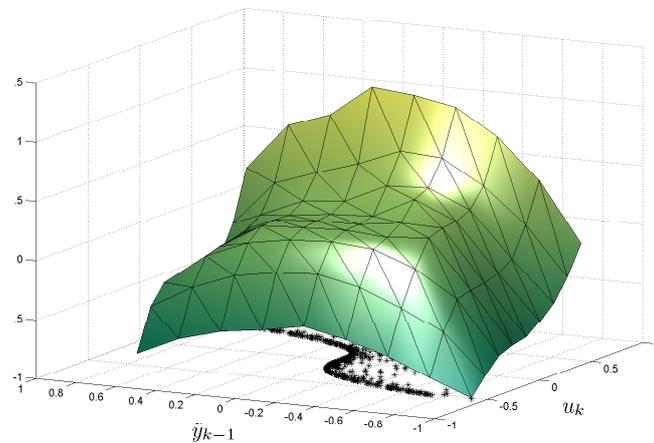
(b) Error de aproximación (muestras 1 a 700) y validación (muestras 701 a 1000).

Figura 3.7: Características del error en el Ejemplo 3.3.4 para el caso con ruido

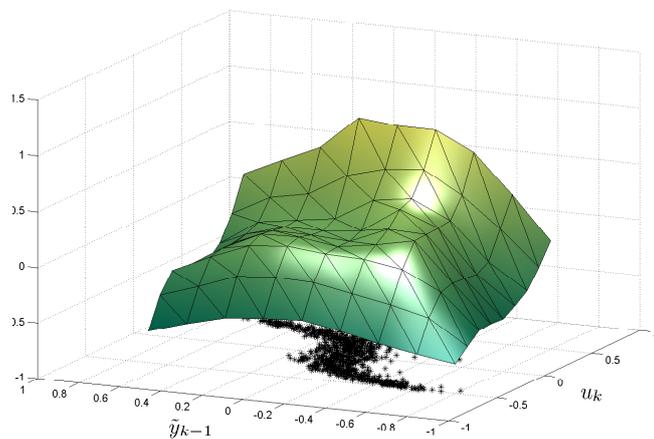
Si bien el modelo fue definido para sistemas SISO, puede ser generalizado fácilmente a sistemas MISO y MIMO.

El objetivo de este algoritmo es lograr la identificación y simulación de un modelo no lineal para aplicaciones que requieren capacidad de procesamiento rápido y gran cantidad de datos. Es sabido que un algoritmo de identificación NOE puede requerir mucho tiempo de cómputo. La plataforma de hardware CLATAN no sólo produciría una reducción drástica de dicho tiempo sino que además, la flexibilidad de modelado permitiría la identificación de una gran variedad de sistemas dinámicos. El modelado y síntesis de circuitos de sistemas dinámicos no lineales ha atraído mucho la atención en los últimos años ya que la simplicidad de integración en gran escala (VLSI) de las funciones LAT posibilita el desarrollo de estructuras de identificación orientadas a hardware [33, 45, 76, 75].

Por otra parte, el uso de bases de funciones CLATAN posibilita el abordaje sistemático



(a) Caso sin ruido.

(b) Caso con ruido para  $n_{div} = 8$  divisiones. En ambas figuras, el conjunto de datos utilizado para el proceso de identificación es indicado con +.Figura 3.8:  $f_{lat}$  para el Ejemplo 3.3.4

y riguroso de problemas de aproximación de sistemas no lineales de dimensión  $n$ , tanto en aplicaciones relativas al modelado como al control de sistemas. Esta técnica conduce a problemas de optimización de probada robustez y de eficiencia computacional, como son cuadrados mínimos y programación lineal. La reducción del tamaño de la grilla de subdivisión simplicial del dominio de las funciones CLATAN permite diseñar y obtener modelos de aproximación progresivos que se inician con uno lineal, y abordar el modelado mediante estructuras NOE, los cuales son sumamente difíciles de evaluar con las técnicas disponibles en la actualidad.

Los ejemplos presentados ilustran el potencial de la estructura de identificación propuesta. Los dos primeros muestran la capacidad para mantener la aproximación lograda cada vez que se incrementa el número de divisiones  $ndiv$  del dominio de las funciones CLATAN. Esta propiedad acelera el proceso de identificación ya que el número de parámetros aumenta gradualmente; el costo computacional del algoritmo es muy pequeño en las primeras iteraciones y aumenta sólo si se necesita una mejor aproximación.

Los dos últimos ejemplos muestran la propiedad de generalización del modelo, es decir, que la discrepancia entre el sistema real y el modelo es pequeña en el conjunto de validación.



# Capítulo 4

## Estructuras de Identificación tipo Wiener

En este capítulo describimos los métodos de aproximaciones funcionales basados en series de Volterra y de Wiener y los modelos generales orientados a bloques. En este marco, presentamos los esquemas de identificación propuestos en esta tesis, que pueden considerarse dentro de estos últimos. En su trabajo, Wiener [101] propuso aproximar un sistema no lineal continuo por una combinación polinomial de sistemas de Laguerre, utilizando para ello polinomios de Hermite. Las estructuras que aquí presentamos se diferencian de una Wiener por la manera en que dichos sistemas están relacionados. Nuestros modelos aproximan un sistema no lineal, discreto por una combinación no lineal de un conjunto finito de sistemas discretos de Laguerre o de Kautz, utilizando funciones CLATAN. Demostramos que estas estructuras de modelado son capaces de aproximar sistemas que pueden caracterizarse como no lineales, discretos, causales, TI y que poseen un cierto tipo de continuidad denominada *memoria evanescente*. Concluimos el capítulo con tres ejemplos de aplicación, cada uno de ellos con diferentes no linealidades, que muestran las capacidades de aproximación de los modelos propuestos y las ventajas de proponer sistemas de Laguerre o de Kautz para la identificación de los sistemas considerados, de acuerdo con sus características.

### 4.1. Diferentes Modelos de Aproximación E/S

En esta sección presentamos los métodos de aproximaciones funcionales basados en series de Volterra y de Wiener, los modelos orientados a bloques en serie, entre los que se encuentran los modelos de Wiener y de Hammerstein descritos en el Capítulo 2, y los modelos orientados a bloques en paralelo.

### 4.1.1. Métodos de las Series de Volterra y de Wiener

Los modelos de Hammerstein y de Wiener orientados a bloques vistos en el Capítulo 2, Sección 2.6 se pueden relacionar con los métodos de aproximaciones funcionales basados en series de Volterra y de Wiener.

Se denomina expansión funcional de un sistema a una representación E/S del mismo de la forma

$$y(k) = \text{Fun}[u(k')], \quad k' \leq k, \quad (4.1)$$

donde  $u(k)$ ,  $y(k)$  son la entrada y la salida, respectivamente, del sistema real en el tiempo  $k$  y  $\text{Fun}[\cdot]$  es la funcional que se desea identificar. Entonces, es posible evaluar la salida en cualquier instante  $k$ , conociendo la evolución de la entrada  $u(k')$ , para todo tiempo  $k' \leq k$ . La implementación de estas estrategias de aproximación en el caso lineal, lleva a representaciones como la expansión de un sistema en una base ortonormal formada por retardos unitarios (modelo FIR) o en bases de Laguerre, descritas en el Capítulo 2, ecuaciones (2.33) y (2.39). Cuando se trata de métodos de expansiones funcionales en el contexto de identificación de sistemas no lineales, se habla naturalmente de las series de Volterra y de Wiener.

El estudio de funcionales no lineales (4.1) comenzó en 1887 cuando Volterra [95] investigó funcionales analíticas e introdujo la representación

$$y(k) = \sum_{n=1}^{\infty} \int_{\Omega} \dots \int h_n(\tau_1, \tau_2, \dots, \tau_n) \prod_{i=1}^n u(k - \tau_i) d\tau_i \quad (4.2)$$

que luego se conoció como *serie de Volterra*. Las funciones  $h_n(\tau_1, \tau_2, \dots, \tau_n)$  en las variables  $\tau_i$  de la ecuación (4.2) se denominan núcleos de Volterra válidos en el dominio  $\Omega$  [2]. Estos núcleos están acotados en cada  $\tau_j$ , son funciones simétricas respecto de su argumento y, para sistemas causales, se verifica que  $h_n(\tau_1, \tau_2, \dots, \tau_n) = 0$  para cualquier  $\tau_j < 0$ . La identificación por medio de series de Volterra requiere entonces la estimación de dichos núcleos. Dado que la ecuación (4.2) involucra una serie, para realizar la identificación es necesario truncarla.

Fréchet [36] consideró esta representación y generalizó el teorema de aproximación de Weierstrass probando que cualquier funcional continua puede ser representada por medio de una serie de Volterra que converge uniformemente en cualquier conjunto compacto de funciones continuas.

La serie de Volterra discreta está dada por

$$y(k) = \sum_{m=0}^{m_s} \sum_{i_1=0}^{lm-1} \dots \sum_{i_m=0}^{lm-1} h_m(i_1, i_2, \dots, i_m) \prod_{j=1}^m u(k - i_j), \quad (4.3)$$

donde  $m_s$  es el orden de la serie que tiene memoria finita  $lm$  y  $h_m$  es un núcleo simétrico. Esta serie puede aproximar uniformemente cualquier sistema discreto no lineal, con memoria

finita, causal y TI, sobre un conjunto de entradas uniformemente acotadas, siempre que se considere un orden  $m_s$  suficientemente alto. Precisamente éste es uno de los inconvenientes que presenta la aplicación práctica del método, aún para núcleos de baja dimensión. El otro inconveniente es la dificultad para identificar el núcleo  $h_m$ , lo que ha restringido las aplicaciones a expansiones con núcleo de dimensión menor o igual a 2.

Los trabajos de Volterra fueron continuados por Wiener en la década de 1950. Wiener fue uno de los primeros autores en considerar la identificación de sistemas no lineales. Dado un sistema no lineal continuo, causal y TI con entrada  $u(k)$  y salida  $y(k)$  en el instante  $k$ , la funcional utilizada por Wiener para expandir al sistema se puede expresar como

$$y(k) = \sum_{n=1}^{\infty} \sum_{m_0=0}^{\infty} \cdots \sum_{m_n=0}^{\infty} c_{m_0 m_1 \dots m_n} p h_{m_0}(v v_0(k)) p h_{m_1}(v v_1(k)) \dots p h_{m_n}(v v_n(k)),$$

donde  $c_{m_0 m_1 \dots m_n}$  representa a los coeficientes de la expansión y  $v v_j$  son variables auxiliares obtenidas por combinación de los polinomios de Hermite  $p h_{m_j}(\cdot)$ . Estas variables representan la convolución de la señal de entrada  $u(k)$  con los sistemas de Laguerre  $l_j(k)$  de orden  $p$

$$v v_p(k) = \int_0^{\infty} l_p(\tau) u(k - \tau) d\tau, \quad j = 0, 1, \dots$$

El polinomio de Hermite  $p h_{m_j}(\cdot)$  parcialmente normalizado está definido por

$$p h_{m_j}(u) = (-1)^{m_j} 2^{m_j/2} \frac{e^{u^2}}{\sqrt{m_j!}} \frac{d^{m_j}}{du^{m_j}} \{e^{-u^2}\}.$$

El método de identificación propuesto por Wiener puede ser visualizado como una cascada de tres bloques diferentes. El primero de ellos representa un sistema SIMO (Simple-Input-Multiple-Output) por medio de filtros de Laguerre; el segundo de los bloques es sintetizado por medio de una transformación no lineal sin memoria, representada por polinomios de Hermite y, finalmente, los sumadores y amplificadores que representan los coeficientes de la expansión. En la Figura 4.1 puede verse la versión discretizada de esta representación. Aquí  $L_0, L_1, \dots, L_m$  representan los filtros de Laguerre discretizados, dados por

$$\begin{aligned} L_0[u(z)] &= K \frac{1}{z-a} u(z), \\ L_i[u(z)] &= \frac{(1-az)}{(z-a)} L_{i-1}[u(z)], \quad \text{para } i \geq 1. \end{aligned}$$

Si bien Wiener no menciona en su libro el trabajo previo de Volterra, se puede demostrar que la formulación de Wiener está relacionada con una expansión del tipo de Volterra [80].

A pesar de que la formulación de Wiener es teóricamente muy elegante, es poco práctica y difícil de utilizar debido a la cantidad excesiva de coeficientes que se requieren para lograr una aproximación deseada [8]. Por ejemplo, si se utilizan  $m$  filtros de Laguerre

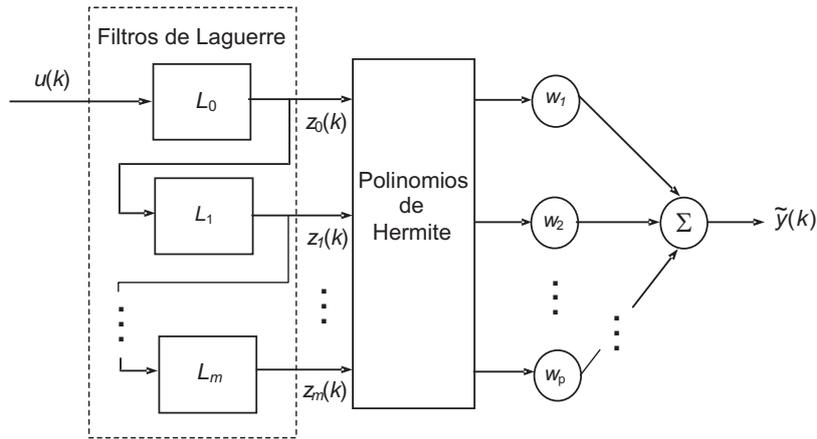


Figura 4.1: Representación del esquema de Wiener discreto

y polinomios de Hermite de grado  $n_m$ , entonces serán necesarios  $(n_m)^m$  coeficientes para caracterizar el sistema. De esta forma, si utilizáramos 10 filtros de Laguerre para un sistema con una única linealidad de segundo orden, harían falta  $2^{10}$  coeficientes. De hecho, esta complejidad numérica es el principal obstáculo de este esquema de aproximación [8] y es lo que ha restringido las aplicaciones a sistemas cuadráticos. Existen otros resultados teóricos que muestran las dificultades de calcular núcleos de orden mayor o igual a tres [71]. Como consecuencia de esto, las versiones discretizadas del modelo de Wiener han tenido muy pocas aplicaciones [54].

Las expansiones de Volterra y Wiener se han limitado a realizaciones donde el orden del núcleo es reducido, debido a la gran cantidad de parámetros asociados que es necesario estimar.

#### 4.1.2. Modelos Generales Orientados a Bloques

Los sistemas representados por bloques en cascada, o modelos orientados a bloques, comenzaron a ser estudiados por varios autores en una búsqueda por reducir la alta carga computacional asociada a las aproximaciones mediante series funcionales. La Figura 4.2 muestra estos modelos que consisten en una cascada integrada por dos sistemas lineales dinámicos representados por las respuestas impulsivas  $hgc_1(k)$  y  $hgc_2(k)$ , donde  $hgc_i(k)$  es la respuesta impulsiva del modelo para la sección  $i$ -ésima, y un sistema no lineal sin memoria  $ff(\cdot)$ . Se trata de aproximar al sistema teniendo en cuenta que las señales  $t(k)$  y  $s(k)$  no son accesibles para la medición.

Los métodos para la estimación de  $hgc_i(k)$  han sido muy variados. Sin embargo, todas estas técnicas tienen en común que requieren un largo tiempo de experimentación y la adquisición de datos que en muchos casos no es posible obtener. Un método más simple de

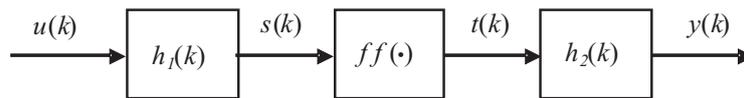


Figura 4.2: Modelo general de bloques en cascada

resolver el problema de identificación fue desarrollado por Korenberg [54], utilizando técnicas de correlación y transformadas de Fourier para identificar los valores de  $hgc_i(k)$ . Una vez estimadas estas respuestas impulsivas se pueden calcular los valores de las señales  $t(k)$  y  $s(k)$  procediéndose a graficar la función  $ff(\cdot)$ . Las desventajas del método son que es necesario graficar a la función  $ff(\cdot)$  y, de acuerdo a que esta función sea par o impar, se deben calcular correlaciones de orden superior para poder identificar completamente al modelo.

Como una forma de simplificar los cálculos, surgieron otros métodos muy utilizados que son subestructuras del modelo general. Entre ellos, podemos citar los modelos de Wiener y de Hammerstein, tomando  $hgc_2(k) = \delta^3(k)$  y  $hgc_1(k) = \delta(k)$ , respectivamente, vistos en el Capítulo 2. Los métodos de bloques en cascada tienen la ventaja que utilizan una menor cantidad de términos que los de expansiones funcionales.

### 4.1.3. Modelos Orientados a Bloques en Paralelo

Un método que mejoró el anterior fue desarrollado más tarde por Korenberg y Paarman [55] y consiste en la conexión en paralelo de varias secciones, compuestas por la cascada de un sistema lineal dinámico y un sistema no lineal estático, como se muestra en la Figura 4.3. Este método puede pensarse entonces como la unión de los dos métodos utilizados con

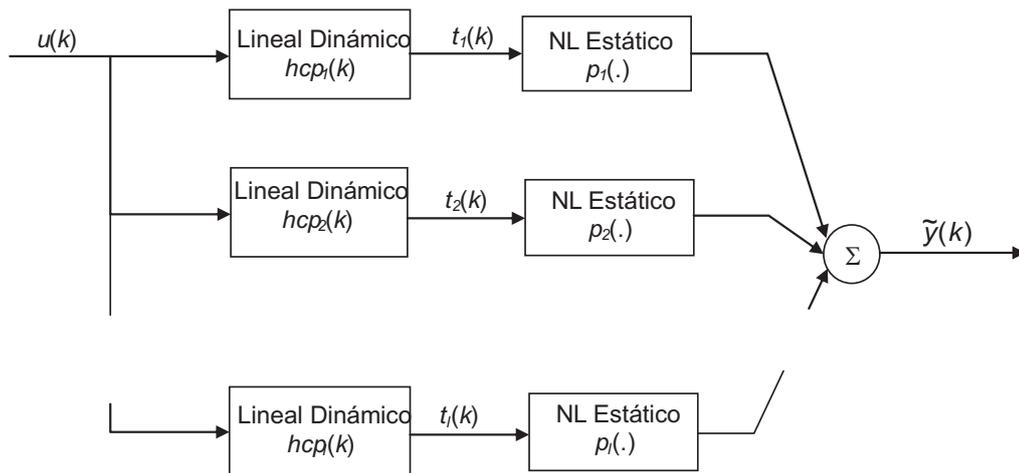


Figura 4.3: Representación del esquema de Korenberg y Paarman

propósitos de aproximación: expansión en serie de funcionales y bloques en cascada. La salida del sistema no lineal se aproxima, en primer lugar, por medio de una cascada de un sistema lineal seguido de una no linealidad estática. Luego se calcula el error entre la salida del sistema verdadero y de esta primera aproximación, ajustando el residuo con una segunda cascada. Este proceso continúa hasta que el error sea tan pequeño como se quiera y está garantizada la convergencia en media cuadrática.

Entre los diferentes esquemas propuestos en la literatura podemos citar, además del modelo de Hammerstein [88], el modelo propuesto por De Figueiredo y Chen [34], el modelo propuesto por Sentoni [83, 84], denominado LNet, y el modelo propuesto por Castro *et al.* [15].

En los artículos [8, 40, 86, 106] se puede encontrar una revisión minuciosa de los diferentes métodos a través de los años.

## 4.2. Estructuras Laguerre-CLATAN y Kautz-CLATAN

En esta sección presentamos las estructuras de aproximación Laguerre-CLATAN y Kautz-CLATAN propuestas en esta tesis, conjuntamente con una metodología para su utilización en la identificación de sistemas no lineales, discretos. Estas estructuras se encuadran dentro de la clase de modelos orientados a bloques que describimos anteriormente y están compuestas por un conjunto finito de sistemas discretos de Laguerre o de Kautz, relacionados de manera no lineal por funciones CLATAN. Con el fin de estimar los parámetros de las funciones CLATAN, formulamos el problema en el contexto de la *Teoría de estimación con conjuntos de membresía* o *Teoría de estimación SM* (por sus siglas en inglés, Set Membership Theory). La Subsección 4.3.1 comprende los principales tópicos de esta teoría; para una revisión sobre este tema puede consultarse, por ejemplo, [12, 61, 66].

Sean  $\mathbf{u}$ ,  $\mathbf{y}$  los vectores E/S correspondientes a un sistema dinámico no lineal SISO. Un sistema no lineal BIBO estable puede representarse por la siguiente estructura de modelo

$$\tilde{y}(k+1) = f(P_1[u(k)], P_2[u(k)], \dots), \quad (4.4)$$

donde  $\tilde{y}(k+1)$  es el valor estimado de la salida del sistema, que corresponde a la entrada  $u(k)$  en el instante  $k$ ,  $f: \ell^\infty \rightarrow \mathbb{R}$  es una transformación estática no lineal y  $P_j[u(k)]$  es la salida de la  $j$ -ésima componente del sistema lineal elegido [80].

En esta tesis proponemos un sistema finito de bases discretas de Laguerre, o de Kautz con dos parámetros, para representar la dinámica lineal y una transformación no lineal estática representada usando funciones CLATAN [1].

Formalmente, el modelo propuesto está dado por

$$\tilde{y}(k+1) = f_{lat}(z_1[u(k)], z_2[u(k)], \dots, z_n[u(k)]), \quad (4.5)$$

donde  $z_i[u(k)]$ ,  $i = 1, \dots, n$  son las salidas de las funciones de transferencia discretizadas y  $f_{lat}$  representa una función en  $PWL_H[\mathbf{S}]$ . Podemos visualizar esta estructura en la Figura 4.4.

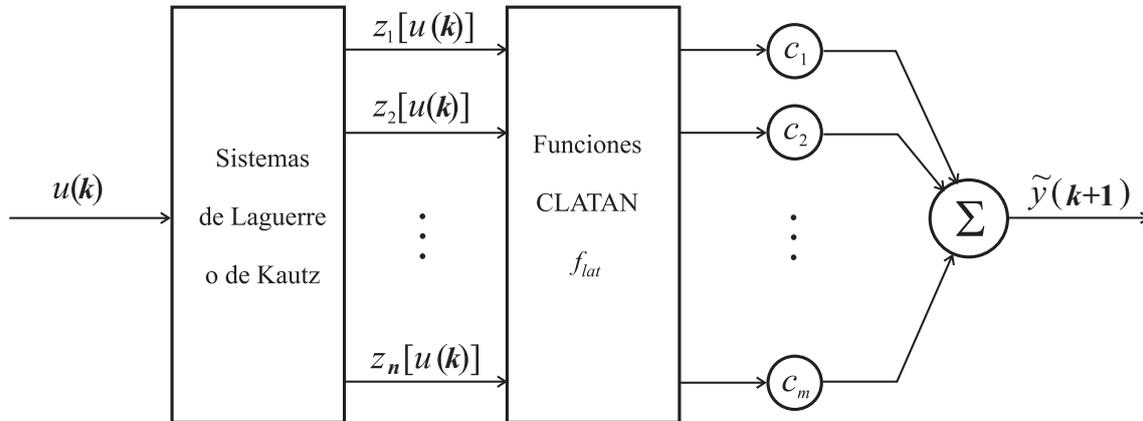


Figura 4.4: Estructura Laguerre-CLATAN y Kautz-CLATAN

Una pregunta que surge es

*¿Por qué elegir un modelo tipo Wiener para problemas de identificación?*

Una ventaja importante de este tipo de estructura es que el problema de identificación del sistema se reduce a dos procedimientos concretos:

- (a) la identificación de un sistema lineal,
- (b) la interpolación de una función no lineal estática utilizando alguna estructura adecuada.

Para lograr el paso (a) utilizamos un número finito de elementos denominado el *orden* del modelo.

El paso (b) requiere la elección de una representación adecuada que debe satisfacer varias restricciones. En primer lugar, debe ser capaz de representar uniformemente una función continua y, en segundo lugar, debe ser una representación eficiente desde el punto de vista de la cantidad de parámetros mínima y necesaria para aproximar la función dada. Otro factor relevante es la necesidad de contar con un algoritmo adecuado que permita ajustar los parámetros.

En esta tesis proponemos la identificación del bloque lineal dinámico mediante funciones de Laguerre o de Kautz, según las características conocidas *a priori* del proceso real. En particular, utilizamos funciones CLATAN para la representación del bloque no lineal estático. El modelo que resulta es no lineal, fácilmente identificable a partir de los datos E/S y

puede predecir la respuesta futura del sistema sobre horizontes de tiempo arbitrariamente grandes. Estas características hacen que esta estructura de aproximación sea particularmente adecuada para control predictivo no lineal.

En las próximas subsecciones describimos las dos propuestas. Luego, desarrollamos los métodos para la aproximación lineal dinámica y presentamos el esquema de aproximación no lineal que utilizamos para ambas estructuras de identificación, conjuntamente con la teoría de estimación con conjuntos de membresía. Esta teoría nos permitirá ajustar los parámetros del bloque no lineal estático suponiendo sólo que el ruido es desconocido, pero acotado.

### 4.2.1. Estructura Laguerre-CLATAN

En esta estructura proponemos utilizar filtros Laguerre para identificar el bloque lineal dinámico y funciones CLATAN para la parte no lineal estática representada por la función  $f$  como se describió en la Sección 2.7 del Capítulo 2. Otros modelos propuestos en la literatura en los cuales la transformación no lineal se representa por medio de redes neuronales fueron presentados en [8, 9, 22, 71, 83, 84, 93].

Con respecto a la entrada que debemos proveer al sistema para identificar la parte no lineal estática mediante funciones CLATAN, ésta debe ser lo suficientemente rica para permitir poner de manifiesto la no linealidad del sistema. Por ejemplo, algunas de las que comúnmente se utilizan son las funciones *chirp*, las funciones aleatorias con probabilidad uniforme y también las funciones equiprobabilidad. Estas últimas se caracterizan porque la probabilidad de cambio al final de cada intervalo de muestreo y la amplitud varían, según la especificación del usuario, de acuerdo a una distribución normal.

Una vez determinados los parámetros de la función  $f$  por medio del algoritmo elegido, tenemos identificada la parte no lineal estática del sistema. Esto, junto con los parámetros de la parte lineal del sistema, que son los polos de los filtros aplicados, conforman la estructura de identificación propuesta.

### 4.2.2. Estructura Kautz-CLATAN

Para esta estructura proponemos utilizar sistemas de Kautz para identificar el bloque lineal dinámico y funciones CLATAN para la identificación de la parte no lineal estática. Otras estructuras que mantienen los filtros de Laguerre como en la formulación original de Wiener y utilizan funciones CLAT de alto orden (pero no funciones CLAT ortonormales) han sido propuestas en [8, 15, 90]. Ahora bien, en [8] y [90] no fueron formulados esquemas numéricos o esquemas prácticos para hallar los coeficientes de la aproximación por funciones LAT de alto orden.

Las entradas al sistema no lineal estático son las que provienen del sistema lineal dinámico. En el caso de esta estructura no es tan fácil decidir qué entrada debemos utilizar de

manera tal que se pueda garantizar la identificación del sistema no lineal. En este caso debemos tener en cuenta que las salidas del sistema lineal cubran de manera efectiva el dominio de las funciones CLAT para que no queden simplices sin identificar pues esto lleva a tener problemas en la generalización del modelo. En la Sección 4.5 proponemos un algoritmo para solucionar este inconveniente. Una vez hallados los parámetros de las funciones CLATAN asociados a la parte no lineal estática, éstos conforman, junto con los parámetros de la parte lineal dinámica, el sistema identificado.

Las ventajas de utilizar funciones CLATAN son las siguientes:

- (i) Es posible aproximar uniformemente cualquier función no lineal, Lipschitz continua, definida sobre un conjunto compacto. Mas aún, es posible determinar *a priori* la cantidad de términos necesarios para lograr un error deseado.
- (ii) El cálculo de los coeficientes de la aproximación es directo.
- (iii) Es posible detectar los simplices que contribuyen a la aproximación.
- (iv) Es posible detectar los simplices que no están identificados.
- (v) Es posible obtener una identificación robusta mediante la identificación de dos sistemas: uno que acote por encima al sistema verdadero y otro que lo acote por debajo, como se describe en el Capítulo 5.

### 4.2.3. Metodología de Modelado

En general, la construcción de un modelo a partir de datos del sistema real involucra tres pasos básicos: la recolección de los datos, la selección de la clase de modelos o *tipo de conjunto de modelos* [58] candidatos, *i.e.* el tipo de grupo de modelos que comparten una misma estructura, y las reglas con las cuales es posible determinar un modelo de dicho conjunto, utilizando los datos recolectados.

En particular, si consideramos la clase de modelos Laguerre-CLATAN y Kautz-CLATAN, cada estructura de identificación Laguerre-CLATAN queda definida por un conjunto finito de filtros de Laguerre generados por un polo real que identifica el bloque lineal, y una función CLATAN  $f_{lat}$  que identifica la parte no lineal estática. Similarmente, un esquema de identificación Kautz-CLATAN queda determinado por un número finito de filtros de Kautz generados por un par de polos complejos conjugados y una función CLATAN  $f_{lat}$ . En ambos casos, el número de parámetros de la transformación  $f_{lat}$  depende del número de filtros seleccionados para modelar la parte lineal dinámica y de la cantidad de divisiones del dominio de dicha función. La etapa que sucede a la selección de una clase de modelos es la determinación de un modelo que represente al sistema, por medio de algún índice capaz de medir su calidad en base al error de estimación y a la cantidad de parámetros incluidos en él. La calidad de un modelo puede definirse en base al error de predicción o al error de

estimación [58]. Este proceso comprende la identificación de un bloque lineal dinámico y de uno no lineal estático.

#### 4.2.4. Identificación del Bloque Lineal Dinámico

La identificación del bloque lineal dinámico del sistema bajo estudio comprende la estimación del polo real de los sistemas de Laguerre, o de los dos polos complejos conjugados de los sistemas de Kautz que, en ambos casos, representan los parámetros del modelo lineal dinámico elegido y la evaluación del número de funciones base de Laguerre o de Kautz *i.e.* el orden del modelo lineal dinámico.

Tomemos como bloques de la parte lineal dinámica, las funciones de transferencia de Laguerre o de Kautz definidas en el Capítulo 2, Sección 2.4.

En términos de la transformada  $\mathcal{Z}$ , el conjunto de filtros de Laguerre puede ser expresado por

$$L_j(z, a) = z^{-nd} \frac{\sqrt{(1-a^2)T}}{z-a} \left( \frac{1-az}{z-a} \right)^{j-1}, \quad j = 1, 2, \dots \quad (4.6)$$

donde el parámetro  $a \in \mathbb{R}$  es el polo generador, con  $|a| < 1$ .

Análogamente, los sistemas de Kautz se definen como

$$K_{2j}(z, b, c) = \frac{\sqrt{(1-c^2)(1-b^2)}}{z^2 + b(c-1)z - c} \left( \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{j-1}$$

$$K_{2j-1}(z, b, c) = \frac{(z-b)\sqrt{(1-c^2)}}{z^2 + b(c-1)z - c} \left( \frac{-cz^2 + b(c-1)z + 1}{z^2 + b(c-1)z - c} \right)^{j-1}, \quad (4.7)$$

donde  $b = (\beta + \bar{\beta})/(1 + \beta\bar{\beta})$ ,  $c = -\beta\bar{\beta}$  son constantes reales tales que

$$|b| < 1, \quad |c| < 1, \quad j = 1, 2, \dots$$

y  $\beta, \bar{\beta} \in \mathbb{C}$  son los polos de Kautz en la región  $\|z\| < 1$ .

Ahora bien, las preguntas que surgen inmediatamente de esta propuesta son las siguientes

*¿Cómo elegimos los polos de los filtros de Laguerre o de Kautz, que son los parámetros del modelo lineal dinámico?*

*¿Qué orden elegimos para el bloque lineal, *i.e.* cuántos filtros utilizamos?*

*¿Cuáles son las ventajas de proponer sistemas Laguerre o Kautz para la identificación de la parte lineal dinámica?*

### Polos del Sistema Elegido

En la formulación original de Wiener, un problema relacionado con la identificación de la parte lineal dinámica es la elección del polo de los filtros de Laguerre. En general, como el sistema es no lineal, el valor óptimo del polo no sólo depende de la amplitud de la señal sino también del orden elegido para el modelo. Debido a estas razones, la identificación del polo no resulta un problema trivial (ver, por ejemplo, [15]). Por otra parte, es necesario destacar que la formulación original de Wiener no depende del valor específico del polo, pero una elección adecuada del mismo puede reducir el orden del modelo en forma significativa.

Con el esquema que proponemos no tenemos este problema dado que utilizamos la salida del sistema para determinar las funciones de transferencia necesarias para la representación lineal, tal como se describirá en la próxima sección. Por lo tanto, la contribución frecuencial que efectúan las diferentes funciones de transferencia ayudan a cubrir todo el espectro de frecuencias de la salida del sistema no lineal. Es decir, esta es la forma en que incorporamos el conocimiento *a priori* del sistema que estamos considerando.

Ahora bien, para realizar la selección de las funciones de transferencia es necesario conocer la ventana tiempo-frecuencia de los filtros y, por lo tanto, debemos determinar los parámetros de los mismos. Para ello, excitamos el sistema, por ejemplo con una entrada tipo escalón de corta duración y con una amplitud adecuada según el sistema que estemos considerando. Observemos que no podemos generalizar acerca del tipo de señal a tomar ya que depende del sistema que debemos identificar. Hallamos entonces una aproximación ARX de primer orden de la salida obtenida, si utilizamos los sistemas de Laguerre, o de segundo orden en el caso que utilizemos los sistemas de Kautz y calculamos los polos de este modelo, los cuales utilizamos como parámetros del modelo lineal.

El polo generador de los sistemas de Laguerre también puede ser estimado de acuerdo a

$$a = 1 - \frac{T}{t_d}, \quad (4.8)$$

donde  $t_d$  es la constante de tiempo dominante del sistema y  $T$  representa el período de muestreo. Para sistemas discretos  $T = 1$ . Un método que puede utilizarse por su sencillez para estimar la constante  $t_d$  consiste en tomar muestras de la respuesta al escalón del sistema y luego aproximar dicha respuesta en algún sentido óptimo mediante el término pasabajos  $L_0(q^{-1})$  de los sistemas de Laguerre.

Una vez estimados los polos, el conjunto de sistemas elegidos queda totalmente definido. Resta ahora evaluar el número de dichos sistemas.

### Orden de la Base Elegida

El número de funciones de las bases Laguerre o Kautz y el número de entradas al bloque no lineal estático es el mismo. Para obtener este número utilizamos un método similar al

sugerido en [42] para determinar la dimensión de una función multivariable en modelos no lineales E/S, para lo cual se tiene en cuenta el coeficiente de Lipschitz. Este método fue presentado en [84] y [92] para el caso Laguerre y Kautz, respectivamente, y se basa en el principio de que toda función continua que toma valores en una región cerrada es Lipschitz acotada.

Debido a que los modelos Laguerre-CLATAN y Kautz-CLATAN realizan un mapeo E/S con la combinación de dos estructuras, el análisis se centra en determinar cómo cambian las características de continuidad de estas estructuras con la variación del número de sistemas de Laguerre o de Kautz.

En el Apéndice B describimos el algoritmo utilizado para seleccionar el orden del modelo lineal utilizando el coeficiente de Lipschitz.

### Filtros utilizados en la identificación de la parte lineal dinámica del sistema

Nosotros proponemos la elección de cualquiera de los dos sistemas, Laguerre o Kautz vistos en el Capítulo 2; cuál de ellos utilizamos dependerá del sistema a identificar.

Aunque el uso de filtros de Laguerre en la identificación de sistemas no lineales es muy común, un inconveniente que presentan es que los sistemas sub-amortiguados son difíciles de aproximar utilizando un número razonable de términos de esta base. Entonces, en esos casos es preferible utilizar las bases ortonormales de funciones de Kautz [105, 44, 56]. Estos filtros permiten aproximar señales con un comportamiento fuertemente oscilatorio [97, 98]. La ventaja de usar bases de Kautz para los sistemas con resonancia es que se necesita una cantidad menor de filtros, *i.e.* un modelo de orden reducido, para la aproximación de la parte lineal.

Tenemos entonces identificada la parte lineal dinámica de nuestro sistema no lineal. Debemos proceder ahora a identificar la parte no lineal estática del sistema, para lo cual proponemos la utilización de funciones lineales a tramos de alto nivel y la estimación de los parámetros de dichas funciones utilizando teoría de estimación SM. Este procedimiento de estimación tiene una propiedad muy interesante y es que los parámetros se calculan sólo con la suposición que el ruido es desconocido, pero acotado. Con ello quedan determinadas dos estructuras de identificación que llamamos Laguerre-CLATAN y Kautz-CLATAN, respectivamente, según el conjunto de funciones base que hayamos utilizado para identificar el bloque lineal dinámico.

## 4.3. Identificación de la Parte no Lineal Estática

Comenzamos la sección describiendo algunos tópicos de la teoría de estimación SM que serán necesarios para estimar los parámetros de las funciones CLATAN que identifican el bloque no lineal estático de las dos clases de modelos presentadas.

### 4.3.1. Teoría de Estimación con Conjuntos de Membresía

La teoría con conjuntos de membresía o teoría SM nació a fines de los años sesenta y fue aplicada a problemas de estimación de los estados de sistemas dinámicos [81, 103].

En los años ochenta, esta teoría comenzó a despertar interés debido al desarrollo de la teoría de control robusto, ya que la propuesta de identificación de sistemas con conjuntos de membresía proporciona modelos que permiten hallar cotas fuertes para la incertidumbre [35, 70].

El objetivo de un problema de estimación es obtener un modelo dinámico para un sistema desconocido del que sólo disponemos de datos E/S medidos con ruido. De acuerdo a las hipótesis que se hagan sobre el ruido, se puede distinguir entre una propuesta estadística y una propuesta determinística. La diferencia principal entre la estimación clásica, estadística, y la estimación con conjuntos de membresía, determinística, es que en la primera el ruido se representa como un proceso estocástico (usualmente ruido blanco filtrado), mientras que en estimación SM el ruido se supone desconocido pero acotado o ruido UBB (por las siglas en inglés de Unknown But Bounded error), es decir, el único conocimiento acerca del ruido consiste en una cota del mismo evaluada en alguna norma dada.

En los problemas de identificación estándar, el error tiene un término en varianza debido al ruido que afecta a los datos, y un término de sesgo (bias) debido a la dinámica del sistema que no es capturada por el modelo nominal estimado. Claramente, la naturaleza de esos dos términos del error es muy diferente, el primero generalmente no está correlacionado con la señal de entrada, mientras que el segundo depende fuertemente del modelo nominal estimado y de la entrada utilizada en el proceso de identificación. En los primeros trabajos sobre identificación SM se formulan algoritmos eficientes para estimar el conjunto de modelos posibles, compatible con la hipótesis de ruido UBB, sin separar dinámica no modelada y ruido (o, equivalentemente, asumiendo que la planta que genera los datos pertenece a la familia de modelos considerada).

En estimación estadística, la incertidumbre se describe en términos de intervalos de confianza. Por el contrario, en estimación SM se halla un conjunto que contiene todas las posibles soluciones del problema que permite la evaluación de las cotas de incertidumbre asociadas al problema de estimación. Por esta razón esta propuesta se denomina SM. Además, mientras la estimación estadística trata con casos promedio, la teoría determinística considera el peor caso, que es el estimador que muestra el mejor desempeño en el peor caso considerado (modelo de peor caso o modelo de incertidumbre). El siguiente propósito es presentar una descripción general de los principales conceptos y aspectos de la teoría SM para la estimación e identificación de sistemas y un tratamiento detallado de algunos problemas específicos de interés en identificación SM.

En teoría de estimación SM, la incertidumbre se describe utilizando una cota para el ruido o error UBB adicionado a las salidas del sistema. En este contexto, el proceso de estimación del modelo que minimiza el error de identificación de peor caso es usualmente denominado *identificación con conjuntos de membresía*. En esta tesis, describimos el problema

de la estimación de los parámetros de las funciones CLATAN en el formalismo de teoría de identificación SM, asumiendo un error UBB [63, 82].

Este procedimiento de estimación tiene una propiedad muy interesante que es el cálculo de los parámetros suponiendo únicamente que el ruido es desconocido, pero acotado. La teoría de identificación SM proporciona algoritmos eficientes para la estimación del conjunto de modelos posibles, compatibles con la hipótesis de error UBB. Una ventaja importante es que los parámetros de la expresión CLATAN asociada, se pueden obtener mediante la resolución de  $2m$  problemas de programación lineal, siendo  $m$  la dimensión del espacio de los parámetros desconocidos.

### Conceptos Básicos

Un problema de estimación general puede establecerse como sigue.

*Dado un elemento desconocido  $x$ , hallar un estimador de la función  $S(x)$ , basado en una información a priori  $K$  y en datos de la función  $f(x)$  medidos con ruido.*

Por supuesto que la solución de este problema depende de las funciones  $S(\cdot)$  y  $F(\cdot)$ , de la información a priori  $K$  y, en particular, de las hipótesis que se hagan sobre el ruido  $e$ . La teoría SM trata con ruido desconocido pero acotado, que es ruido acotado en alguna norma dada.

Un problema de estimación genérico puede formularse en los siguientes espacios

$X$ : espacio de los elementos del problema, (estimar);

$Y$ : espacio de las medidas,

$Z$ : espacio de las soluciones.

Suponemos que  $X, Y$  y  $Z$  son espacios normados lineales, de dimensión  $n, m$  y  $p$  respectivamente, y  $\|\cdot\|_X, \|\cdot\|_Y$  y  $\|\cdot\|_Z$  son las normas asociadas a cada espacio.

Definimos el operador solución  $S : X \rightarrow Z$ , que es la función que asocia a cada elemento  $x \in X$  la cantidad que se quiere estimar,

$$z = S(x).$$

La estimación se basa en la información disponible sobre  $x$ , la que es fundamentalmente de dos tipos diferentes.

Una información *a priori*, usualmente representada por un subconjunto  $K \subseteq X$  al cual el elemento debe pertenecer *i.e.*,  $x \in K \subseteq X$ .

Este conjunto tiene generalmente alguna característica estructural como convexidad o simetría. Un ejemplo común es

$$K = \{x \in X : \|L(x - x_0)\|_X \leq 1\},$$

donde  $L$  es un operador lineal y  $x_0 \in X$ .

Una información *a posteriori* representada por el conocimiento de una cierta función  $F(x)$ , donde  $F : X \rightarrow Y$  se denomina *operador de información*. En general, no se conoce exactamente el valor de  $F(x)$ , porque en el proceso de medición está medido con ruido. Si asumimos ruido aditivo, las observaciones disponibles  $y \in Y$  son

$$y = F(x) + e.$$

Un aspecto crucial del problema de estimación es la hipótesis sobre el ruido. El procedimiento adoptado para resolver el problema así como las características de la solución (optimalidad, comportamiento asintótico, etc.) dependen fuertemente de esta hipótesis.

Se asume que el ruido  $e$  es UBB en norma

$$\|e\|_Y \leq \epsilon,$$

para alguna constante  $\epsilon > 0$ .

Definimos un *algoritmo de identificación o estimador* como un operador  $\Phi(\cdot)$ ,  $\Phi : Y \rightarrow Z$ , el cual da una aproximación  $\Phi(y)$  de la cantidad  $S(x)$  a ser estimada. En otras palabras, debemos determinar un algoritmo  $\Phi$ , tal que  $\Phi \simeq S(x)$ . Una representación esquemática del problema que se describe puede verse en la figura 4.5.

En general, la función  $\Phi(\cdot)$  puede ser monovaluada o multivaluada. Si  $\Phi(\cdot)$  asocia un elemento (punto) en el espacio de soluciones a cada vector en el espacio de las medidas entonces  $\Phi(\cdot)$  se dice que es un *estimador puntual*. Una función  $\Phi(\cdot)$  que determina un conjunto de elementos en el espacio  $Z$  se denomina un *conjunto estimador*. En lo que sigue  $\Phi(\cdot)$  es un estimador puntual genérico.

Debido a las hipótesis de ruido acotado, los elementos del problema de estimación pueden ser caracterizados por medio de conjuntos adecuados (propuesta SM). Se definen los siguientes conjuntos:

*Conjunto de medidas de incertidumbre*

$$MUS_y = \{\tilde{y} \in Y : \|\tilde{y} - y\|_Y \leq \epsilon\}. \quad (4.9)$$

Es el conjunto de todas las medidas cuya distancia a la observación es menor o igual a  $\epsilon$ . Este conjunto contiene toda la información exacta que pudo haber generado la medida con ruido  $y$ . Observemos que si  $x$  es un elemento desconocido a ser estimado, e  $y$  es la observación disponible, entonces  $F(x) \in MUS_y$  pues  $\|y - F(x)\|_Y = \|e\|_Y \leq \epsilon$ .

*Conjunto de estimaciones de incertidumbre*

$$EUS_\Phi = \Phi(MUS_y). \quad (4.10)$$

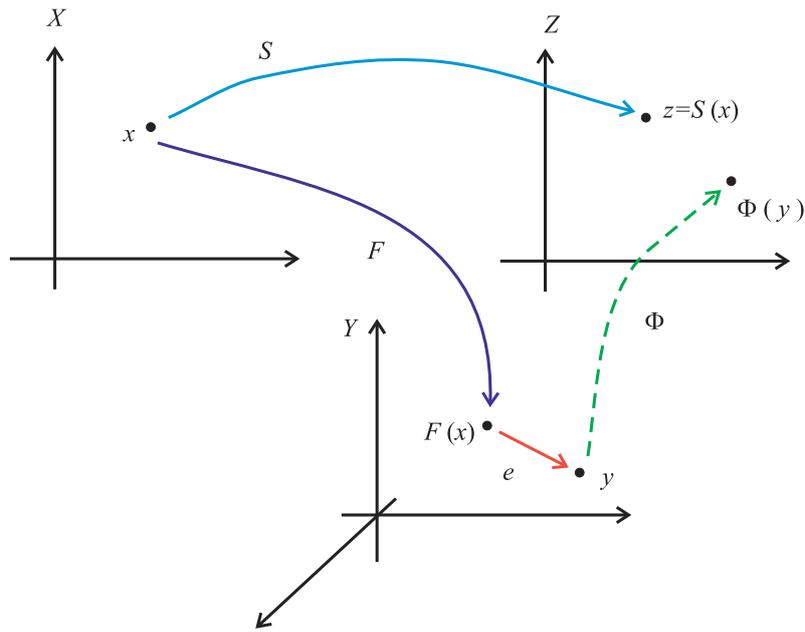


Figura 4.5: Gráfico ilustrativo de un problema de estimación genérico

Para un estimador fijo  $\Phi$ , éste es el conjunto que contiene todos los estimadores que pueden obtenerse con las medidas pertenecientes a  $MUS_y$ .

*Conjunto de parámetros posibles:*

$$FPS_y = \{x \in K : \|y - F(x)\|_Y \leq \epsilon\}. \quad (4.11)$$

Éste es el conjunto de los elementos del problema que son compatibles con toda la información disponible: la estructura del operador  $F(\cdot)$ , la hipótesis UBB del ruido, la información a priori  $K$  y las medidas observadas  $y$ . Notemos que si este conjunto es vacío, las observaciones no son consistentes con la formulación del problema, esto es, con  $F(\cdot)$ ,  $K$  y  $\epsilon$ .

*Conjunto de soluciones posibles:*

$$FSS_y = S(FPS_y). \quad (4.12)$$

Es el conjunto de soluciones admisibles, compatible con la información disponible en el problema de estimación.

En algunos problemas de estimación paramétrica y no paramétrica, el operador solución  $S(\cdot)$  es el operador identidad,  $S(x) = x$ . En ese caso, el conjunto  $FPS_y$  coincide con el conjunto  $FSS_y$ , y es llamado *conjunto posible*.

En general, los conjuntos definidos previamente pueden tener una estructura muy compleja (no convexos, no conexos, etc.). Sin embargo, en muchos problemas de interés el operador

$F$  es lineal, *i.e.*,  $F(x) = Fx$ . En este caso, si  $K = X$ , entonces  $FPS_y$  es la imagen inversa por  $F$  del conjunto  $MUS_y \cap Im(F)$ .

Si el operador  $F$  es lineal y  $K = X$ , entonces  $FPS_y = F^{-1}(MUS_y \cap Im(F))$ .

La estructura de  $FPS_y$  depende de la norma utilizada en el espacio de las medidas  $Y$ , bajo la hipótesis UBB.

Por ejemplo, si  $F$  es lineal, la siguiente tabla describe la estructura del conjunto  $FPS_y$  según la norma definida en el espacio  $Y$ . Si  $\dim(Y) < \dim(X)$ , entonces el conjunto  $FPS_y$

$\ \cdot\ _Y$	$FPS_y$
$\ell^\infty$	politopo
$\ell^2$	elipsoide
$\ell^1$	politopo

Tabla 4.1: Estructura del conjunto  $FPS_y$  según la  $Y$ -norma, caso lineal

es no acotado. Por el contrario, si  $\dim(Y) \geq \dim(X)$ , que es la situación más frecuente en problemas de estimación, entonces  $FPS_y$  es acotado si y sólo si  $\text{rango}(f) = n$ . Esta condición se suele denominar *información suficiente* (ver [62]).

Para analizar estos conjuntos y sus propiedades se necesitan los conceptos de centro y radio de Chebyshev que definimos a continuación.

**Definición 4.3.1** Sea  $I$  un subconjunto de  $Z$ .

$$cen(I) = \arg \inf_{z \in Z} \sup_{\tilde{z} \in I} \|z - \tilde{z}\|_Z. \quad (4.13)$$

se dice el *centro de Chebyshev* de  $I$ , y

$$rad(I) = \sup_{\tilde{z} \in I} \|cen(I) - \tilde{z}\|_Z, \quad (4.14)$$

se dice el *radio de Chebyshev* de  $I$ .

Dicho de otro modo, el centro de Chebyshev es el centro de la mínima esfera, en la norma  $Z$ , que contiene a  $I$ .

En general, el  $cen(I)$  no es único y puede no pertenecer a  $I$ . Si el conjunto  $I$  tiene un centro geométrico,  $cen(I)$  coincide con él.

## Errores y Radio de Información

Sea  $\Phi$  un estimador puntual. Una medida de la calidad de la estimación de  $\Phi$  viene dada por la distancia

$$\|S(x) - \Phi(y)\|_Z,$$

que depende del elemento desconocido  $x$  y de la observación disponible  $y$ . En el contexto SM, es común evaluar la distancia anterior con respecto al peor caso del elemento del problema y/o la peor medida. Esto es, utilizamos el error medido en el peor caso.

Es posible definir los siguientes errores

*Error local en Y*

$$E_y(\Phi, \epsilon) = \sup_{x \in FPS_y} \|S(x) - \Phi(y)\|_Z. \quad (4.15)$$

Es una medida *a posteriori* de la bondad de la estimación pues se basa en las medidas observadas y en el conjunto  $FPS_y$ .

*Error local en X*

$$E_x(\Phi, \epsilon) = \sup_{y \in MUS_{f(x)}} \|S(x) - \Phi(y)\|_Z. \quad (4.16)$$

Es una medida *a priori* de la estimación pues depende de los elementos del problema y no se basa en las observaciones.

*Error global*

$$E(\Phi, \epsilon) = \sup_{x \in X} E_x(\Phi, \epsilon) = \sup_{y \in Y_0} E_y(\Phi, \epsilon), \quad (4.17)$$

donde  $Y_0 = \{y \in Y : FPS_y \neq \emptyset\}$ .

En lo que sigue notamos el conjunto de sistemas posibles como  $FSS_y^m$  y los errores como  $E(\Phi, \epsilon, m)$ , donde  $m$  es el número de observaciones disponibles, porque es importante enfatizar la dependencia de la cantidad de observaciones disponibles.

El *error global mínimo* se define como el *radio de información*

$$R(\epsilon) = \inf_{\Phi} E(\Phi, \epsilon). \quad (4.18)$$

El nombre de radio se debe a que  $R(\epsilon)$  puede ser calculado de la siguiente manera

### Proposición 4.3.2

$$R(\epsilon) = \sup_{y \in Y_0} \text{rad}(FSS_y),$$

donde  $\text{rad}$  es el radio de Chebyshev.

**Demostración:** Ver [67].

□

### Propiedades del Algoritmo de Estimación

La optimalidad de un algoritmo depende del tipo de error que se minimiza.

**Definición 4.3.3** Un algoritmo  $\Phi^*$  es *X-localmente optimal* si

$$E_x(\Phi^*, \varepsilon) \leq E_x(\Phi, \varepsilon), \quad \forall x \in X, \forall \Phi.$$

**Definición 4.3.4** Un algoritmo  $\Phi^*$  es *Y-localmente optimal* si

$$E_y(\Phi^*, \varepsilon) \leq E_y(\Phi, \varepsilon), \quad \forall y \in Y_0, \forall \Phi.$$

**Definición 4.3.5** Un algoritmo  $\Phi^*$  es *globalmente optimal* si

$$E(\Phi^*, \varepsilon) \leq E(\Phi, \varepsilon), \quad \forall \Phi.$$

De esta definición y de (4.18) sigue inmediatamente que para un algoritmo globalmente optimal, el error global es igual al radio de información, es decir

$$E(\Phi^*, \varepsilon) = R(\varepsilon).$$

Observemos la diferencia entre los conceptos de optimalidad local en las Definiciones 4.3.3 y 4.3.4. En un problema de estimación donde se dispone de las medidas  $y(k)$ , el mejor algoritmo es el que minimiza el error de estimación con respecto al peor elemento  $x$  compatible con las observaciones, para todos los posibles valores de  $y(k)$  (*Y-optimalidad local*).

Por el contrario, si no se dispone de las medidas  $y(k)$ , entonces un algoritmo *X-localmente optimal* da la mejor estimación con respecto al peor valor  $y(k)$ , para todos los elementos desconocidos  $x$  en  $X$ .

Debemos remarcar que la optimalidad local es más fuerte que la optimalidad global.

**Proposición 4.3.6** Si  $\Phi$  es un algoritmo *Y-localmente optimal* o *X-localmente optimal* entonces es *globalmente optimal*.

**Demostración:** Sea  $\Phi^*$  un algoritmo *Y-localmente optimal*. Si

$$\bar{y} \in \arg \sup_{y \in Y_0} E_y(\Phi^*, \varepsilon)$$

tenemos

$$E(\Phi^*, \varepsilon) = E_{\bar{y}}(\Phi^*, \varepsilon) \leq E_{\bar{y}}(\Phi, \varepsilon) \leq E(\Phi, \varepsilon), \quad \forall \Phi,$$

donde la primera desigualdad se desprende de la *Y-optimalidad local*, y la segunda de la definición de error global. Si  $\Phi^*$  es *X-localmente optimal* la demostración es similar.  $\square$

En general, es posible probar que un algoritmo globalmente optimal no es necesariamente localmente optimal.

## Conjunto de Estimadores

Como vimos en la Sección 4.3.1, en teoría SM la información se describe por medio de conjuntos; luego se puede representar el estimador como un conjunto de soluciones admisibles. Estos estimadores pueden clasificarse en *algoritmos exactos y algoritmos aproximados*.

Los algoritmos exactos describen exactamente el conjunto de soluciones posibles, *i.e.*  $\Phi(y) = FSS_y$ .

Sin embargo, en muchos casos el conjunto  $FSS_y$  tiene una estructura compleja, como por ejemplo no lineal o no convexa. Por esta razón se trabaja con algoritmos aproximados. Estos algoritmos pueden dividirse en aproximaciones internas y externas. La propuesta más común para aproximar el conjunto de soluciones posibles es utilizar elipsoides u ortotopos como regiones aproximantes.

**Definición 4.3.7** Se llama *paralelotopo* a todo politopo unitario y a todo politopo  $k$ -dimensional, con  $k > 1$ , que tiene dos hipercaras, paralelas y congruentes entre sí, ambas de dimensión  $k - 1$ . En particular, son paralelotopos los segmentos, los paralelogramos y los paralelepípedos.

**Definición 4.3.8** Un paralelotopo se llama *ortotopo* si dos aristas adyacentes, situadas en una misma cara, son ortogonales. Son ejemplos de ortotopos los rectángulos planos y los prismas rectangulares, llamados también cajas (boxes).

Las definiciones anteriores implican que todo politopo convexo, que es la cápsula convexa de un hipercubo  $m$ -dimensional, es regular y al mismo tiempo un ortotopo.

Si  $F$  es lineal, utilizando la norma  $\|\cdot\|_1$  o la norma  $\|\cdot\|_\infty$  en el espacio de las medidas  $Y$ , se sabe que el conjunto  $FSS_y$  es un politopo, ver Tabla (4.1).

La descripción de un politopo genérico puede ser computacionalmente compleja, debido al número de vértices y de caras. Con la aproximación recursiva por medio de paralelotopos, se logra un balance entre calidad de la aproximación y esfuerzo computacional requerido (ver [25, 94]).

Entonces el conjunto  $FPS_y \in \mathbb{R}^m$  puede ser aproximado por un paralelepípedo exterior de mínimo volumen o MOB (por sus siglas en inglés, Minimal volume Outer Box) de caras paralelas a los ejes coordenados. Este paralelepípedo se construye hallando el máximo rango de posibles variaciones de los valores  $x_j$  de los parámetros  $\mathbf{x} = (x_1, \dots, x_m) \in FPS_y$ , dado por los *intervalos de incertidumbre*  $VUI_j$  (por sus siglas en inglés, Values Uncertainty Intervals) definidos como

$$VUI_j = [\inf x_j, \sup x_j], \quad j = 1, 2, \dots, m. \quad (4.19)$$

Para cada  $j$ , la medida del conjunto  $VUI_j$  da la longitud del lado del MOB que contiene a  $FPS_y$ , a lo largo del correspondiente  $j$ -ésimo eje coordenado.

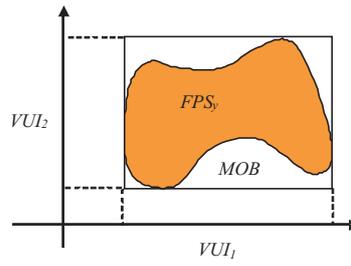


Figura 4.6: Aproximación exterior del conjunto  $FSS_y$  por medio de ortotopos

### Estimador Puntual

La teoría SM permite caracterizar el conjunto de todas las posibles soluciones de un problema de estimación; entonces la definición de un estimador consiste esencialmente en la selección de un elemento representativo dentro de un conjunto. Una posible elección es tomar el centro de Chebyshev del conjunto.

**Definición 4.3.9** Un algoritmo central  $\Phi_c$  se define como

$$\Phi_c(y) = cen(FSS_y).$$

De esta definición se obtiene que

$$E_y(\Phi_c, \varepsilon) = rad(FSS_y).$$

Las características del algoritmo central dependen, obviamente, de la estructura de  $FSS_y$ .

**Teorema 4.3.10** Si  $\|\cdot\|_Z = \ell^\infty$  y  $F, S$  son operadores lineales entonces el algoritmo central se define como:

$$\Phi_{c,j}(y) = cen_j(FSS_y) = \frac{(z_j^M + z_j^m)}{2}, \quad j = 1, \dots, m, \quad (4.20)$$

donde

$$\begin{cases} z_j^m &= \inf_{z \in FSS_y} z_j, \\ z_j^M &= \sup_{z \in FSS_y} z_j, \quad j = 1, \dots, m. \end{cases} \quad (4.21)$$

El  $Y$ -error local es

$$rad_{\mathcal{M}}(FSS_y) = \max_{1 \leq j \leq m} \frac{(z_j^M - z_j^m)}{2}, \quad (4.22)$$

**Demostración:** Ver [65].

□

### Ejemplo de un Procedimiento de Identificación SM

En esta sección presentamos un ejemplo de aplicación de las técnicas de identificación SM. En primer lugar damos una descripción del modelo físico a ser identificado. Utilizamos un conjunto estimador MOB y un estimador puntual (algoritmo central) para la identificación SM.

El proceso a ser identificado es un motor DC, donde se asume que el voltaje del motor es la variable de entrada y la velocidad angular es la variable de salida. El proceso está afectado por dos no linealidades. La primera es una saturación que afecta a la entrada en el intervalo  $[-5, 5]$  y la segunda consiste en un escalón en la salida debido a la presencia de fricción.

La clase de modelo utilizado para identificar el sistema es un modelo ARX dado por

$$\tilde{y}(k) = \sum_{i=1}^{n_a} a_i y(k-i) + \sum_{i=1}^{n_b} b_i u(k-i) + \nu(k) \quad (4.23)$$

donde  $u(k)$  es la entrada,  $\tilde{y}(k)$  es la salida y  $\nu(k)$  es el ruido. El proceso de identificación consiste en estimar los parámetros  $a_i$  y  $b_i$  sobre la base del conocimiento de las entradas  $[u(1), u(2), \dots, u(N)]$  y las salidas medidas  $[y(1), y(2), \dots, y(N)]$ .

Los parámetros serán estimados asumiendo que el ruido es UBB en norma  $\ell^\infty$

$$|\nu(k)| \leq \varepsilon, \quad \forall k. \quad (4.24)$$

Llamamos  $X$  al espacio de parámetros desconocidos, que tiene dimensión  $n_a + n_b$ , y cuyos elementos son de la forma

$$x = [a_1, \dots, a_{n_a}, b_1, \dots, b_{n_b}]^T.$$

El espacio de las medidas disponibles  $Y$  tiene dimensión  $N - n_a$  siendo sus elementos

$$y = [y(n_a + 1), \dots, y(N)]^T;$$

el vector de ruidos  $\nu \in Y$ , definido como  $\nu = [\nu(n_a + 1), \dots, \nu(N)]^T$ , verifica  $\|\nu\|_\infty \leq \varepsilon$ .

Como  $S(x) = x$ , el espacio de soluciones coincide con el espacio de parámetros, *i.e.*  $Z = X$ . Además, asumimos que no tenemos información *a priori* sobre los parámetros desconocidos; luego  $K = X$ . El operador de información está dado por

$$F(x) = \begin{bmatrix} \phi_{n_a} \\ \phi_{n_a+1} \\ \vdots \\ \phi_{N-1} \end{bmatrix} x,$$

siendo

$$\phi_k = [y(k), \dots, y(k - n_a + 1), u(k), \dots, u(k + 1 - n_b)]$$

es el regresor. Como  $F(x)$  es lineal,  $FPS_y$  es un politopo convexo. En lo que sigue calculamos un conjunto estimador y un estimador puntual.

Calcular el mínimo ortotopo que contiene al conjunto de soluciones posibles requiere la solución de  $2n$  problemas de programación lineal, dados por

$$\begin{cases} \sup x_i \\ \inf x_i \end{cases}, \quad i = 1, \dots, N$$

sujeto a

$$\|y - F(x)\|_\infty \leq \varepsilon,$$

esto es

$$F(x) = \begin{bmatrix} F \\ -F \end{bmatrix} x \leq \begin{bmatrix} y + \varepsilon \\ -y + \varepsilon \end{bmatrix}.$$

Para obtener un estimador puntual, y en particular el algoritmo central en la norma  $\ell^\infty$ , es suficiente calcular el centro geométrico del MOB hallado previamente.

### 4.3.2. Identificación del Bloque no Lineal utilizando Teoría SM

Formulamos ahora el procedimiento para identificar el bloque no lineal estático de los modelos propuestos basado en teoría de estimación SM, suponiendo que ya identificamos el bloque lineal dinámico.

Para ello debemos determinar una función CLATAN de la forma

$$f_{lat}(\mathbf{x}) = \mathbf{c}^T \Lambda(\mathbf{x}) = \mathbf{c}^T \left[ \Lambda^{0^T}, \Lambda^{1^T}, \dots, \Lambda^{m^T} \right]^T, \quad (4.25)$$

donde  $\mathbf{c} = \left[ \mathbf{c}^{0^T}, \mathbf{c}^{1^T}, \dots, \mathbf{c}^{m^T} \right]^T$ , y cada  $\mathbf{c}^i$  es un vector de parámetros asociado a la función vectorial con nivel de anidamiento  $i$ ,  $\Lambda^i$  (ver Capítulo 2).

Esta función aproxima la transformación estática no lineal  $g$  desconocida entre las salidas de los filtros  $\mathbf{z}(k) = [z_1(k), \dots, z_n(k)]$  y la salida del sistema  $y(k)$  relacionados por

$$y(k+1) = g(\mathbf{z}(k)) + e(k), \quad \forall k = 0, 1, \dots, N-1. \quad (4.26)$$

Para definir el dominio  $\mathbf{S}$  de la función CLATAN asumimos que

$$u \in \mathbf{K} \doteq \{u \in \ell^\infty : \|u\|_\infty \leq U, U > 0\} \subset \ell^\infty. \quad (4.27)$$

Las salidas de los filtros de Laguerre o de Kautz satisfacen

$$\|z_j[u(\cdot)]\|_\infty \leq G \cdot U, \quad \forall j = 1, \dots, n, \quad (4.28)$$

donde  $G \doteq \|z_1\|_1$ . Entonces, el dominio de la función CLATAN se puede definir como

$$\mathbf{S} \doteq \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z}\|_\infty \leq G \cdot U\}, \quad (4.29)$$

el cual es particionado usando una partición simplicial  $H$  con una grilla de medida  $\delta$ .

Asumimos que el ruido es UBB en norma  $\ell^\infty$ , *i.e.*

$$|e_k| \leq \epsilon, \quad \forall k, \quad (4.30)$$

para un  $\epsilon > 0$  dado.

Para aplicar las técnicas de identificación SM definimos el espacio  $X$  de parámetros desconocidos  $\mathbf{c} = [c_1, c_2, \dots, c_m]^T$  cuya dimensión depende del número  $n$  de filtros de Laguerre o de Kautz. En este caso,  $m = (ndiv + 1)^n$ .

Teniendo en cuenta (4.5) y (4.25), obtenemos

$$\begin{aligned} \tilde{y}(k+1) &= f_{lat}(z_1[u(k)], z_2[u(k)], \dots, z_n[u(k)]) \\ &= f_{lat}(\mathbf{z}[u(k)]) \\ &= \mathbf{c}^T \Lambda(\mathbf{z}[u(k)]). \end{aligned} \quad (4.31)$$

En consecuencia, podemos elegir el operador de información como  $F(\mathbf{c}) = \mathbf{c}^T \Lambda(\mathbf{z})$ . En este caso,  $F(\mathbf{c})$  es lineal y resulta que el conjunto  $FPS_y$  es un politopo convexo.

En lo que sigue, calculamos un conjunto estimador y el algoritmo central.

De acuerdo a (4.11) y (4.30)

$$FPS_y = \{\mathbf{c} \in X : \|\mathbf{y} - F(\mathbf{c})\|_\infty \leq \epsilon\}, \quad (4.32)$$

donde  $FPS_y \subseteq \mathbb{R}^m$ .

El MOB que contiene a  $FPS_y$  se puede calcular hallando el máximo rango de posibles variaciones de valores posibles  $c_j$ , dado por los intervalos de incertidumbre  $VUI_j$  definidos en (4.19).

Esto requiere resolver  $2m$  problemas de optimización con  $m$  variables y  $2N$  condiciones de desigualdad cuya función objetivo es

$$f_j(\mathbf{c}) = c_j, \quad j = 1, 2, \dots, m, \quad (4.33)$$

siendo  $\mathbf{c}$  el vector de parámetros desconocidos.

**Theorem 4.3.1** *La optimización de (4.33) sobre todos los valores posibles  $\mathbf{c} \in FPS_y$  es equivalente a resolver los siguientes problemas de programación lineal:*

$$\min f_j(\mathbf{c}) \quad j = 1, 2, \dots, m \quad (4.34)$$

sujeito a

$$\|\mathbf{y} - F(\mathbf{c})\|_{\infty} \leq \epsilon. \quad (4.35)$$

y

$$\max f_j(\mathbf{c}) \quad j = 1, 2, \dots, m \quad (4.36)$$

sujeito a

$$\|\mathbf{y} - F(\mathbf{c})\|_{\infty} \leq \epsilon. \quad (4.37)$$

**Demostración:** La demostración es inmediata teniendo en cuenta las ecuaciones (4.11) y (4.19).  $\square$

Para obtener el algoritmo central utilizando la norma  $\ell^{\infty}$ , es suficiente calcular el centro de Chebyshev del MOB, como se establece en el Teorema 4.3.10. El error de identificación de peor caso está dado por el radio de Chebyshev.

Si los valores de la función se conocen en los vértices de la partición simplicial, entonces el vector de parámetros  $\mathbf{c}$  se puede calcular resolviendo un sistema de ecuaciones lineales caracterizado por una matriz triangular superior, como puede verse en [48]. En nuestro caso, los datos están dados por un conjunto de valores distribuidos uniformemente, sin que tengamos conocimiento de los valores sobre los vértices de la partición.

## 4.4. Propiedades de Aproximación de las Estructuras Laguerre-CLATAN y Kautz-CLATAN

En las secciones anteriores nos hemos dedicado solamente a describir las estructuras de identificación que proponemos. Ahora bien,

*¿Qué tipos de sistemas podemos aproximar con las estructuras propuestas?*

La respuesta a esta pregunta es la siguiente.

*Con la estructura propuesta es posible aproximar uniformemente la evolución dinámica de cualquier sistema no lineal de memoria evanescente.*

Para poder precisar el significado de esta afirmación y probar los teoremas que respaldan nuestra afirmación, necesitamos introducir los sistemas de memoria evanescente y algunas definiciones previas.

### 4.4.1. Sistemas con Memoria Evanesciente

Consideremos los conjuntos

$$\mathbb{Z}^+ = \{k \in \mathbb{Z}, k > 0\}, \quad \mathbb{Z}_0^- = \{k \in \mathbb{Z}, k \leq 0\}.$$

## Consideraciones Previas

**Definición 4.4.1** Una secuencia  $u$  es causal si toma valores sólo para  $k \leq 0$ , *i.e.* si  $u \in \ell^\infty(\mathbb{Z}_0^-)$ .

La definición anterior expresa que una secuencia causal no toma valores para  $k > 0$ . Dado que estamos interesados en operadores definidos en  $\ell^\infty$ , debemos definir una extensión no causal de una secuencia causal. Una manera posible de hacerlo es decir que todos los valores futuros de la secuencia son iguales al valor actual.

Si  $u \in \ell^\infty(\mathbb{Z}_0^-)$ , definimos una extensión continua no causal  $u_e \in \ell^\infty$  de la siguiente manera

$$u_e(k) = \begin{cases} u(k), & k \leq 0 \\ u(0), & k > 0. \end{cases} \quad (4.38)$$

Es posible realizar la operación opuesta, es decir es posible truncar una sucesión no causal  $u \in \ell^\infty$  utilizando el operador de proyección  $P : \ell^\infty \rightarrow \ell^\infty(\mathbb{Z}_0^-)$  definido como

$$Pu(k) = \begin{cases} u(k), & k \leq 0 \\ 0, & k > 0. \end{cases} \quad (4.39)$$

En la Figura 4.7 podemos observar el gráfico de una sucesión causal  $u$ , su extensión no causal  $u_e$  y el operador de proyección aplicado a la sucesión no causal, *i.e.*  $Pu_e$ .

Podemos asociar una funcional  $T$  en  $\ell^\infty(\mathbb{Z}_0^-)$  a cada operador causal, invariante en el tiempo  $f$  de la siguiente manera

$$Tu = fu_e(0), u \in \ell^\infty(\mathbb{Z}^-), \quad (4.40)$$

donde  $u_e$  es la extensión no causal de  $u$  definida en (4.38).

Es fácil ver que el número  $fu_e(0)$  representa el valor actual de la salida del sistema, el cual es cero por convención. Esta consideración permite interpretar que la funcional  $T$  transforma el pasado de la entrada  $u$  en la salida presente del sistema. Entonces el operador  $f$  puede verse como una funcional que transforma al pasado de la entrada en la salida actual para cada instante de tiempo. Esto permite reconstruir al operador  $f$  partiendo de su funcional  $T$  asociada, utilizando la relación

$$fu(k) = TPq_{-k}u(k), \quad (4.41)$$

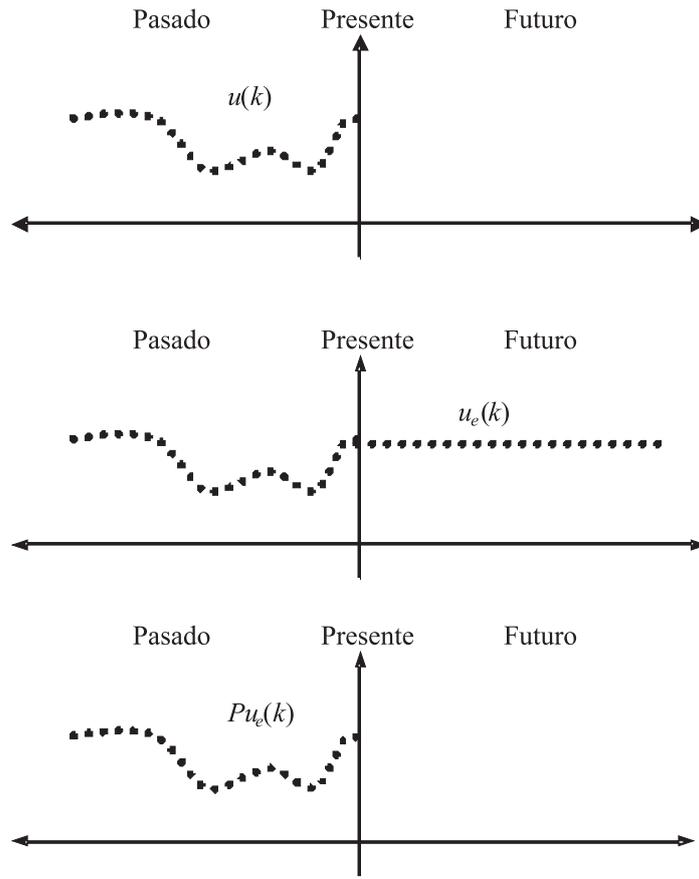


Figura 4.7: Sucesiones  $u$ ,  $u_e$  y  $Pu_e$

siendo  $q$  el operador de retardo que satisface  $q^\tau u(k) = u(k - \tau)$ ,  $k \in \mathbb{Z}$ .

En la Figura 4.8 observamos en (a) la sucesión de entrada  $u(k)$ , en (b) su extensión no causal  $u_e(k)$ , en (c) la respuesta  $f$  a la entrada  $u(k)$  y finalmente en (d) la respuesta  $f$  a la entrada  $u_e(k)$ . En (d) puede observarse resaltado el valor de  $Tu = fu_e(0)$ .

Considerando el caso  $k = -1$  en la ecuación (4.41), en la Figura 4.9 vemos cómo se puede recuperar al operador  $f$  a partir del valor de la funcional  $T$ . En (a) podemos observar la señal  $u(k)$  afectada por el operador de retardo  $q^{-1}$ , sobre la que actúa el operador de truncamiento  $P$ , generando la sucesión  $Pq^{-1}u(k)$  que está esquematizada en (b). La versión no causal  $(Pq^{-1}u)_e(k)$  de la sucesión anterior la podemos ver en (c), mientras que en (d) se ve, resaltado, el valor de la funcional  $TPq^{-1}u$ . Este proceso, visualizado para el caso  $k = -1$ , se aplica a cada valor  $k$  permitiendo recuperar así el operador no lineal  $f$ .

En base a las definiciones y consideraciones previas es posible expresar en una forma compacta la noción de continuidad y causalidad para un operador TI.

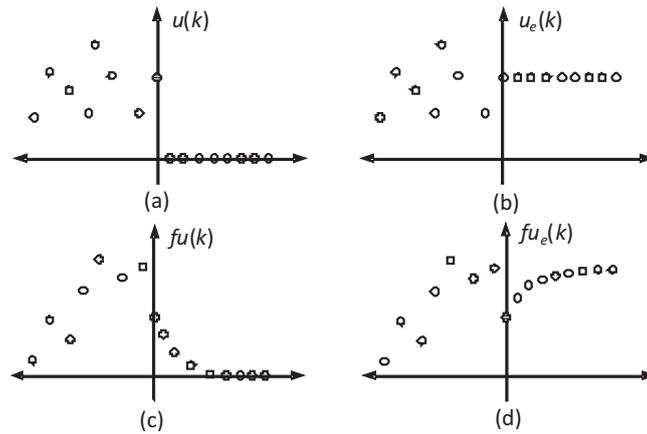


Figura 4.8: Sucesiones causal, extendida y sus respuestas asociadas

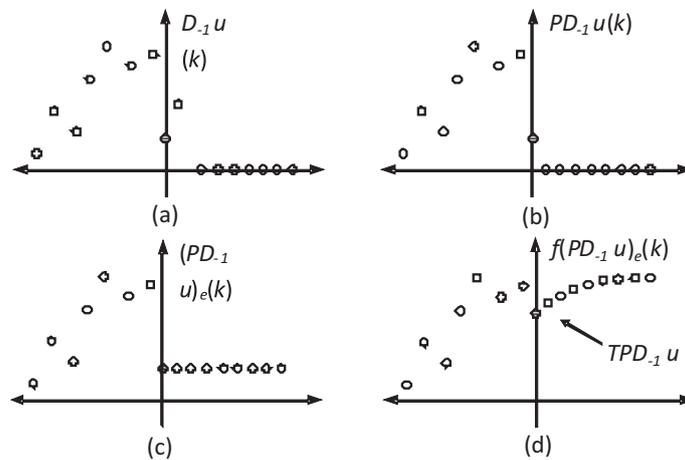


Figura 4.9: Reconstrucción de  $f$  a partir de la funcional asociada  $T$

**Definición 4.4.2** Un operador  $f$  TI es causal y continuo si y sólo si para cada  $u, v \in \ell^\infty(\mathbb{Z})$  y cada  $\varepsilon > 0$ , existe  $\delta > 0$  tal que

$$\sup_{k \leq 0} |u(k) - v(k)| < \delta \Rightarrow |fu(0) - fv(0)| < \varepsilon. \tag{4.42}$$

La definición anterior expresa que para un operador continuo la salida actual del sistema  $fu(0)$  depende continuamente de toda la historia previa de la entrada. es decir, que un suceso acontecido en un pasado remoto ( $K \rightarrow -\infty$ ) afecta de igual manera a la salida actual, que un suceso ocurrido en un pasado reciente (valores de  $k$  cercanos a cero).

Para una gran cantidad de sistemas reales la influencia del pasado de las entradas sobre la salida presente disminuye a medida que se retrocede en el tiempo. Cuando se piensa en este concepto, se arriba de manera natural a la idea de sistemas con *memoria evanescente*.

Intuitivamente, un operador tiene memoria evanescente si dadas dos señales de entrada que están cercanas en algún sentido en el pasado reciente (no necesariamente en el pasado remoto), producen salidas cercanas en el presente. El concepto de memoria evanescente expresa que a medida que se retrocede en el pasado ( $K \rightarrow -\infty$ ), la influencia de la entrada  $u(k)$  sobre la salida presente  $fu(0)$  se hace despreciable o se evanesce. Este efecto se puede expresar generando una secuencia que decrece a medida que aumenta  $k$  y que pesa distinto los efectos de la entrada sobre la salida de acuerdo al tiempo transcurrido. Entonces consideramos una secuencia decreciente  $w : \mathbb{Z}^+ \rightarrow (0, 1]$ , con  $\lim_{k \rightarrow \infty} w(k) = 0$  en base a la cual definimos la norma ponderada o  $w$ -norma en  $\ell(\mathbb{Z}^+)$ .

**Definición 4.4.3** Para  $u \in \ell^\infty$ ,

$$\|u\|_w = \sup_{k \leq 0} |u(k)|w(-k). \quad (4.43)$$

se dice la *norma ponderada de  $u$* .

Es fácil ver que el efecto práctico de la norma (4.43) es hacer desaparecer o evanescer a la entrada  $u(k)$  cuando  $k \rightarrow -\infty$ .

**Definición 4.4.4** Boyd y Chua, [10]. Se dice que un operador causal  $T : \mathbf{K} \rightarrow \ell^\infty$ , siendo  $\mathbf{K} \subset \ell^\infty$ , tiene *memoria evanescente (m.e.) en  $\mathbf{K}$*  si existe una sucesión decreciente  $w : \mathbb{Z}^+ \rightarrow (0, 1]$ ,  $\lim_{k \rightarrow \infty} w(k) = 0$  tal que para cada  $u \in \mathbf{K}$  y  $\varepsilon > 0$ , existe  $\delta > 0$  tal que

$$\sup_{k \leq 0} |u(k) - v(k)|w(-k) < \delta \Rightarrow |Tu(0) - Tv(0)| < \varepsilon, \quad (4.44)$$

$\forall v \in \mathbf{K}$ . La sucesión  $w$  se denomina *sucesión de pesos*.

Esta definición expresa cómo la salida del sistema es influenciada cada vez menos por señales remotas de la entrada. La sucesión de pesos  $w(k)$  simula hacer desaparecer o evanescer al pasado remoto de la entrada hasta que en el límite lo anula.

### Observaciones:

1. De (4.42) y (4.44) es fácil ver que si un operador tiene memoria evanescente entonces es continuo. Sin embargo, la recíproca no es válida como lo muestra el siguiente ejemplo. Sea  $N : \ell^\infty(\mathbb{Z}) \rightarrow \ell^\infty(\mathbb{Z})$ , definido como  $Nu(k) = \sup_{n \leq k} u(n)$ , denominado *operador detector de pico*. Es fácil ver que el operador  $N$  no tiene memoria evanescente dado que, independientemente de lo que suceda en su pasado, la salida mantiene el valor máximo de la señal de entrada. Sin embargo el operador  $N$  es continuo, ya que para todo  $u, v \in \ell^\infty(\mathbb{Z})$  se verifica  $\|Nu - Nv\|_\infty \leq \|u - v\|_\infty$ .
2. Si un operador tiene memoria evanescente con respecto a una sucesión  $w$  entonces tiene memoria evanescente con respecto a cualquier otra sucesión de pesos  $\tilde{w}$  que domine

a  $w$ , *i.e.* tal que  $\tilde{w}(k) \geq w(k) \quad \forall k \in \mathbb{Z}^+$ . Luego, si utilizamos la sucesión de pesos definida por

$$\tilde{w} = \max_{k \in \mathbb{Z}^+} [w(k), (1+k)^{-1}], \quad (4.45)$$

y renombramos esta sucesión como  $w$ , entonces simplemente podemos suponer que la sucesión de pesos satisface  $w(k)^{-1} \leq (1+k) \quad \forall k$ .

El concepto de memoria evanescente no es un concepto reciente. Ya se hacía mención del mismo en los trabajos de Volterra [95]

*Un postulado muy natural es suponer que, en un momento dado, la influencia de entradas lejanas disminuye gradualmente,*

en los de Wiener [102]

*Estamos suponiendo que la salida de la red no depende del pasado remoto. ... Luego, estamos considerando redes en las cuales la salida es asintóticamente independiente del pasado remoto,*

y en otros trabajos posteriores como el de George [38], el de Barrett [4] y más tarde el de Root [78], quien hace mención de los operadores con memoria finita. En la Figura 4.10 ilustramos esta idea.

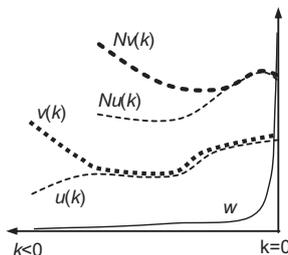


Figura 4.10: Sistema con memoria evanescente

De ahora en más supondremos que las entradas  $u$  al sistema no lineal satisfacen la condición

$$u \in \mathbf{K} \subset \ell^\infty, \text{ siendo } \mathbf{K} = \{u \in \ell^\infty : \|u\|_\infty \leq r, r > 0\}. \quad (4.46)$$

Ahora bien,

*dada una entrada  $u$  que satisface (4.46), ¿cuáles son las salidas del sistema lineal discretizado?*

En el dominio discreto, los operadores Laguerre y Kautz se pueden definir a través de la transformada Zeta como vimos en el Capítulo 2.

Entonces, las salidas del bloque lineal dinámico identificado se pueden escribir matricialmente como

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N]^T, \quad (4.47)$$

donde  $N$  es la cantidad de observaciones disponibles del proceso real y cada fila  $\mathbf{z}_i$  de la matriz  $\mathbf{Z}$  es un vector de longitud igual a la cantidad  $n$  de filtros utilizados para la identificación, *i.e.*

$$\mathbf{z}_k = (L_1 u(k), L_2 u(k), \dots, L_n u(k)), \quad 1 \leq k \leq N. \quad (4.48)$$

O bien, si utilizamos filtros Kautz es

$$\mathbf{z}_k = (K_1 u(k), K_2 u(k), \dots, K_n u(k)), \quad 1 \leq k \leq N. \quad (4.49)$$

Así definido, nuestro sistema lineal tiene memoria evanescente en  $\mathbf{K}$ , donde  $\mathbf{K}$  está definido en (4.46). Esto surge inmediatamente del siguiente teorema (ver [10]), que caracteriza en forma simple a los sistemas lineales invariantes en el tiempo, de memoria evanescente y del hecho que los filtros Laguerre y Kautz pertenecen a  $\ell^\infty$ , como probaremos más adelante.

**Teorema 4.4.5** *Sea  $C(\mathbb{R})$  el espacio de las funciones continuas con la métrica del supremo.*

(I)  *$A : C(\mathbb{R}) \rightarrow C(\mathbb{R})$  es un operador lineal, invariante en el tiempo con memoria evanescente si y sólo si tiene una representación mediante convolución*

$$Au(t) = \int_0^\infty u(t - \tau) h(d\tau),$$

donde  $h$  es una medida acotada definida en  $\mathbb{R}_0^+$ .

(II)  *$A : \ell^\infty \rightarrow \ell^\infty$  es un operador lineal, invariante en el tiempo con memoria evanescente si y sólo si tiene una representación por convolución*

$$Au(n) = \sum_{k=0}^{\infty} h(k) u(n - k), \quad (4.50)$$

donde  $h \in \ell^1(\mathbb{Z}^+)$ .

**Demostración:** Ver Boyd y Chua [10]. □

La propiedad de memoria evanescente puede expresarse en términos de la funcional  $T$  asociada al operador no lineal  $f$ .

Entonces, un operador no lineal  $f$  tiene memoria evanescente en  $\mathbf{K}$  si y sólo si  $T$  es continuo respecto de la  $w$ -norma en  $P\mathbf{K} = \{Pu : u \in \mathbf{K}\}$  (ver [10]).

Estamos ahora en condiciones de probar el teorema de aproximación mencionado al comienzo de esta sección y a ello nos abocaremos en la siguiente.

### 4.4.2. Teorema de Aproximación

En primer lugar, vamos a enunciar el Teorema de aproximación y luego probaremos los lemas que hacen falta para llegar a la tesis del mismo.

**Teorema 4.4.6** *Sea  $\mathbf{K}$  como se definió en (4.46) y sea  $f : \mathbf{K} \rightarrow \ell^\infty$  un operador causal, tiempo invariante con memoria evanescente en  $\mathbf{K}$ . Entonces para un dado  $\varepsilon > 0$ , existen  $n$  operadores Laguerre o Kautz  $\{z_1(\cdot), z_2(\cdot), \dots, z_n(\cdot)\}$  definidos en  $\mathbf{K}$  y una función CLATAN  $f_{lat} \in PWL_H[\mathbf{S}]$ , donde  $\mathbf{S}$  está definido en (4.29), tal que para cualquier  $u \in \mathbf{K}$*

$$\|y - \tilde{y}\|_\infty < \varepsilon$$

siendo  $y$  la salida del sistema definida por (5.8) e  $\tilde{y}$  la salida del modelo definida por (4.5).

**Demostración:** Para el caso de operadores Laguerre, la demostración del teorema es presentada por Sentoni en [84]. Para la demostración del caso de operadores Kautz ver el Apéndice C.  $\square$

La demostración del Teorema 4.4.6 está basada en el Teorema de aproximación de Stone-Weierstrass (ver, por ejemplo, Dieudonné [28]) y que Boyd y Chua [10] enuncian de la siguiente manera.

**Teorema 4.4.7** *Sea  $\mathbf{E}$  un espacio métrico compacto y  $\mathcal{G}$  un conjunto de funcionales lineales continuas definidas en  $\mathbf{E}$  tales que separa puntos en  $\mathbf{E}$ , es decir si  $u, v \in \mathbf{E}$  y  $u \neq v$  entonces existe  $G \in \mathcal{G}$  tal que  $Gu \neq Gv$ . Sea  $T$  cualquier funcional continua en  $\mathbf{E}$  y  $\varepsilon > 0$ . Entonces existe un polinomio  $p : \mathbb{R}^n \rightarrow \mathbb{R}$  y  $n$  funcionales  $G_1, G_2, \dots, G_n \in \mathcal{G}$  tales que, para todo  $u \in \mathbf{E}$ ,*

$$|Tu - p(G_1u, G_2u, \dots, G_nu)| < \varepsilon. \quad (4.51)$$

Además del Teorema de Stone-Weierstrass, para la demostración del Teorema 4.4.6 necesitamos los resultados del Capítulo 2, Sección 2.7.4.

Para poder aplicar el Teorema 4.4.7, son necesarios los siguientes lemas previos, cuyas demostraciones están incluidas en el Apéndice C para que no se pierda la continuidad en la lectura.

**Lema 4.4.8** *El espacio  $\mathbf{K}_- = \{Pu : u \in \mathbf{K}\}$ , siendo  $P : \ell^\infty \rightarrow \mathbf{K}$  el operador proyección, es compacto con la  $w$ -norma con pesos  $\|\cdot\|_w$  definida por (4.43).*

**Demostración:** Ver [84].  $\square$

**Lema 4.4.9** *El conjunto  $\mathcal{G}$  de funcionales  $G_j$  asociadas a los operadores discretos de Kautz  $K_j(z) = \sum_{i=0}^{\infty} k_j(i)z^{-i}$ , definidas como  $G_j u = \sum_{i=0}^{\infty} k_j(i)u(-i)$ , son continuas con la  $w$ -norma  $\|\cdot\|_w$ .*

**Demostración:** Ver Apéndice C. □

**Lema 4.4.10** *El conjunto  $\mathcal{G}$  de funcionales definidas en Lema 4.4.9 separa puntos en  $\ell^\infty(\mathbb{Z}_-)$ .*

**Demostración:** Ver Apéndice C. □

Estamos ahora en condiciones de probar el teorema de aproximación.

### Demostración del Teorema 4.4.6

**Demostración:** De acuerdo al Lema 4.4.8,  $\mathbf{E} = \mathbf{K}_-$  es un espacio métrico compacto con la  $w$ -norma  $\|\cdot\|_w$ . Por los Lemmas 4.4.9 y 4.4.10, las funcionales  $G_j$  asociadas a los operadores discretos de Kautz son continuas en la  $w$ -norma  $\|\cdot\|_w$  y el conjunto  $\mathcal{G}$  separa puntos en  $\ell^\infty(\mathbb{Z}_-)$ . Entonces, por el Teorema de Stone-Wierstrass, cualquier funcional  $T$  asociada a un operador no lineal  $f$ , invariante en el tiempo con memoria evanescente, puede aproximarse por un conjunto de  $n$  funcionales  $G_1, G_2, \dots, G_n$  definidas como en el Lema (4.4.9), seguido por un polinomio  $p : \mathbb{R}^n \rightarrow \mathbb{R}$ , de modo que

$$|Tu - p(G_1u, G_2u, \dots, G_nu)| < \varepsilon/2, \quad (4.52)$$

$\forall \varepsilon > 0, \forall u \in \ell^\infty(\mathbb{Z}_-)$ .

Como las funcionales de Kautz son exponencialmente estables y  $u \in \ell^\infty(\mathbb{Z}_-)$ , existen constantes finitas  $M > 0$  y  $U > 0$  tales que

$$\begin{aligned} |G_j u| &\leq \|u\|_\infty \sum_{i=0}^{\infty} |k_j(i)| \\ &\leq M \|u\|_\infty \\ &\leq MU. \end{aligned}$$

Luego,  $\mathbf{v} = [G_1u, G_2u, \dots, G_nu] \in \mathbf{S}$  y por (2.70) de la Subsección 2.7.1, Capítulo 2, la aproximación CLATAN  $f_{lat}$  sobre el dominio  $\mathbf{S}$  satisface

$$|f_{lat}(\mathbf{v}) - p(\mathbf{v})| \leq \varepsilon(\delta) \doteq \delta L, \quad (4.53)$$

siendo  $L$  la constante de Lipschitz de  $f_{lat}(\cdot)$ .

En consecuencia, para cualquier  $\varepsilon > 0$  dado, siempre es posible elegir  $\delta^*$  tal que  $\varepsilon(\delta^*) < \varepsilon/2$ ; entonces podemos concluir que

$$|Tu - f_{lat}(\mathbf{v})| \leq |Tu - p(\mathbf{v})| + |p - f_{lat}(\mathbf{v})| < \varepsilon \quad (4.54)$$

para cualquier  $\mathbf{v} \in \mathbf{S}$ ,  $u \in \mathbf{K}_-$ .

Finalmente, probamos que  $f_{lat}$  es la aproximación deseada a cualquier operador no lineal  $f$  con memoria evanescente en  $\mathbf{K}$ . Sea  $u \in \mathbf{K}$ ,  $k \in \mathbb{Z}$ .

Como  $Pq^{-k}u \in \ell^\infty(\mathbb{Z}_-)$ , y  $\mathbf{v} = [G_1Pq^{-k}u, G_2Pq^{-k}u, \dots, G_nPq^{-k}u] \in \mathbf{S}$ , entonces por (4.54)

$$\begin{aligned} |y(k) - \tilde{y}(k)| &= |fu(k) - f_{lat}(K_1u, K_2u, \dots, K_nu)| \\ &= |TPq^{-k}u(k) - f_{lat}(\mathbf{v})| < \varepsilon, \end{aligned} \quad (4.55)$$

para cualquier  $\varepsilon > 0$ ,  $u \in \mathbf{K}$ ,  $k \in \mathbb{Z}$  y la demostración del Teorema 4.4.6 sigue inmediatamente de (4.55).  $\square$

## 4.5. Ejemplos de Aplicación

En esta sección presentamos varios ejemplos de aplicación de las estructuras de identificación Laguerre-CLATAN y Kautz-CLATAN propuestas en esta tesis y la metodología empleada para estimar sus parámetros. El primer ejemplo consiste en identificar un sistema discreto SISO no lineal, causal, TI y con memoria evanescente que, por sus características, posee una representación Laguerre-CLATAN. En el segundo ejemplo modelamos un sistema discreto SISO, no lineal, amortiguado y con saturación en la entrada. Este sistema se aproximó por un modelo Kautz-CLATAN, que tiene ventajas sobre uno Laguerre-CLATAN, respecto del orden que es necesario utilizar para aproximar este tipo de sistemas. En el último ejemplo se considera la aplicación de la estructura de identificación Laguerre-CLATAN propuesta, en el modelado de un sistema MIMO.

En las estructuras de identificación que proponemos, las entradas al bloque no lineal del sistema son las salidas del bloque lineal dinámico ya identificado, dispuestas en una matriz de datos  $\mathbf{Z}$ , *i.e.*, es la señal  $\mathbf{u}$  de entrada al sistema, previamente filtrada a través de las funciones de transferencia que conforman la parte lineal dinámica de la estructura. Las salidas de este bloque son las medidas disponibles  $y(k)$ .

Para construir los dos primeros modelos, utilizamos un algoritmo de selección [14], que determina qué datos utilizar de la matriz  $\mathbf{Z}$ , definida en (4.47), (4.48) y (4.49), para poder garantizar que los símplexes en los que se divide la región determinada por las salidas del bloque lineal dinámico van a quedar identificados. El algoritmo se basa en la ventaja de que la transformación no lineal entrada-salida es estática y por lo tanto se puede utilizar esta característica para seleccionar las muestras convenientemente.

Podemos describir matemáticamente dicho algoritmo de la siguiente manera. Si  $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_p\} \in \mathbf{S}$  es la matriz de las salidas del bloque lineal identificado, entonces elegimos sólo aquellas muestras  $\mathbf{z}_i(k)$  tales que

$$|\mathbf{z}_i(k) - \mathbf{z}_j(k')| > \xi, \quad \forall i, j, k, k'.$$

La constante  $\xi$  es un número pequeño prefijado, que tiene relación con el tamaño de la grilla utilizada para definir los símplexes en los cuales queda dividida la región  $\mathbf{S}$ .

Utilizando este algoritmo de selección, se obtiene una cantidad de muestras uniformemente distribuidas sobre todos los símplexes, sin necesidad de que la cantidad de las mismas sea una dificultad para el cálculo. Una vez aplicado el algoritmo, hallamos los parámetros de las funciones CLATAN utilizando la metodología SM descrita en el este capítulo, implementada en MATLAB.

Las siguientes consideraciones se aplican a todos los ejemplos.

1. Generamos la entrada  $\mathbf{u}$  como una señal aleatoria, con distribución uniforme, de longitud  $N$ . Aplicamos estas señales a los sistemas dinámicos considerados, obteniendo una señal de salida  $\mathbf{y}$ .
2. En los dos primeros ejemplos utilizamos 80% de los datos E/S para identificación y el resto para validación. En el último ejemplo, tomamos 2/3 de los datos disponibles para el proceso de identificación y el resto para el de validación.
3. Con el objetivo de mostrar el desempeño del modelo, adicionamos ruido a las muestras utilizadas en la identificación.
4. Identificamos los polos de las funciones base elegidas.
  - a) Caso Laguerre: el polo se evalúa simulando un modelo ARX de primer orden con los datos E/S utilizados en la identificación.
  - b) Caso Kautz: los polos se evalúan simulando un modelo ARX de segundo orden con los datos E/S que se utilizan para identificar el modelo.
5. Utilizando el método de los coeficientes de Lipschitz dado en el Apéndice B, estimamos el orden del bloque lineal dinámico, *i.e.* la cantidad de funciones base. A continuación propagamos las entradas a través de las funciones base, obteniendo la matriz  $\mathbf{Z}$ , que es la entrada al bloque no lineal estático.
6. Aplicamos el algoritmo de selección descripto al comienzo de la sección.
7. Identificamos el bloque no lineal estático, utilizando funciones CLATAN y teoría de estimación SM para evaluar los parámetros.

8. En todos los casos presentamos los resultados de evaluación de la estructura en presencia de ruido aditivo de salida y la validación cruzada de los modelos obtenidos. En los dos primeros ejemplos, testeamos los resultados de la aproximación sobre un nuevo conjunto de datos, que llamamos *conjunto de prueba*.
9. Calculamos el *coeficiente de correlación FIT* y la raíz cuadrada del *error cuadrático medio RMSE* (por sus siglas en inglés, Root Mean Square Error) sobre el conjunto de datos para identificación y para validación de la estructura de modelo propuesta.

El coeficiente de correlación *FIT* se define como

$$FIT = \left( 1 - \frac{\|\tilde{\mathbf{y}} - \mathbf{y}\|_2}{\|\mathbf{y} - \bar{\mathbf{y}}\|_2} \right) 100, \quad (4.56)$$

donde  $\tilde{\mathbf{y}}$  es la salida estimada e  $\bar{\mathbf{y}}$  es la media del vector de salidas disponibles  $\mathbf{y}$ , expresado en porcentaje (ver, por ejemplo, [58]). Es una medida que sirve para evaluar la bondad de ajuste del modelo a los datos, ya que mide la capacidad predictiva del modelo ajustado.

Calculamos el error de peor caso  $E_y(\Phi_c, \epsilon)$  para el modelo identificado.

Debemos remarcar que los modelos discretos SISO y MIMO utilizados para identificación, así como los sistemas continuos originales, no son estructuras tipo Wiener.

#### 4.5.1. Ejemplo 1: Sistema SISO no Lineal Discreto

Este primer ejemplo, debido a Narendra y Parthasarathy [68], corresponde a un sistema discreto SISO, no lineal, causal, invariante en el tiempo y con memoria evanescente, dado por la siguiente ecuación en diferencias no lineal

$$y(k+1) = \frac{y(k)}{1+y^2(k)} + u^3(k) + e(k+1), \quad e(k+1) = 0,75(e(k) + r(k)), \quad (4.57)$$

donde  $r(k)$  es una señal uniforme, generada aleatoriamente entre los valores  $-0,5$  y  $0,5$ , por lo que  $e(k)$  puede considerarse como un ruido coloreado adicionado a la salida  $y(k)$ .

Para realizar la identificación se generó una entrada aleatoria  $\mathbf{u}$  de 6000 valores entre 1 y 2 con distribución uniforme. Al aplicar dicha entrada al sistema (4.57) y registrar la salida se obtuvo un conjunto de muestras  $y(k)$ , con  $k = 1, \dots, 6000$ .

El ruido fue adicionado a las primeras 4800 muestras utilizadas para el proceso de identificación, y las 1200 muestras sin ruido restantes fueron utilizadas para validación.

Con estos valores entrada-salida, aplicamos una de las estructuras de aproximación SISO descriptas en la Sección 4.2 y el método SM propuesto en la Subsección 4.3.1, para estimar los parámetros del modelo construido con las funciones CLATAN.

En este caso, por las características del sistema, para identificar el bloque lineal dinámico elegimos funciones de transferencia de Laguerre. Evaluamos el polo de los sistemas de Laguerre utilizados en la identificación del subsistema lineal dinámico. Dicho parámetro fue estimado como  $a = 0,0873$ . En base al análisis de los coeficientes de Lipschitz del sistema (ver Apéndice B), seleccionamos una expansión Laguerre de orden 2.

Con la estimación del polo y seleccionada la base, formamos la matriz de datos  $\mathbf{Z}$  de orden  $2 \times 4800$  con los 4800 valores de la entrada, filtrados por los sistemas de Laguerre  $L_1u(k)$  y  $L_2u(k)$  con polo estimado  $a = 0,0873$ .

Una vez determinado el modelo del bloque lineal dinámico del sistema, debemos determinar los parámetros del bloque no lineal estático que vamos a aproximar utilizando funciones CLATAN.

En este ejemplo, el conjunto de datos seleccionados para la identificación del bloque no lineal, luego de haber aplicado el algoritmo de selección descripto, fue de 2141 muestras  $\mathbf{z}_i(k)$ ,  $i = 1, 2$ . En consecuencia, la matriz  $\mathbf{Z}$  es ahora de orden  $2 \times 2141$ .

Entonces el esquema de aproximación es

$$y(k) = f_{lat}(L_1u(k), L_2u(k)) = f_{lat}(z_1, z_2), \quad (4.58)$$

donde  $f_{lat}$  es una función CLATAN definida sobre el dominio

$$\mathbf{S} = \{\mathbf{z} \in \mathbb{R}^2 : \alpha_i \leq z_i \leq \beta_i, i = 0, 1\}$$

Utilizamos como región  $\mathbf{S}$  para hallar la aproximación CLATAN al conjunto

$$\mathbf{S} = [1, 2] \times [1, 2],$$

con un paso de grilla  $ndiv = 2$  en cada intervalo.

El modelo nominal  $f_{lat}$  fue identificado utilizando estimadores basados en el conjunto de sistemas posibles (4.11) y tomando como entradas al sistema no lineal, las salidas de los filtros de Laguerre. El número de parámetros identificados para esta estructura fue 9, que es el número de funciones CLATAN usadas para la representación del sistema no lineal. La cota para el ruido fue elegida como  $\epsilon = 1$ .

Para validar el modelo utilizamos los 1200 datos restantes, a los que filtramos con los sistemas de Laguerre de segundo orden con el polo previamente estimado. Luego de aplicar el algoritmo de selección, la matriz  $\mathbf{Z}$  de salidas del bloque lineal que se obtuvo fue de orden  $2 \times 956$ . En particular, calculamos el coeficiente de correlación  $FIT$  sobre el conjunto de datos para validación de la estructura de modelo propuesta.

Finalmente, el modelo fue testeado utilizando un nuevo conjunto de datos E/S. Para ello excitamos al sistema utilizando una señal de entrada aleatoria con distribución uniforme de 1000 datos, obteniendo la salida  $\mathbf{y}$  como respuesta del sistema. Formamos la matriz de datos

$\mathbf{Z}$  aplicando los sistemas de Laguerre a la secuencia de entrada y seleccionamos 803 muestras que pertenecen al conjunto  $\mathbf{S}$ .

En la Tabla 4.2, se indica la raíz cuadrada del error cuadrático medio ( $RMSE$ ) y el  $FIT$  obtenidos para el modelo identificado sobre el conjunto de validación y para el nuevo conjunto de datos de prueba. El error optimal de peor caso obtenido en la identificación fue  $E_y(\Phi_c, \epsilon) = 5,4654$ .

	Error $RMSE$	$FIT$
Identificación	0,3195	82,8875 %
Validación	0,1027	94,4948 %
Test	0,3197	82,7381 %

Tabla 4.2: Error  $RMSE$  y coeficiente de correlación  $FIT$  para el Ejemplo 1

Figuras 4.11 y 4.13 muestran los resultados de la validación y del test, respectivamente, junto con la salida del sistema real y el comportamiento dinámico de la salida, obtenido con los parámetros de la estructura Laguerre-CLATAN. En esas figuras, la línea continua (azul) representa la respuesta del sistema y la línea punteada (rojo) la salida del modelo. Para tener una mejor visión de los resultados, presentamos una ampliación de la validación del modelo.

Luego, podemos concluir que esta estructura no modela ruido.

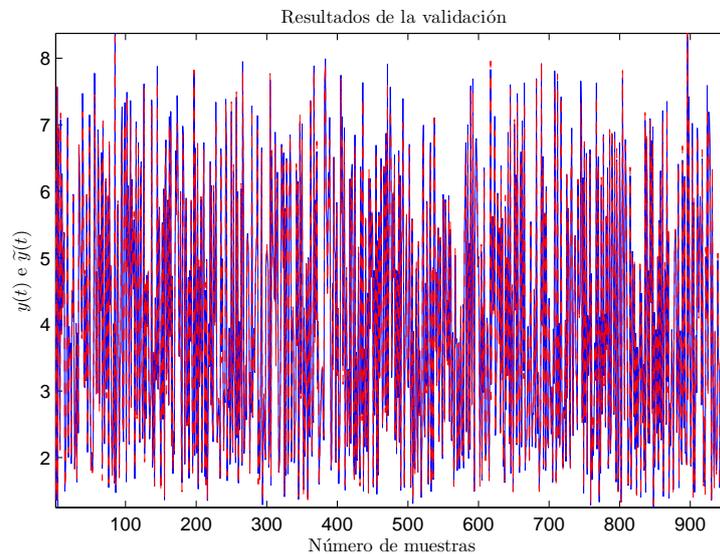


Figura 4.11: Validación del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

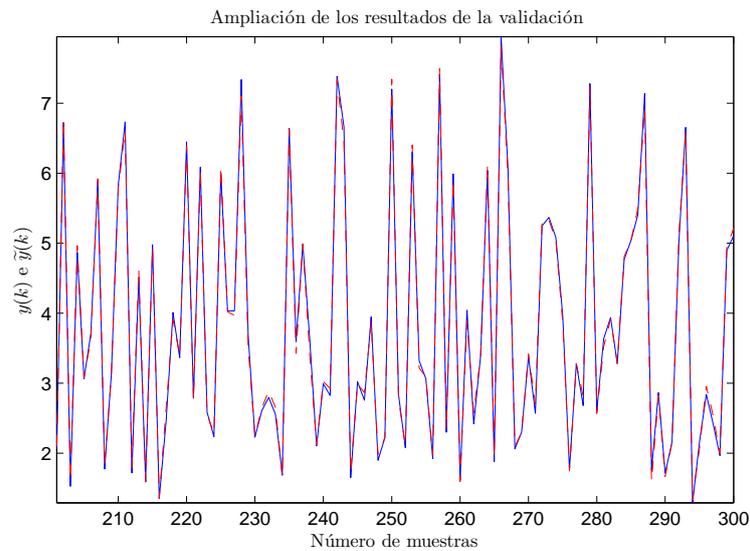


Figura 4.12: Validación del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

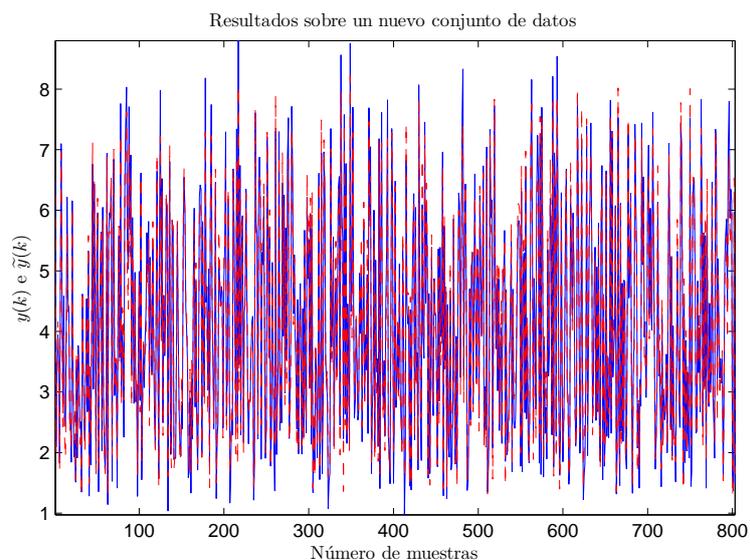


Figura 4.13: Resultados sobre el conjunto de datos de prueba del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

#### 4.5.2. Ejemplo 2: Sistema Mecánico con Saturación en la Entrada

En este ejemplo se modela la posición de la masa de un sistema masa-resorte amortiguado. El sistema es SISO, no lineal discreto, invariante en el tiempo y cumple con las condiciones

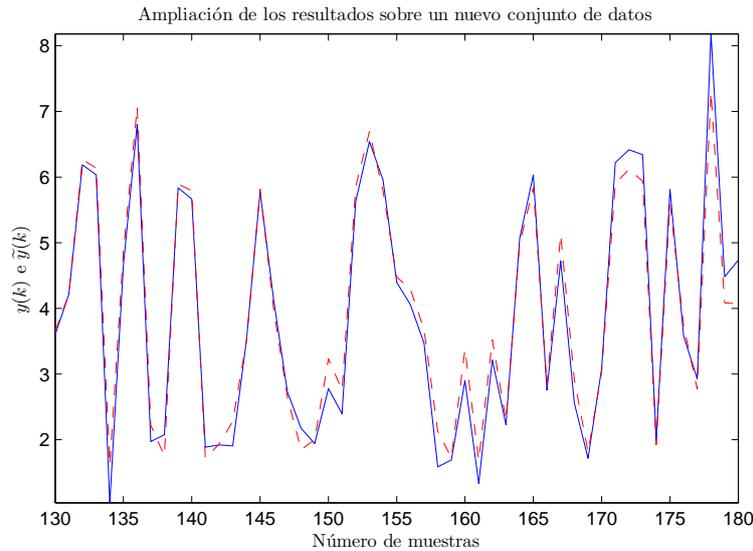


Figura 4.14: Resultados sobre el conjunto de datos de prueba del modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

de causalidad y memoria evanescente. Además, por su estructura posee una representación Kautz-CLAT, es decir, el bloque lineal es expandido mediante sistemas de Kautz. En la Figura 4.15 se puede ver un esquema del mismo.

Consideramos el sistema SISO, no lineal discreto, similar al presentado en [64], descrito por la siguiente ecuación en diferencias:

$$y(k+1) = 0,002y(k) - 0,2y(k-1) + 0,1\text{sen}(y(k-1)) + \tanh(3u(k)) + e(k+1), \quad (4.59)$$

que representa una aproximación discreta de un sistema masa-resorte amortiguado con estiramiento o fuerza del resorte lineal, fuerza de amortiguación no lineal y una no linealidad estática en la entrada (saturación en la entrada). La entrada  $u(k)$  es una fuerza que actúa sobre la masa y la salida  $y(k)$  es la posición de la masa en el instante  $k$ .

El ruido  $\mathbf{e}$  es una señal uniforme generada aleatoriamente de amplitud  $\|\mathbf{e}\|_{\infty} \leq 0,025$  y fue adicionada a las muestras utilizadas en el proceso de identificación. Exitamos el sistema con una señal aleatoria con distribución uniforme cuya amplitud variaba en el intervalo  $[-1, 1]$ . Generamos un conjunto de 8000 datos entrada-salida utilizando el sistema (4.59). Los primeros 6400 datos fueron utilizados para el proceso de identificación, mientras que los 1600 datos restantes, sin ruido, fueron usados para validar el modelo.

El proceso de identificación fue realizado siguiendo la descripción dada al comienzo de la sección. Como el sistema no lineal dado es amortiguado, en este caso proponemos el uso de sistemas de Kautz con dos parámetros para aproximar el bloque lineal dinámico. La ventaja

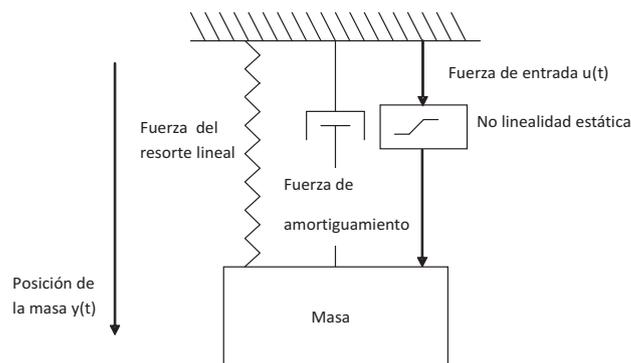


Figura 4.15: Ejemplo 2. Sistema masa-resorte amortiguado

de utilizar bases de Kautz en vez de bases de Laguerre para sistemas amortiguados, es que se necesita un menor número de filtros para aproximar el subsistema lineal.

Debemos estimar los polos de los sistemas de Kautz. Los valores de estos parámetros fueron identificados como  $\beta = -0,0042 + 0,3248j$  y  $\bar{\beta} = -0,0042 - 0,3248j$ .

Para evaluar el número de funciones de la base de Kautz utilizamos el método de los coeficientes de Lipschitz descrito en el Apéndice B. En este caso seleccionamos una base de Kautz de orden 2.

Identificamos el bloque lineal dinámico con los sistemas de Kautz de segundo orden y obtenemos una matriz de datos de salida  $\mathbf{Z}$ , que son los datos de entrada al sistema no lineal. Aplicando el algoritmo de selección, se seleccionaron 1292 muestras de la matriz de salidas del modelo lineal identificado.

La estructura de aproximación ahora es

$$y(k) = f_{lat}(K_1 u(k), K_2 u(k)) = f_{lat}(z_1, z_2), \quad (4.60)$$

donde  $f_{lat}$  es una función CLATAN definida sobre el dominio

$$\mathbf{S} = \{\mathbf{z} \in \mathbb{R}^2 : \alpha_i \leq z_i \leq \beta_i, i = 0, 1\}.$$

Utilizamos como región  $\mathbf{S}$  para hallar la aproximación CLATAN al conjunto

$$\mathbf{S} = [-0,5, 0,5] \times [-0,5, 0,5],$$

con  $ndiv = 2$  divisiones en cada intervalo.

Identificamos los parámetros del modelo nominal  $f_{lat}$ , usando estimadores basados en el conjunto de sistemas posibles (4.11), donde la entrada al sistema no lineal es la salida de los filtros de Kautz. El número de parámetros identificados para esta estructura es 9, que es el número de funciones CLATAN utilizadas para la representación. La cota del ruido seleccionada fue  $\epsilon = 0,3$ .

Para validar el modelo, utilizamos los 1600 datos restantes sin ruido, y luego de identificar el subsistema lineal, mediante el algoritmo de selección obtuvimos 377 muestras distribuidas uniformemente sobre todos los símlices en los que queda dividido el dominio  $\mathbf{S}$  de la función estática no lineal  $f_{lat}$ .

Como en el ejemplo anterior, el modelo fue testeado utilizando un nuevo conjunto de datos. Exitamos el sistema con una señal aleatoria con distribución uniforme de 1000 datos. Identificamos el bloque lineal dinámico filtrando los datos de entrada con los sistemas de Kautz, cuyos polos fueron estimados en la identificación. Después de seleccionar las salidas del modelo lineal, utilizamos 69 muestras que pertenecen al conjunto  $\mathbf{S}$ .

En la Tabla 4.3, se muestran el  $RMSE$  y el coeficiente de correlación  $FIT$  (ver (4.56)) obtenidos para el modelo identificado sobre el conjunto de validación y sobre el conjunto de datos de prueba. El error optimal de peor caso para la identificación fue  $E_y(\Phi_c, \epsilon) = 2,5648$ .

	Error $RMSE$	$FIT$
Identificación	0,1076	82,1147 %
Validación	0,1061	82,1308 %
Test	0,1660	71,4035 %

Tabla 4.3: Error  $RMSE$  y coeficiente de correlación  $FIT$  para el Ejemplo 2

Figuras 4.16 y 4.18 muestran los resultados de la validación del modelo y sobre el conjunto de datos de prueba, respectivamente, junto con la salida del sistema real y el comportamiento dinámico de la salida obtenida con los parámetros de la estructura Kautz-CLAT. En esas figuras, la línea sólida (azul) representa la respuesta del sistema y la línea punteada (rojo) representa la salida del modelo. En orden de tener una mejor visión de los resultados, presentamos una ampliación de la validación del modelo.

Podemos concluir que esta estructura no modela la salida del ruido.

### 4.5.3. Ejemplo 3: Modelo MIMO de una Unidad de Generación de Vapor

En este ejemplo mostramos cómo es posible extender las ideas de modelación presentadas para contemplar modelos MIMO. Aplicaremos la metodología propuesta en el modelado de la presión y el nivel de agua de una unidad de generación de vapor tipo domo (UGVD). El proceso consiste en una caldera tipo domo para una central termoeléctrica de 200 MW. El

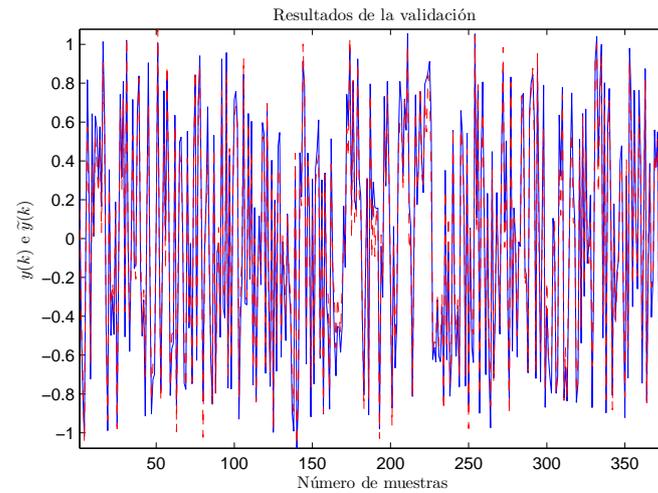


Figura 4.16: Validación del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

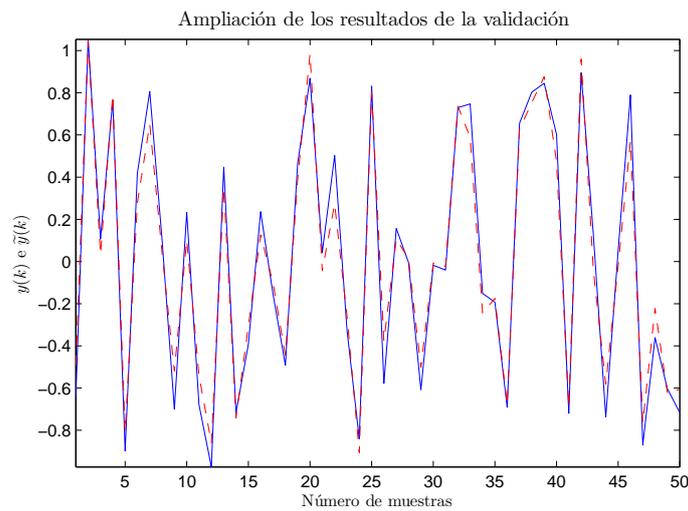


Figura 4.17: Validación del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

modelo fue desarrollado por [74]

$$\begin{aligned}
 \frac{dP}{dt} &= -0,00193SP^{1/8} - 0,000736w_c + \\
 &\quad 0,014524F + 0,00121L + 0,000176T_e \\
 \frac{dS}{dt} &= 10c_vP^{1/2} - 0,78571S \\
 \frac{dL}{dt} &= 0,00893w_c + 0,002F + 0,463c_v - 610^{-6}P^2 - \\
 &\quad 0,00914L - 8,210^{-5}L^2 - 0,007328S.
 \end{aligned} \tag{4.61}$$

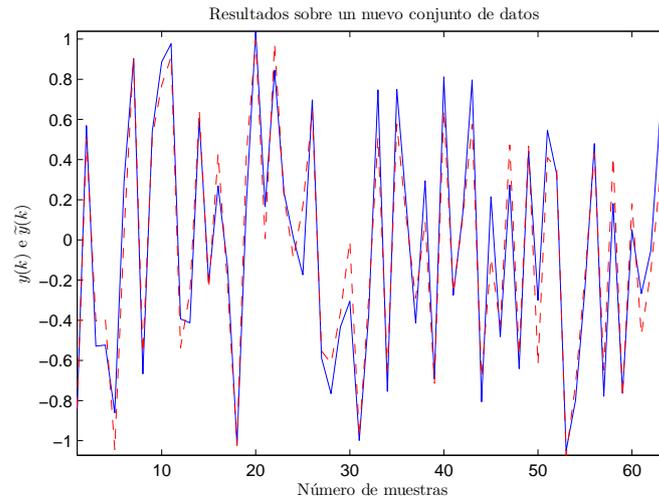


Figura 4.18: Resultados sobre el conjunto de datos de prueba del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

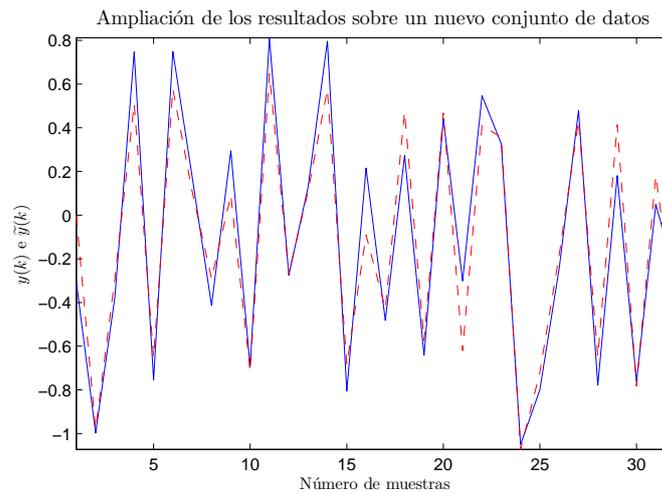


Figura 4.19: Resultados sobre el conjunto de datos de prueba del modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo)

Los estados del modelo no lineal del generador de vapor con domo son la presión del domo ( $P$ ), el caudal de vapor a la turbina de alta presión ( $S$ ) y el nivel del domo ( $L$ ). Los estados  $P$  y  $L$  son las variables controladas. Existen dos variables manipuladas que son el caudal de combustible ( $F$ ) y el caudal de alimentación de agua ( $w_c$ ), y dos perturbaciones que son la temperatura del agua ( $T_e$ ) y la posición de la válvula de control ( $c_v$ ). Los valores de las variables en el estado estacionario considerado se presentan en la Tabla 1 4.4. En

lo sucesivo llamaremos  $y = [P, L]^T$  y  $u = [F, w_c]^T$  a los vectores de variables controladas y manipuladas, respectivamente. Para ello, generamos un conjunto de 30000 datos con un

VARIABLE	VALOR
$F(kg/s)$	38,5736
$w_c(kg/s)$	190,9620
$T_e(K)$	310
$c_v$	0,8

Tabla 4.4: Variables de la UGVD

tiempo de muestreo de 20 segundos. Con la finalidad de excitar al sistema, se utilizan entradas (variables manipuladas) que se generan sumando a sus valores nominales perturbaciones aleatorias. Dichas perturbaciones aleatorias pueden tomar valores comprendidos entre  $\pm 20\%$  del respectivo valor nominal. Cada valor se mantiene durante una cantidad de muestras, a fin de poder identificar la ganancia en estado estacionario. Para ilustrar la situación desfavorable de adquisición de medidas con ruido, se consideró que los datos de salidas del sistema  $y_1$  e  $y_2$  se encuentran corruptos con un ruidos acotados por  $\epsilon_1 = 0,5 \text{ randn}(\text{long.}(y_1))$  y  $\epsilon_2 = 0,5 \text{ randn}(\text{long.}(y_2))$ , respectivamente. La Figura 4.20 muestra el conjunto de datos de salida ( $P$  y  $L$ ) sin el ruido aditivo, de los cuales las primeras 20000 muestras se utilizan para identificar y el resto para validar el modelo obtenido.

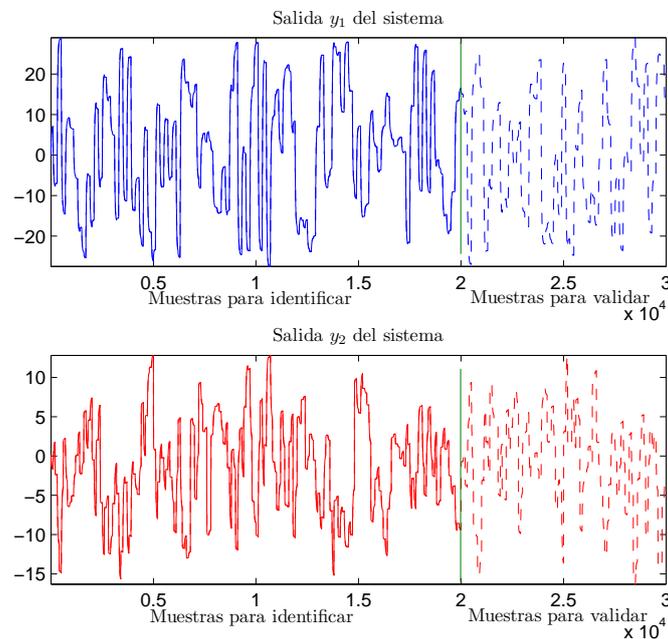


Figura 4.20: Salidas de la unidad de generación de vapor.  $y_1$ : presión,  $y_2$ : nivel

En este caso, seleccionamos una expansión de Laguerre de orden 1 con polos estimados  $a_1 = 0,95$  y  $a_2 = 0,8$  para cada entrada al sistema real.

El modelo nominal  $f_{lat}$  fue identificado utilizando teoría de estimación SM para la evaluación de los parámetros. Tomamos las salidas de los sistemas de Laguerre como las entradas al bloque no lineal estático. El número de parámetros identificados para cada una de las estructuras no lineales fue 9. Las cotas estimadas para el ruido fueron  $\epsilon_1 = 3,54$  y  $\epsilon_2 = 2,71$ , para las medidas disponibles  $\mathbf{y}_1$ ,  $\mathbf{y}_2$ , respectivamente.

Para la respuesta  $\mathbf{y}_1$  del sistema, la raíz cuadrada del error cuadrático medio y el coeficiente de correlación  $FIT$  obtenidos para el modelo identificado fueron  $RMSE = 0,8866$ ,  $FIT = 94,08$  y  $RMSE = 1,0174$ ,  $FIT = 93,5216$  sobre el conjunto de datos disponibles para identificar y para validar el modelo, respectivamente. El error optimal de peor caso en la identificación fue  $E_y(\Phi_{cc}, \epsilon) = 2,3974$ .

Para la respuesta  $\mathbf{y}_2$  del sistema, el  $RMSE = 0,5947$  y el  $FIT = 90,3804$  sobre el conjunto de datos disponibles para identificar, y  $RMSE = 0,6464$ ,  $FIT = 89,7152$  en la validación. El error optimal de peor caso dió  $E_y(\Phi_{cc}, \epsilon) = 2,4240$ .

las Figuras 4.21 y 4.22 presentan el comportamiento dinámico de los modelos  $\tilde{\mathbf{y}}_1$  e  $\tilde{\mathbf{y}}_2$  comparados con las salidas del sistema  $\mathbf{y}_1$  e  $\mathbf{y}_2$ . Realizamos un estudio de validación con

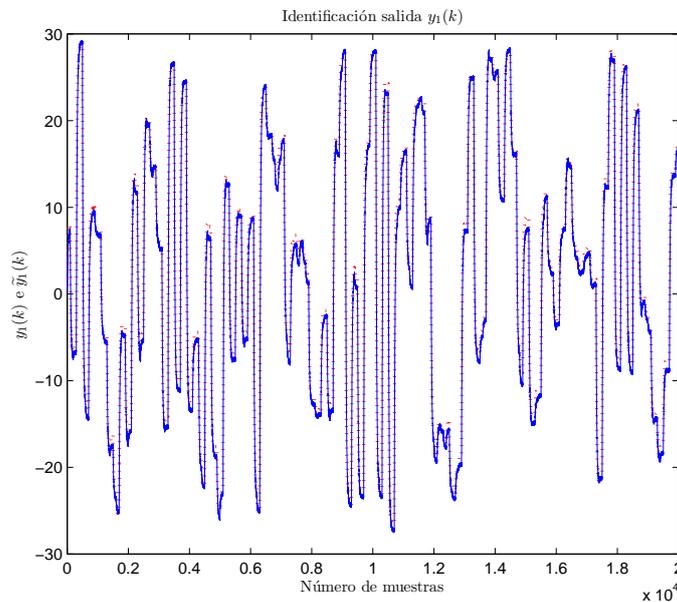


Figura 4.21: Modelo Laguerre-CLATAN. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo)

los datos restantes no utilizados en la identificación del modelo. Las Figuras 4.23 y 4.24 muestran los resultados.

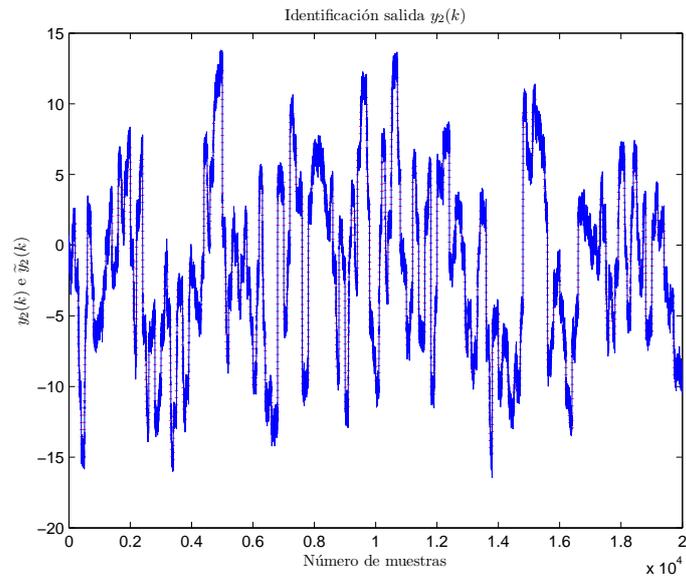


Figura 4.22: Modelo Laguerre-CLATAN. Sistema ( $y_2$ : azul), modelo ( $\tilde{y}_2$ : rojo)

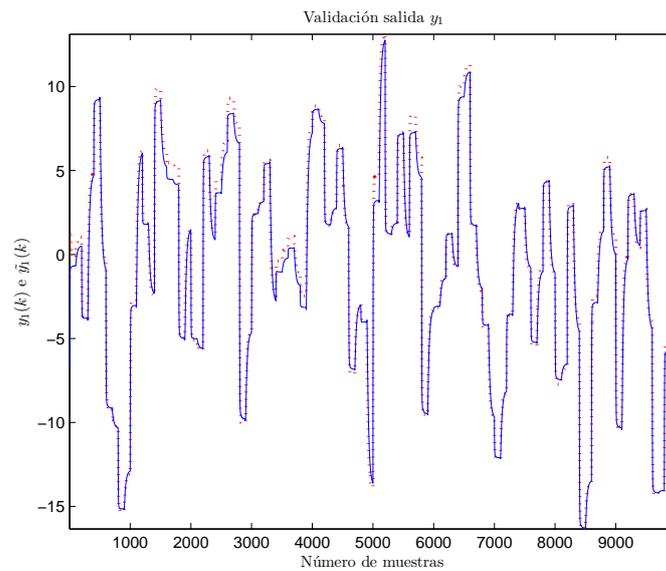


Figura 4.23: Modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $y_1$ : azul), modelo ( $\tilde{y}_1$ : rojo)

## 4.6. Conclusiones

En este capítulo hemos presentado las estructuras Laguerre-CLATAN y Kautz-CLATAN que proponemos para la identificación de sistemas no lineales causales, invariantes en el tiempo.

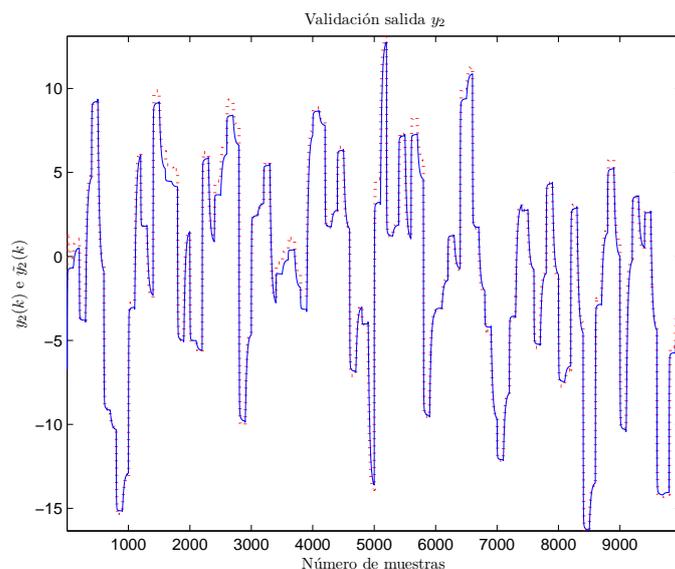


Figura 4.24: Modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $y_2$ : azul), modelo ( $\tilde{y}_2$ : rojo)

po, de memoria evanescente. Hemos enunciado y demostrado los teoremas que aseguran que nuestra estructura es capaz de aproximar esa clase de sistemas. Desarrollamos tres ejemplos de aplicación diferentes para mostrar la capacidad de identificación de nuestras estructuras. El primer ejemplo, por sus características, posee una representación Laguerre-CLATAN. En el segundo ejemplo modelamos un sistema discreto SISO, no lineal, sub-amortiguado y con saturación en la entrada. Este sistema se aproximó por un modelo Kautz-CLATAN. En el último ejemplo se considera la aplicación de la estructura de identificación Laguerre-CLATAN propuesta, en el modelado de un sistema MIMO. En todos los casos se presentan los resultados de la evaluación de la estructura en presencia de ruido aditivo de salida y la validación cruzada de los modelos obtenidos.

## Capítulo 5

# Identificación Robusta de los Modelos Laguerre-CLATAN y Kautz-CLATAN

En este capítulo presentamos una metodología para la identificación de los modelos Laguerre-CLATAN y Kautz-CLATAN con incertidumbre. La estructura de modelo nominal está dada por la representación paramétrica Laguerre-CLATAN o Kautz-CLATAN, ambas presentadas en el Capítulo 4, y el método se basa en la identificación de cotas de los parámetros para dichos modelos. Por lo tanto, describimos la incertidumbre asociada al modelo nominal como un conjunto de parámetros del bloque no lineal estático, cuyos valores se obtienen resolviendo un problema de optimización. La metodología resultante es robusta, en el sentido que la totalidad de los datos utilizados para la identificación del sistema en estudio, puede ser reproducida por al menos uno de los modelos generado con parámetros del conjunto identificado. Hemos abordado el problema para sistemas SISO y su extensión a sistemas MIMO. Ilustramos el algoritmo de identificación desarrollado con tres ejemplos, cuyos modelos nominales fueron obtenidos en el Capítulo 4, Sección 4.5.

### 5.1. Introducción

Una vez identificado el sistema real con los datos disponibles, es importante asegurar la robustez del modelo pues ello permite tener en cuenta la presencia de incertidumbre asociada al modelo nominal y evaluar cómo ésta podría afectar el modelo estimado.

Un enfoque típico para la identificación robusta consiste en definir un conjunto de posibles modelos que representen el comportamiento incierto del mismo. Si se asume el mismo formato para todos los modelos posibles en el conjunto de incertidumbre, entonces esta familia de modelos se define en términos de un conjunto de parámetros. En este contexto, es posible realizar una identificación robusta teniendo en cuenta un conjunto  $\Theta$  de parámetros del modelo, de manera que cuando se utilizan parámetros  $\theta \in \Theta$ , las entradas  $\mathbf{u}$  al proceso se mapean en un conjunto de salidas contenido en el conjunto de la totalidad de los datos de

salida  $\mathbf{y}$  del sistema. De esta manera, la incertidumbre se representa mediante una familia de modelos, todos con la misma estructura, pero que difieren en los valores de sus parámetros. Por otra parte, es posible tener una idea del conservativismo de esta descripción del modelo, comparando la salida  $\tilde{\mathbf{y}}$  estimada por el modelo identificado con la salida  $\mathbf{y}$  del proceso.

Existen en la literatura métodos diferentes para la identificación con incertidumbre de modelos orientados a bloques [5, 20, 87]. En [7] se presenta un algoritmo no iterativo para la identificación de sistemas SISO Wiener y Hammerstein con incertidumbre, que requiere que la no linealidad estática sea *invertible*. El método se basa en la identificación de las cotas de los parámetros de estos modelos; de esta manera, el conjunto de datos utilizados en la identificación puede ser reproducido por alguno de los modelos dentro de estas cotas. Esta metodología se extiende a sistemas MIMO [6] con la condición de invertibilidad de la transformación no lineal estática del modelo orientado a bloques.

En esta tesis asumimos que la no linealidad estática de la estructura nominal del modelo está representada por las funciones CLATAN, mientras que el bloque dinámico lineal se compone de un número finito de filtros de Laguerre o filtros de Kautz. Para estimar los parámetros de las funciones CLATAN, el problema se formula en el contexto de la Teoría de estimación SM desarrollada en el Capítulo 4. Comenzamos considerando modelos inciertos SISO tipo Wiener y luego extendemos la metodología presentada a sistemas tipo Wiener MIMO. La metodología resultante es robusta dado que la totalidad de los datos utilizados para la identificación del sistema, puede reproducirse por al menos uno de los modelos en el conjunto de parámetros identificados. Finalmente, presentamos tres ejemplos que ilustran las capacidades de identificación de las estructuras propuestas, los dos primeros para el caso SISO y el último para el caso MIMO.

## 5.2. Formulación del Problema

Los datos disponibles de un proceso real siempre se conocen con cierto grado de incertidumbre, por lo que es de gran interés evaluar cómo esta incertidumbre afecta al modelo nominal estimado.

De acuerdo a lo visto en el Capítulo 4, un modelo nominal Laguerre-CLATAN o Kautz-CLATAN se compone de un conjunto finito de filtros lineales de Laguerre o de Kautz, seguido de una transformación estática no lineal  $f_{lat} \in PWL_H[\mathbf{S}]$ . Es decir, el modelo lineal mapea la entrada  $u(k) \in \mathbb{R}$  en la señal intermedia  $\mathbf{z}(k) = \mathbf{z}[u(k)] \in \mathbb{R}^n$ , y la salida del modelo global es el resultado del bloque no lineal

$$\tilde{y}(k+1) = f_{lat}(\mathbf{z}[u(k)]) \in \mathbb{R}. \quad (5.1)$$

Recordemos que (Capítulo 2) las funciones CLATAN se definen en el dominio compacto

$$\mathbf{S} \doteq \{\mathbf{x} \in \mathbb{R}^n : a_i \leq x_i \leq a_i + \delta \text{ndiv}, i = 1, 2, \dots, n\}, \quad (5.2)$$

donde  $a_i \in \mathbb{R}$ ,  $\delta$  es el tamaño de la grilla,  $ndiv \in \mathbb{Z}_+$  es el número de divisiones asociados con el eje  $x_i$  y  $\mathbf{S}$  se subdivide utilizando una configuración simplicial  $H$ , como se vió en el Capítulo 2.

Cualquier  $f_{lat} \in PWL_H[\mathbf{S}]$  se puede escribir como

$$f_{lat}(\mathbf{x}) = \mathbf{c}^T \Lambda(\mathbf{x}), \quad (5.3)$$

donde  $\mathbf{c}$  es el vector de parámetros y  $\Lambda$  es la matriz de las funciones base definidas en  $\mathbf{S}$ .

Entonces, la ecuación (5.1) se convierte en

$$\begin{aligned} \tilde{y}(k+1) &= \mathbf{c}^T \Lambda(\mathbf{z}[u(k)]) \\ &= \mathbf{c}^T \Lambda(z_1[u(k)], z_2[u(k)], \dots, z_n[u(k)]), \end{aligned}$$

donde  $z_j[u(k)]$ , para  $j = 1, 2, \dots, n$  son las salidas de los filtros de Laguerre o de Kautz y  $\mathbf{c} \in \mathbb{R}^m$  es el vector de parámetros desconocidos del bloque no lineal que debe ser estimado. Una vez obtenido el modelo nominal, nuestro objetivo es realizar una identificación robusta del mismo.

En general, un modelo no lineal incierto está caracterizado por una familia de modelos, todos con la misma estructura, de la forma

$$\mathcal{F} = \{h : D \rightarrow \mathbb{R} : h(v) = h_N(v) + \Delta(v)\}, \quad (5.4)$$

donde  $h_N$  es el modelo nominal del proceso real y  $\Delta$  satisface  $\sup_{v \in D} \|\Delta(v)\| \leq K$ . Además, suponemos que se dispone de un conjunto  $H_s = \{h_1, h_2, \dots, h_s\}$  de datos medidos de miembros de  $\mathcal{F}$ , es decir,  $h_k = h(v_k)$ , con  $h \in \mathcal{F}$  y  $v_k \in D$ ,  $\forall k = 1, \dots, s$ .

En el Capítulo 4 estudiamos cómo obtener un modelo nominal  $h_N$  Laguerre-CLATAN o Kautz-CLATAN con datos disponibles del sistema. Ahora deseamos una representación del conjunto completo de modelos inciertos; para ello un planteo natural es hallar un modelo “superior”  $h_{sup}$  y un modelo “inferior”  $h_{inf}$  que satisfacen

$$h_{inf}(v) \leq h(v) \leq h_{sup}(v), \quad \forall v \in D. \quad (5.5)$$

En esta tesis consideramos que un modelo no lineal de incertidumbre para un sistema dinámico se puede describir mediante la definición de un conjunto de parámetros  $\mathcal{P}$  para el bloque estático no lineal, de la siguiente manera

$$\mathcal{P} = \{\mathbf{c} : \mathbf{c}^l \leq \mathbf{c} \leq \mathbf{c}^u\}, \quad (5.6)$$

donde las desigualdades se aplican entrada por entrada, *i.e.*,

$$c_j^l \leq c_j \leq c_j^u, \quad 1 \leq j \leq m. \quad (5.7)$$

Con el fin de obtener un modelo de incertidumbre, debemos determinar las cotas de los parámetros  $\mathbf{c}^l$  y  $\mathbf{c}^u$ .

Consideremos un sistema dinámico no lineal, SISO, Lipschitz continuo. Los datos disponibles se componen de  $N$  pares E/S  $\{(u(k), y(k)), k = 0, \dots, N - 1\}$  relacionados por

$$y(k + 1) = f(u(k)) + e(k), \quad (5.8)$$

donde  $f$  es una transformación no lineal desconocida y  $e(k)$  es ruido UBB.

Consideramos un esquema de aproximación tipo Wiener para (5.8) dado por (5.1).

Para definir el dominio  $\mathbf{S}$  de la función CLATAN asumimos que

$$u \in \mathbf{K} \doteq \{u \in \ell^\infty : \|u\|_\infty \leq U, U > 0\} \subset \ell^\infty. \quad (5.9)$$

Las salidas de los filtros de Laguerre o de Kautz satisfacen

$$\|z_j [u(\cdot)]\|_\infty \leq G \cdot U, \quad \forall j = 1, \dots, n, \quad (5.10)$$

donde  $G \doteq \|z_1\|_1$ . Entonces, el dominio de la función CLATAN se puede definir como

$$\mathbf{S} \doteq \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z}\|_\infty \leq G \cdot U\}, \quad (5.11)$$

el cual es particionado usando una partición simplicial  $H$  con una grilla de medida  $\delta$ .

Las entradas al mapeo no lineal, *i.e.* las salidas de los filtros lineales, deben estar presentes en todos los símlices. Dependiendo de la entrada, es posible tener una gran concentración de datos E/S en algunos pocos símlices y un pequeño número de datos en la mayoría de ellos. Para superar este problema, hemos utilizado el algoritmo de selección detallado en el Capítulo 4, Sección 4.5, que elimina datos de los símlices logrando una distribución uniforme.

Describimos la incertidumbre mediante ruido aditivo desconocido pero del cual conocemos una cota en la norma  $\ell^\infty$ , *i.e.*

$$|e(k)| \leq \epsilon, \quad \epsilon > 0, \forall k. \quad (5.12)$$

para un  $\epsilon$  dado.

Definimos ahora los espacios que necesitamos para aplicar la teoría de estimación SM a la estimación de las cotas de los parámetros  $\mathbf{c}$ . Como antes, el espacio normado finito dimensional  $X$  es el espacio de los parámetros desconocidos  $\mathbf{c} = [c_1, c_2, \dots, c_m]^T$ , y su dimensión depende del número  $n$  de filtros de Laguerre o de Kautz utilizados. En este caso,  $m = (ndiv + 1)^n$ , donde  $ndiv$  es el número de celdas asociadas con el eje  $x_i$  en la ecuación (2.59). De acuerdo con las ecuaciones (5.2) y (5.3), obtenemos

$$\tilde{y}(k + 1) = \mathbf{c}^T \Lambda(\mathbf{z}(k)), \quad (5.13)$$

donde  $\mathbf{z}(k) = [z_1(k), \dots, z_n(k)]$ . Consecuentemente, el operador de información puede elegirse como  $F(\mathbf{c}) = \mathbf{c}^T \Lambda(\mathbf{z})$ . Luego, en este caso,  $F(\mathbf{c})$  es lineal.

El vector  $\mathbf{y}$  de salidas del sistema pertenece al espacio normado de salidas disponibles

$$Y = \{\tilde{\mathbf{y}} : \tilde{\mathbf{y}} = \mathbf{c}^T \Lambda(\mathbf{z}[\mathbf{u}]), \text{ para algún } \mathbf{c} \in \mathcal{P}\}, \quad (5.14)$$

donde la norma en el espacio  $Y$  es  $\ell^\infty$ .

En lo que sigue, calculamos un conjunto estimador y el algoritmo central.

De acuerdo con la ecuación (5.12), definimos el conjunto de parámetros posibles como

$$FPS_y = \{\mathbf{c} \in X : \|\mathbf{y} - F(\mathbf{c})\|_\infty \leq \epsilon\} \subseteq \mathbb{R}^m, \quad (5.15)$$

y podemos asegurar que el conjunto  $FPS_y$  es un politopo convexo, ya que  $F(\mathbf{c})$  es lineal y la norma del espacio  $Y$  es  $\ell^\infty$  o  $\ell^1$  (ver Capítulo 4).

Para aproximar el conjunto  $FPS_y$  aplicamos un algoritmo de aproximación exterior. Como vimos en el Capítulo 4, Subsección 4.3.1, la propuesta más común es utilizar ortotopos como regiones aproximantes. En particular, en esta tesis utilizamos cajas exteriores de mínimo volumen (MOB) o cajas interiores de máximo volumen (MIB) con caras paralelas a los ejes coordenados.

El MOB que contiene a  $FPS_y$  puede calcularse hallando el máximo rango de variaciones posibles del conjunto de valores  $c_j$ , provisto por los intervalos de valores de la incertidumbre  $VUI_j$ , definidos en el Capítulo 4.

Para obtener el algoritmo central usando la norma  $\ell^\infty$ , es suficiente calcular el centro de Chebyshev del paralelotopo calculado previamente, como se estableció en el Capítulo 4. El error de identificación de peor caso viene dado por el radio de Chebyshev.

## 5.3. Modelo de Incertidumbre

Para obtener el modelo de incertidumbre, debemos determinar las cotas de los parámetros  $\mathbf{c}^l$  y  $\mathbf{c}^u$ . Entonces, debe existir algún  $\mathbf{c} \in \mathcal{P}$  tal que

$$\|\tilde{\mathbf{y}} - \mathbf{y}\|_\infty < \epsilon, \quad (5.16)$$

para  $\tilde{\mathbf{y}} \in Y$ , donde  $\epsilon$  es la cota del vector de ruidos  $\mathbf{e}$ .

Por lo tanto,  $\tilde{y}_{\min}(k+1)$  y  $\tilde{y}_{\max}(k+1)$  satisfacen

$$\begin{aligned} \tilde{y}_{\min}(k+1) - \epsilon &= \min_{\mathbf{c} \in \mathcal{P}} \mathbf{c}^T \Lambda(\mathbf{z}[u(k)]) - \epsilon \leq y(k+1) \\ \tilde{y}_{\max}(k+1) + \epsilon &= \max_{\mathbf{c} \in \mathcal{P}} \mathbf{c}^T \Lambda(\mathbf{z}[u(k)]) + \epsilon \geq y(k+1). \end{aligned}$$

Luego, podemos describir cualquier salida medida del sistema real en función del modelo de incertidumbre. Para que el modelo sea lo menos conservador posible, proponemos minimizar la amplitud entre las cotas  $\mathbf{c}^l$  y  $\mathbf{c}^u$ , *i.e.*, minimizar  $\|\mathbf{c}^l - \mathbf{c}^u\|$ .

Observemos que para cada entrada  $u(k)$ ,  $\Lambda(\mathbf{z}[u(k)]) \in \mathbb{R}^m$ . Definimos los vectores  $\Lambda^+(\mathbf{z}[u(k)])$  y  $\Lambda^-(\mathbf{z}[u(k)])$  con entradas positivas y con entradas negativas de  $\Lambda(\mathbf{z}[u(k)])$ , respectivamente. Esto es,

$$\begin{aligned}\Lambda^+(\mathbf{z}[u(k)]) &= \text{máx}(\Lambda(\mathbf{z}[u(k)]), 0) \\ \Lambda^-(\mathbf{z}[u(k)]) &= \text{mín}(\Lambda(\mathbf{z}[u(k)]), 0).\end{aligned}$$

Entonces, las desigualdades (5.3) se pueden escribir como

$$\mathbf{c}^{lT} \Lambda^+(u(k)) + \mathbf{c}^{uT} \Lambda^-(u(k)) \leq y(k+1) + \epsilon$$

$$\mathbf{c}^{lT} \Lambda^-(u(k)) + \mathbf{c}^{uT} \Lambda^+(u(k)) \geq y(k+1) - \epsilon,$$

de donde resulta que el problema de identificación robusta se puede formular como un problema de programación lineal. Vale entonces el siguiente teorema.

**Teorema 5.3.1** *Las cotas  $\mathbf{c}^l$  y  $\mathbf{c}^u$  de los parámetros del modelo de incertidumbre se obtienen resolviendo el siguiente problema de optimización*

$$\min \sum_{j=1}^m (\mathbf{c}^u - \mathbf{c}^l)_j \tag{5.17}$$

*sujeito a*

$$\mathbf{c}^{lT} \Lambda^+(u(k)) + \mathbf{c}^{uT} \Lambda^-(u(k)) \leq y(k+1) + \epsilon$$

$$\mathbf{c}^{lT} \Lambda^-(u(k)) + \mathbf{c}^{uT} \Lambda^+(u(k)) \geq y(k+1) - \epsilon.$$

Observemos que el problema de hallar  $\mathbf{c}^l$  y  $\mathbf{c}^u$  se reduce a resolver un problema de programación lineal. Este problema de optimización (5.17) producirá un modelo incierto que minimiza la magnitud de la incertidumbre y asegura la descripción del comportamiento de todos los datos disponibles.

La dimensión del vector de parámetros desconocidos  $\mathbf{c} = [c_1, c_2, \dots, c_m]^T$ , que depende del número  $n$  de filtros de Laguerre o de Kautz, entonces el número de variables de optimización  $c_j^l$  y  $c_j^u$  es  $2m$  y el número de restricciones es  $2mN$ .

### 5.3.1. Caso MIMO

Los resultados presentados pueden extenderse para identificar las cotas de incertidumbre de sistemas discretos, no lineales, acoplados, MIMO. Proponemos un esquema de modelo para este sistema como un conjunto de modelos E/S MISO, cada uno de ellos compuesto

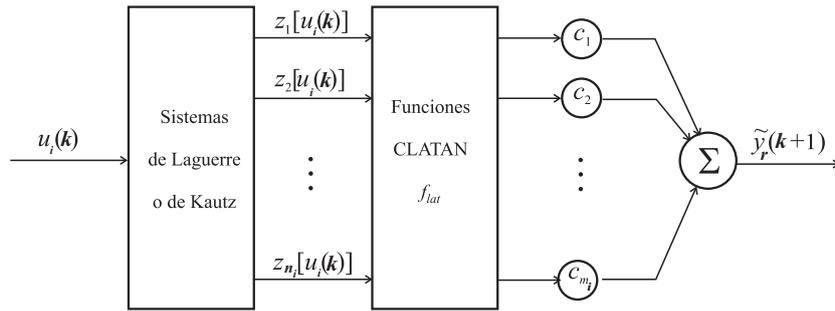


Figura 5.1: Estructura tipo Wiener SISO.

por la unión de conjuntos finitos de filtros discretos de Laguerre o de Kautz aplicados a cada vector de entrada  $\mathbf{u}_i$  del sistema MISO, seguido por una función CLATAN que representa la estática no lineal sin memoria del sistema. Observemos que la parte lineal dinámica del esquema MISO está representada por la unión de los conjuntos de filtros, cada uno de los cuales identifica el bloque lineal de cada sistema SISO contenido en él.

Para cada uno de estos modelos, estimamos los parámetros de las funciones CLATAN mediante la Teoría de estimación SM. De esta manera, cada modelo MISO tipo Wiener se describe como un conjunto de parámetros para el subsistema no lineal estático, cuyos valores se obtienen resolviendo un problema de optimización. La estructura de modelo tipo Wiener MIMO se representa como un conjunto de modelos MISO, cada uno de los cuales relaciona todas las entradas al sistema  $u_i(k)$  con cada salida  $y_r(k)$ . Luego la identificación de un sistema acoplado MIMO se reduce a la identificación de todos los posibles sistemas MISO contenidos en él. A fin de ilustrar el método de identificación de las cotas de incertidumbre de un sistema con estas características, aplicamos la metodología propuesta al sistema no lineal MIMO dado en el Capítulo 4, cuyas ecuaciones diferenciales describen el comportamiento de una unidad de generación de vapor tipo domo. El modelo nominal fue obtenido en dicho capítulo.

Suponemos que un sistema dinámico no lineal MIMO, acoplado es excitado por las múltiples entradas  $\mathbf{u}(k) = (u_1(k), u_2(k), \dots, u_p(k)) \in \mathbb{R}^p$ ; e  $\mathbf{y}(k) = (y_1(k), y_2(k), \dots, y_q(k)) \in \mathbb{R}^q$  son las respectivas salidas del proceso real en cada instante  $k$ , siendo  $0 \leq k \leq N - 1$ .

Sean  $\mathbf{u}_i = (u_i(0), u_i(1), \dots, u_i(N - 1))$ ,  $\mathbf{y}_r = (y_r(0), y_r(1), \dots, y_r(N - 1))$ , los vectores E/S que corresponden a cada sistema no lineal SISO contenido en el sistema dado.

La Figura 5.1 muestra la estructura de modelo SISO tipo Wiener, que consiste de un número finito de filtros lineales para cada conjunto de datos de entrada  $\mathbf{u}_i$ ,  $1 \leq i \leq p$ , seguido de una transformación estática no lineal  $f_{lat}$ , donde  $f_{lat}$  es una función CLATAN. Como describimos en el Capítulo 4, el modelo lineal  $\mathcal{L}$  mapea el vector de entrada  $\mathbf{u}_i \in \mathbb{R}^N$  en la señal intermedia  $\mathbf{Z} = \mathbf{z}[\mathbf{u}_i]$ , donde  $\mathbf{Z}$  es una matriz  $n_i \times N$ . El número de filas depende del número  $n_i$  de filtros aplicados a cada vector de entrada  $\mathbf{u}_i$ , y la salida del modelo es la salida  $\mathbf{y}_r$  del bloque no lineal. Asumimos que cada vector de salida  $\mathbf{y}_r$  de datos disponibles

contiene una componente de ruido aditivo  $\mathbf{e}_r$  UBB. Entonces, la estructura de modelo SISO está dada por

$$\tilde{y}_r(k+1) = f_{lat}(\mathbf{z}[u_i(k)]), \quad (5.18)$$

para  $1 \leq i \leq p$  y para cada  $1 \leq r \leq q$ .

En vista de la ecuación (5.18), proponemos un esquema de modelo para el sistema MIMO como un conjunto de modelos E/S MISO compuestos por la unión de  $q$  conjuntos finitos de filtros discretos de Laguerre o de Kautz aplicados a cada entrada  $\mathbf{u}_i$ ,  $1 \leq i \leq p$ , seguida por una función CLATAN. La salida del modelo viene dada por

$$\tilde{y}_r(k+1) = f_{lat}(\mathbf{z}[u_1(k)], \mathbf{z}[u_2(k)], \dots, \mathbf{z}[u_p(k)]), \quad (5.19)$$

para  $1 \leq r \leq q$ .

Luego, de las ecuaciones (5.3) y (5.19) obtenemos

$$\tilde{y}_r(k+1) = \mathbf{c}_r^T \Lambda(\mathbf{z}[u_1(k)], \dots, \mathbf{z}[u_p(k)]),$$

donde  $\mathbf{z}[u_i(k)] = (z_1[u_i(k)], z_2[u_i(k)], \dots, z_{n_i}[u_i(k)])$ , para  $i = 1, 2, \dots, p$  son las salidas de los filtros para cada entrada  $u_i(k)$  y  $\mathbf{c}_r \in \mathbb{R}^{m_r}$  es el vector de parámetros desconocidos del bloque no lineal que debemos estimar.

Para obtener la estructura MISO determinamos el vector de parámetros de la parte no lineal sin memoria  $\mathbf{c}_r$  utilizando el método de estimación SM. Luego, el ruido es UBB en la norma  $\ell^\infty$ ,

$$|e_r(k)| \leq \epsilon_r, \quad \text{para un dado } \epsilon_r > 0, \forall k. \quad (5.20)$$

Debemos determinar una función CLATAN para aproximar la transformación  $g$  entre la salida de los filtros  $\mathbf{z}[u_1(k)], \dots, \mathbf{z}[u_p(k)]$  y la salida del proceso  $y_r(k+1)$  relacionadas por

$$y_r(k+1) = g(\mathbf{z}[u_1(k)], \dots, \mathbf{z}[u_p(k)]) + e_r(k), \quad (5.21)$$

para todo  $0 \leq k \leq N-1$ . Con la finalidad de aplicar el método SM para ajustar los parámetros de las funciones CLATAN, definimos los siguientes espacios.

Sea  $(Y, \|\cdot\|_\infty)$  el espacio normado de dimensión  $N$  que contiene los datos medidos  $y_r(k)$  con ruido. El espacio normado finito dimensional  $X$  es el espacio de los parámetros  $\mathbf{c}_r = [c_r^1, c_r^2, \dots, c_r^{m_r}]^T$ , donde  $m_r = (ndiv + 1)^n$  y  $n = \sum_{i=1}^q n_i$ .

Como

$$\tilde{y}_r(k+1) = \mathbf{c}_r^T \Lambda(\mathbf{z}[u_1(k)], \dots, \mathbf{z}[u_p(k)]), \quad (5.22)$$

elegimos el operador de información como  $F(\mathbf{c}_r) = \mathbf{c}_r^T \Lambda(\mathbf{z}[u_1(k)], \dots, \mathbf{z}[u_p(k)])$ . Luego resulta que  $F(\mathbf{c}_r)$  es lineal.

En lo que sigue calculamos un conjunto estimador y el algoritmo central para la estimación de los parámetros  $\mathbf{c}_r$ .

De acuerdo con (5.20), definimos el conjunto de parámetros posibles como

$$FPS_{y_r} = \{\mathbf{c}_r \in X : \|\mathbf{y}_r - F(\mathbf{c}_r)\|_\infty \leq \epsilon_r\} \subseteq \mathbb{R}^{m_r}, \quad (5.23)$$

y podemos asegurar que el conjunto  $FPS_{y_r}$  es un politopo convexo, debido a que  $F(\mathbf{c}_r)$  es lineal y la norma del espacio  $Y$  es  $\ell^\infty$ .

Aproximamos el conjunto  $FPS_{y_r}$  con el MOB de mínimo volumen que lo contiene, de lados paralelos a los ejes coordenados.

Para cada  $1 \leq r \leq q$ , calculamos el MOB hallando el máximo rango de posibles variaciones de los valores  $c_r^j$ , dados por los intervalos de incertidumbre  $VUI_j^r$ , que en este caso quedan definidos como

$$VUI_j^r = \left[ \inf_{\mathbf{c}_r \in FPS_{y_r}} c_r^j, \sup_{\mathbf{c}_r \in FPS_{y_r}} c_r^j \right], \quad j = 1, 2, \dots, m_r. \quad (5.24)$$

Las medidas de los conjuntos  $VUI_j^r$  dan la longitud de cada lado del MOB que contiene al conjunto  $FPS_{y_r}$ , a lo largo del correspondiente eje coordenado  $x_j$ . Esto requiere, para cada  $r$ , la solución de  $2m_r$  problemas de optimización lineal con  $m_r$  variables y  $2N$  condiciones de desigualdad, como lo establece el siguiente teorema.

**Teorema 5.3.2** *Sea  $f_j(\mathbf{c}_r) = c_r^j$ ,  $j = 1, 2, \dots, m_r$ ,  $i = 1, \dots, p$ . Para cada  $1 \leq r \leq q$ , los intervalos  $VUI_j^r$  se pueden determinar resolviendo los siguientes problemas de programación lineal.*

$$\min_{\mathbf{c}_r \in FPS_{y_r}} f_j(\mathbf{c}_r), \quad j = 1, 2, \dots, m_r, \quad (5.25)$$

*sujeito a*

$$\|\mathbf{y}_r - F(\mathbf{c}_r)\|_\infty \leq \epsilon_r. \quad (5.26)$$

$$\max_{\mathbf{c}_r \in FPS_{y_r}} f_j(\mathbf{c}_r), \quad j = 1, 2, \dots, m_r, \quad (5.27)$$

*sujeito a*

$$\|\mathbf{y}_r - F(\mathbf{c}_r)\|_\infty \leq \epsilon_r. \quad (5.28)$$

**Demostración:** La demostración es inmediata teniendo en cuenta las ecuaciones (5.23) y (5.24).  $\square$

Para obtener el algoritmo central considerando la norma  $\ell^\infty$ , es suficiente calcular el centro de Chebyshev del MOB determinado previamente como vimos en el Capítulo 4. El error de identificación de peor caso viene dado por el radio de Chebyshev.

Para obtener el modelo de incertidumbre, debemos determinar las cotas de los parámetros  $\mathbf{c}_r^l$  y  $\mathbf{c}_r^u$ . Para ello aplicamos el teorema visto en la sección anterior para cada  $j$ .

**Teorema 5.3.3** *Las cotas  $\mathbf{c}_r^l$  y  $\mathbf{c}_r^u$  de los parámetros del modelo de incertidumbre se obtienen resolviendo el siguiente problema de optimización*

$$\min \sum_{j=1}^{m_r} (\mathbf{c}_r^u - \mathbf{c}_r^l)_j \quad (5.29)$$

sujeeto a

$$\mathbf{c}_r^{lT} \Lambda^+(u(k)) + \mathbf{c}_r^{uT} \Lambda^-(u(k)) \leq y(k+1) + \epsilon$$

$$\mathbf{c}_r^{lT} \Lambda^-(u(k)) + \mathbf{c}_r^{uT} \Lambda^+(u(k)) \geq y(k+1) - \epsilon.$$

El problema de optimización (5.29) producirá un modelo incierto que minimiza la magnitud de la incertidumbre y asegura la descripción del comportamiento de todos los datos disponibles.

La dimensión del vector de parámetros desconocidos  $\mathbf{c}_r = [c_r^1, c_r^2, \dots, c_r^{m_r}]^T$ , que depende del número  $n_i$  de filtros de Laguerre o de Kautz aplicados a cada entrada  $\mathbf{u}_i$ , es  $m_r$ . Entonces, el número de variables de optimización  $c_r^{j_l}$  y  $c_r^{j_u}$  es  $2m_r$  y el número de restricciones es  $2m_r N$ , para cada  $1 \leq r \leq q$ .

## 5.4. Ejemplos de Aplicación

### 5.4.1. Modelo de Incertidumbre del Sistema Masa-Resorte Amortiguado

En el Capítulo 4, Ejemplo 2, modelamos la posición de la masa de un sistema masa-resorte amortiguado. El sistema discreto es SISO, no lineal, amortiguado oscilatorio y con saturación en la entrada, por lo que aproximamos este sistema por un modelo Kautz-CLATAN.

La Figura 5.2 presenta el comportamiento dinámico del modelo  $\tilde{y}(k)$  comparado con la salida  $y(k)$  del sistema y pueden apreciarse también las cotas dinámicas. Es claro que la evolución dinámica del sistema está siempre dentro de esas cotas.

Hicimos un estudio de validación sobre el conjunto de datos no utilizados en la identificación del modelo. En la Figura 5.3 puede observarse una ampliación de los resultados.

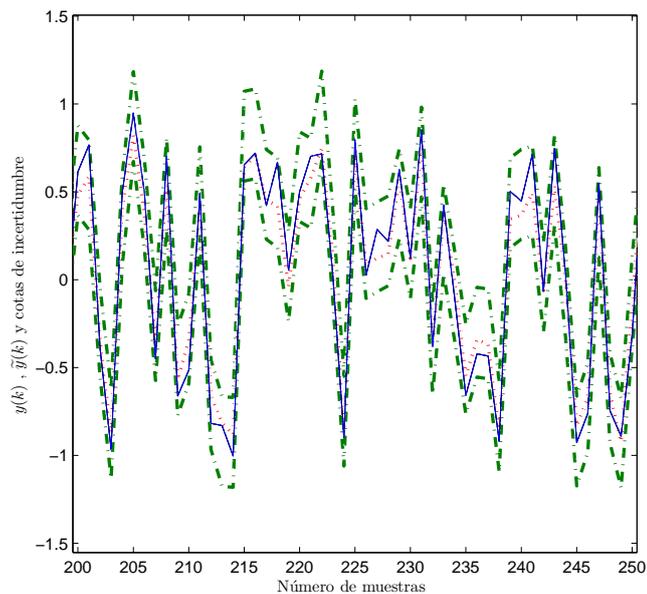


Figura 5.2: Cotas de incertidumbre para el modelo Kautz-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde).

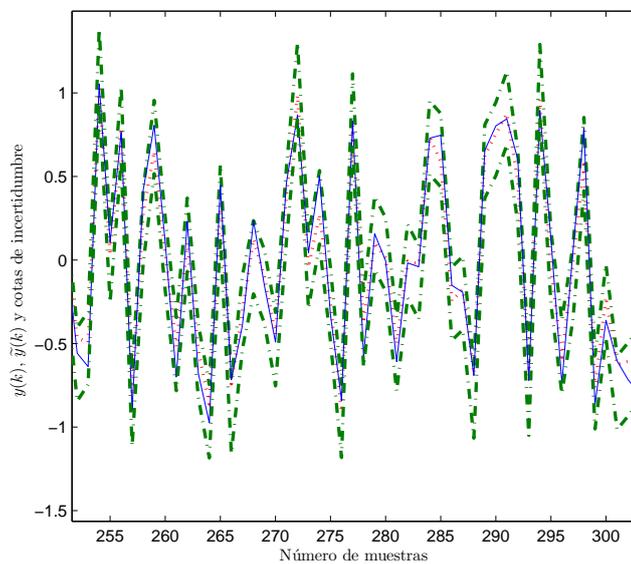


Figura 5.3: Cotas de incertidumbre para el modelo Kautz-CLATAN sobre el conjunto de validación. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde).

### 5.4.2. Modelo de incertidumbre de un Sistema SISO no Lineal Discreto

En este ejemplo, elegimos el sistema no lineal debido a Narendra y Parthasarathy [68], cuyo proceso de identificación del modelo nominal se llevó a cabo en el Capítulo 4.

La Figura 5.4 presenta una ampliación de la respuesta dinámica del sistema  $y(k)$ , el modelo Laguerre-CLATAN  $\tilde{y}(k)$  y las cotas dinámicas. Los resultados muestran que el conjunto de todos los datos disponibles está contenido en las bandas que son generadas por las cotas de incertidumbre.

Realizamos un test de validación con el conjunto de datos no utilizados en la identificación, en la Figura 5.5 mostramos un resultado representativo. Podemos ver que todos los datos se mantienen siempre dentro de esas cotas.

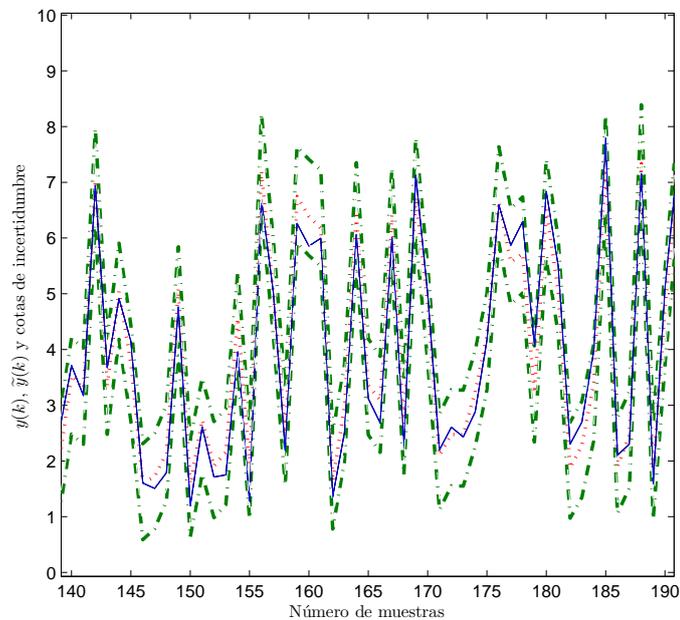


Figura 5.4: Cotas de incertidumbre para el modelo Laguerre-CLATAN. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde).

### 5.4.3. Modelo de Incertidumbre de la Unidad de Generación de Vapor

En este ejemplo aplicamos la metodología propuesta al sistema no lineal MIMO dado en el Capítulo 4, cuyas ecuaciones diferenciales describen el comportamiento de una unidad de

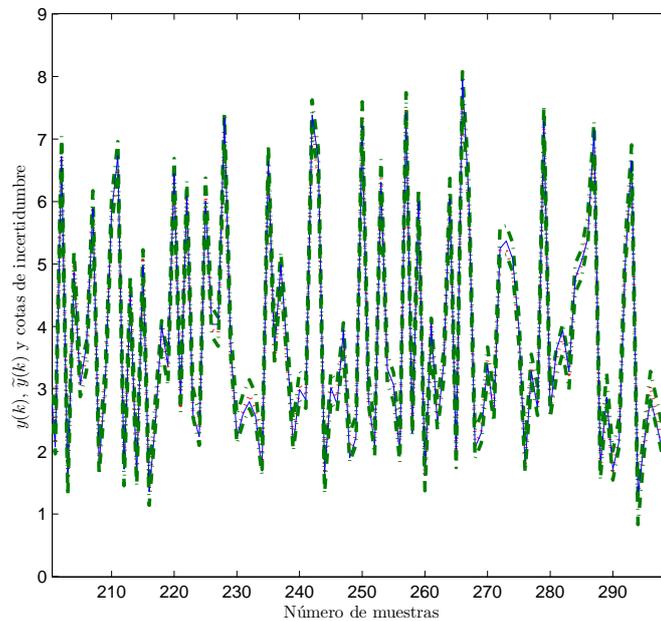


Figura 5.5: Cotas de incertidumbre para el modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $y$ : azul), modelo ( $\tilde{y}$ : rojo), cotas (verde).

generación de vapor tipo domo. Las variables de salida son la presión y el nivel de agua y el proceso de identificación del modelo nominal se llevó a cabo en dicho capítulo. A través de este ejemplo mostramos que es posible extender el método propuesto para la identificación robusta de modelos MIMO.

Las Figuras 5.6 y 5.7 muestran los datos de salida, la estimación y las cotas de incertidumbre. De estos gráficos resulta claro que los datos registrados están entre las cotas inferior y superior obtenidas mediante identificación robusta.

Realizamos un test de validación con el conjunto de datos no utilizados en la identificación, ver Figuras 5.8, 5.9, y un test de prueba sobre un nuevo conjunto de datos. Para este propósito generamos 10000 nuevos datos E/S en las mismas condiciones experimentales que generamos las muestras para la identificación. En las Figuras 5.10 y 5.11 pueden observarse los resultados obtenidos.

## 5.5. Conclusiones

En este capítulo presentamos la evaluación de un modelo de incertidumbre de los datos de E/S de un sistema SISO dinámico no lineal, mediante el uso de teoría de estimación SM

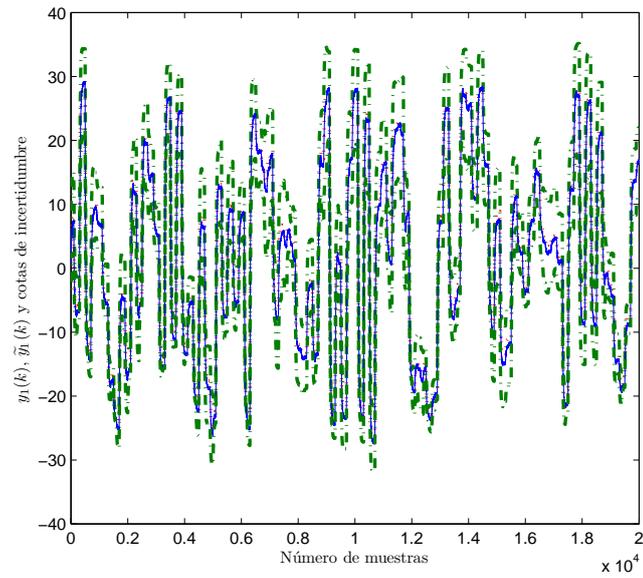


Figura 5.6: Cotas de incertidumbre para la salida  $y_1$  con ruido del modelo Laguerre-CLATAN. Sistema ( $y_1$ : azul), modelo ( $\tilde{y}_1$ : rojo), cotas (verde).

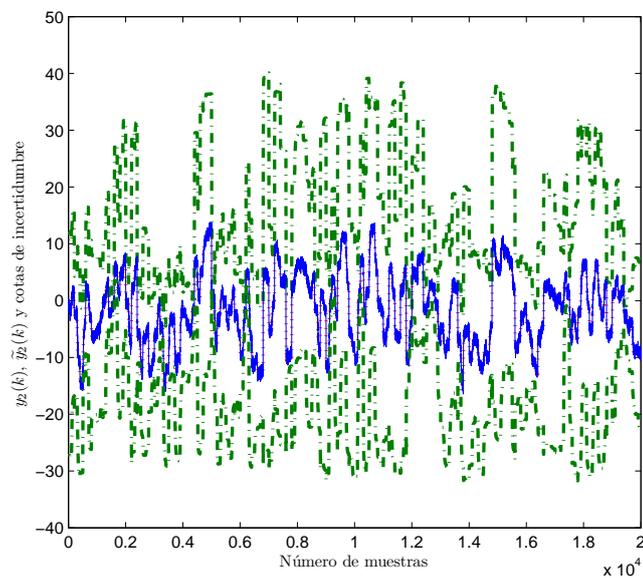


Figura 5.7: Cotas de incertidumbre para la salida  $y_2$  con ruido del modelo Laguerre-CLATAN. Sistema ( $y_2$ : azul), modelo ( $\tilde{y}_2$ : rojo), cotas (verde).

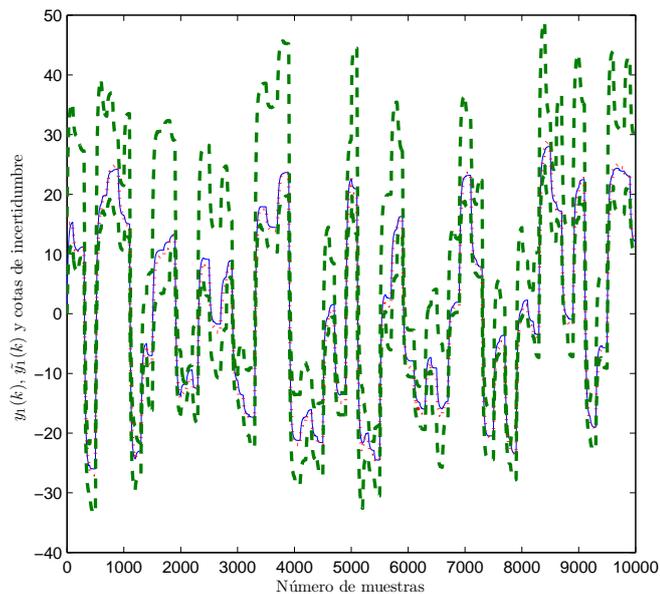


Figura 5.8: Cotas de incertidumbre para la salida  $y_1$  del modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $y_1$ : azul), modelo ( $\tilde{y}_1$ : rojo), cotas (verde).

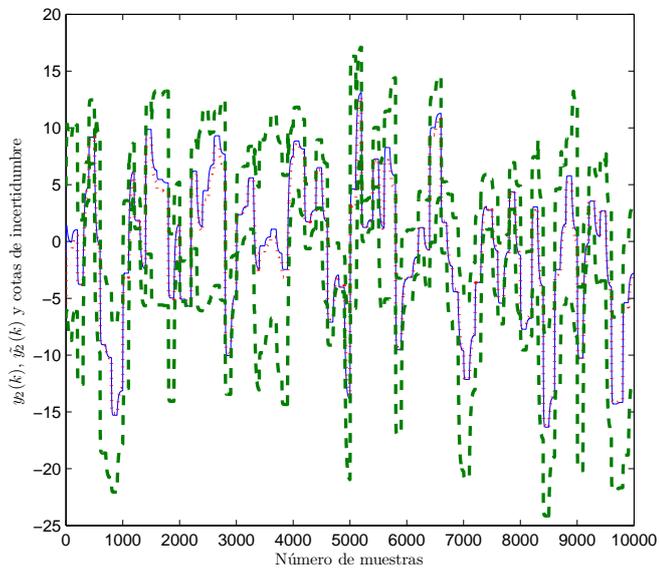


Figura 5.9: Cotas de incertidumbre para la salida  $y_2$  del modelo Laguerre-CLATAN sobre el conjunto de validación. Sistema ( $y_2$ : azul), modelo ( $\tilde{y}_2$ : rojo), cotas (verde).

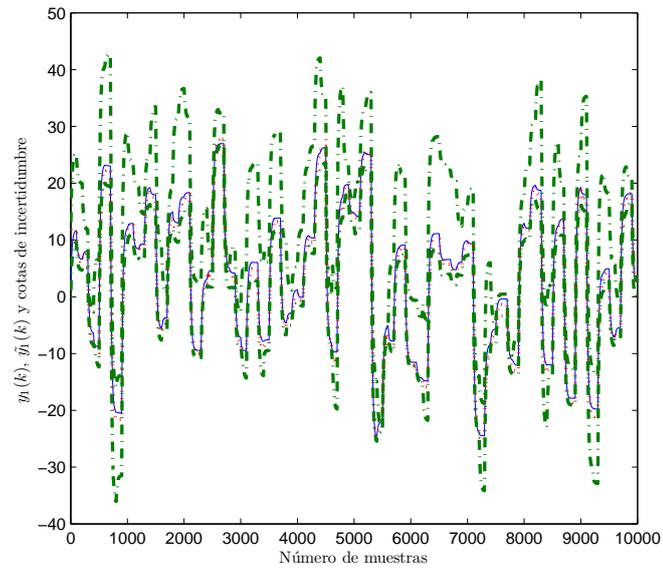


Figura 5.10: Cotas de incertidumbre para el modelo Laguerre-CLATAN sobre un nuevo conjunto de datos. Sistema ( $\mathbf{y}_1$ : azul), modelo ( $\tilde{\mathbf{y}}_1$ : rojo), cotas (verde).

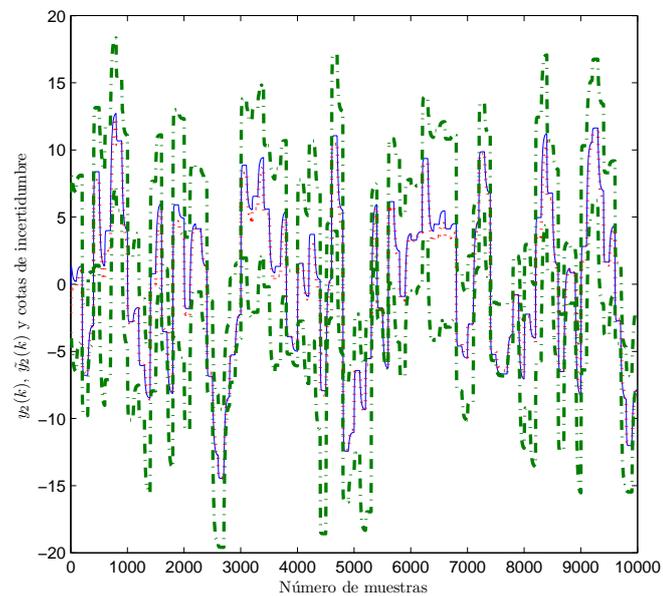


Figura 5.11: Cotas de incertidumbre para el modelo Laguerre-CLATAN sobre un nuevo conjunto de datos. Sistema ( $\mathbf{y}_2$ : azul), modelo ( $\tilde{\mathbf{y}}_2$ : rojo), cotas (verde).

y lo extendemos a sistemas MIMO. El modelo nominal está dado a través de una estructura de tipo Wiener, Laguerre-CLATAN o bien Kautz-CLATAN, que son un tipo especial de modelos orientados a bloques. La parte lineal es un sistema LTI, que puede ser representado utilizando bases racionales ortonormales, y la no linealidad estática está dada por una representación paramétrica de funciones base. En este caso, utilizamos un número finito de filtros de Laguerre o de Kautz para describir la dinámica y funciones CLATAN para describir la no linealidad estática y su incertidumbre. Establecimos la estimación de las cotas de los parámetros como un problema de optimización. La metodología resultante es robusta dado que el conjunto de los parámetros identificados es tal que la totalidad de los datos E/S del proceso se puede reproducir por al menos uno de los modelos en el conjunto. Los resultados muestran la viabilidad de la propuesta técnica de modelado. Realizamos diferentes pruebas de validación y la evolución dinámica del sistema siempre fue dentro de estos límites de incertidumbre. Es importante señalar que las funciones CLATAN que utilizamos tienen una implementación electrónica simple, por lo que sería posible desarrollar una aplicación en tiempo real de esta estructura de modelado para una gran clase de sistemas. La metodología se extiende a los sistemas MIMO.



# Aportes y Trabajo Futuro

En esta tesis hemos presentado dos estructuras de identificación para sistemas dinámicos no lineales. La primera de ellas, denominada NOE-CLATAN, fue implementada usando un algoritmo de identificación basado en funciones lineales a tramos. El objetivo de este algoritmo es lograr la identificación y simulación de un modelo no lineal para aplicaciones que requieren capacidad de procesamiento rápido y gran cantidad de datos. La metodología propuesta permite una evaluación gradual de un modelo NOE, a partir de un modelo lineal OE. Probamos, además, que la estabilidad BIBO del algoritmo está garantizada si la aproximación lineal inicial OE es BIBO estable. Esta propiedad permite hallar un buen modelo sin sobreajuste.

Por otra parte, son conocidos los problemas de evaluar el orden y el costo computacional de actualizar los parámetros para las estructuras de modelado NOE debido a su naturaleza recursiva. El algoritmo de identificación que proponemos en esta tesis ofrece un mecanismo muy simple para aumentar el orden del modelo (y, por lo tanto su capacidad de aproximación) reteniendo la aproximación inmediata anterior. El aumento en el orden del modelo se basa en el refinamiento de la grilla donde están definidas las funciones CLATAN, obteniendo modelos no lineales cada vez más complejos. Es importante señalar que este tipo de modelos está orientado a aplicaciones que necesitan un horizonte de simulación amplio como, por ejemplo, el MPC (por sus siglas en inglés, Model Predictive Control). Un ejemplo de aplicación en este campo puede consultarse en [16].

Además, el modelo NOE-CLATAN que propusimos presenta interesantes propiedades de generalización en el sentido que es posible lograr una buena aproximación en el conjunto de validación, aún en regiones donde no existen datos disponibles durante el proceso de identificación. Por último, si bien el modelo fue definido para sistemas SISO, puede ser generalizado fácilmente a sistemas MISO y MIMO.

El segundo esquema de identificación que presentamos es un modelo orientado a bloques compuesto por un bloque lineal dinámico representado por sistemas de funciones base ortonormales Laguerre o Kautz, y un bloque no lineal estático para el cual hemos utilizado funciones CLATAN. El ajuste de los parámetros de estos esquemas lo realizamos en el contexto de teoría de estimación con conjuntos de membresía para lo cual sólo necesitamos conocer un cota del ruido en alguna norma dada. La utilización de esta técnica dentro de las

estructuras de modelado tipo Wiener, nos permitió el desarrollo de esquemas de modelado de incertidumbre relativamente simples de calcular.

Probamos, además, que este tipo de estructura permite aproximar la evolución dinámica de cualquier sistema no lineal, causal, invariante en el tiempo, con memoria evanescente. La construcción del modelo, en cuanto a la elección de los elementos que componen la parte lineal dinámica, se basa en los datos de entrada/salida disponibles y en cierto conocimiento *a priori* del sistema.

Para estos esquemas, otra de las ventajas de utilizar bases de funciones CLATAN reside en que es posible detectar con exactitud qué símplices contribuyen a la aproximación así como aquellos que aún no han sido identificados.

Es nuestra intención, en un trabajo futuro, tener en cuenta los siguientes aspectos.

- Un inconveniente del modelo NOE-CLATAN es que el número de parámetros crece exponencialmente cuando el número de divisiones de la grilla aumenta. Por ello, proponemos analizar la utilización de grillas adaptivas, *i.e.* que puedan ser refinadas de diferente manera según el grado de no linealidad del sistema, en la definición de las funciones lineales a tramos de alto nivel.
- En el modelo NOE-CLATAN proponemos un algoritmo que permite ajustar fácilmente los parámetros de las funciones lineales a tramos cuando se refina la grilla respectiva. Un problema que proponemos es la formulación de un algoritmo para realizar el proceso inverso, es decir obtener una aproximación sobre una grilla más gruesa a partir de la aproximación obtenida sobre una grilla fina.
- Colaborar en la implementación electrónica del modelo NOE-CLATAN, lo cual permitiría obtener un hardware orientado a la identificación.
- Dijimos que el modelo NOE-CLATAN que propusimos presenta interesantes propiedades de generalización en el sentido que es posible lograr una buena aproximación en el conjunto de validación, aún en regiones donde no existen datos disponibles durante el proceso de identificación. Continuar con el análisis de estas propiedades constituye una línea de trabajo futura.
- Basándonos en el método de los coeficientes Lipschitz, estimamos el orden de las bases de funciones ortonormales necesario para la identificación de la parte lineal dinámica del sistema. Este criterio puede conducir a una estructura sobredimensionada en cuanto a la cantidad de funciones de la base a utilizar. Ahora bien, cuando el número de filtros es grande, la combinación de éstos con funciones lineales a tramos presenta limitaciones computacionales; es necesario entonces desarrollar alguna estrategia de selección más rigurosa del número de funciones de transferencia.
- Continuaremos estudiando la aproximación robusta de peor caso en los modelos cuando utilizamos funciones CLATAN para aproximar la parte no lineal estática del sistema.

Un inconveniente que se nos presenta es el costo computacional cuando el número de funciones de transferencia utilizadas en la parte lineal es grande debido al incremento de problemas de programación lineal a resolver. Este hecho también nos motiva a desarrollar estrategias para reducir en forma eficiente el número de sistemas lineales a utilizar sin perder calidad en el modelo.

- La posibilidad de emplear otras funciones base admisibles en el modelado del bloque lineal dinámico, para ampliar el espectro de sistemas que podamos modelar con nuestras estructuras de identificación.
- La aplicación de Teoría SM en la evaluación del dominio de generalización de modelos de sistemas dinámicos ([93]). Una vez obtenido el modelo ajustando los parámetros de la estructura elegida, otro de los objetivos es lograr un buen comportamiento del mismo ante entradas no utilizadas en la identificación del sistema, *i.e.* que dicho modelo responda adecuadamente a un conjunto de situaciones diferentes del sistema real. Con esta finalidad deseamos generar, utilizando teoría SM, un método que nos permita obtener una descripción de la zona de validez del modelo sobre su espacio de entrada, donde el error de aproximación se mantenga acotado, para entradas no utilizadas en la identificación del sistema. Esta zona de validez de la estructura sobre el espacio de las entradas se define como el dominio de generalización del modelo.
- El estudio de las propiedades de robustez de la estructura ante variaciones en los parámetros y su vinculación con los aspectos numéricos de la identificación, es decir el sobreajuste de los datos y el nivel de ruido presente en los mismos.



# Bibliografía

- [1] M.P. Álvarez, L.R. Castro, and O.E. Agamennoni. Set-membership estimation theory for Wiener modelling using HLCPWL functions. *IJMIC*, 14(1/2):13–26, 2011.
- [2] C.R. Arnold and K.S. Narendra. The characterization and identification of systems. *Technical Report 471, Harvard University, New York*, 1965.
- [3] K.J. Åström and P. Eykhoff. System identification-a survey. *Automatica*, 7:123–162, 1971.
- [4] J.F. Barrett. The use of functionals in the analysis of nonlinear physical systems. *Int. J. Electron. Control*, 5:567–615, 1963.
- [5] G. Belforte and P. Gay. Discrete time Hammerstein model identification with unknown but bounded errors. *IEE Proc. Control Theory Applications*, 148:523–529, 2001.
- [6] S.I. Biagiola and J.L. Figueroa. Feasible parameter set approach for uncertain Wiener models identification. In *Anales de la XIII Reunión en Procesamiento, Información y Control, RPIC'09*, pages 494–500, Rosario, Argentina, 2009.
- [7] S.I. Biagiola and J.L. Figueroa. Wiener and Hammerstein uncertain models identification. *Mathematics and Computers in Simulation*, 79:3296–3313, 2009.
- [8] S.A. Billings. Identification of nonlinear systems-a survey. volume 6, pages 272–284, 1980.
- [9] S.A. Billings, H. Jamaluddin, and S. Chen. Properties of neural networks with applications to modelling non-linear dynamical systems. *Int. J. Control*, 56:319–346, 1992.
- [10] S. Boyd and L.O. Chua. Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Trans. on Cir. and Syst.*, CAS-32:1150–1161, 1985.
- [11] C. Bruni. Analysis of approximation of linear and time-invariant systems pulse response by means of Laguerre finite term expansion. *IEEE Trans. on Automatic Control*, 9:580–581, 1964.

- [12] M. Casini. *Set-Membership Estimation: an Advanced Tool for Systems Identification*. Tesis Dottorato in Ingegneria dell'Informazione, Università degli Studi di Siena, Siena, Italy, 2004.
- [13] L. Castro, O. Agamennoni, and C. D'Attellis. Rational wavelets in Wiener-like modelling. *Mathematical and Computer Modelling*, 35:991–1006, 2002.
- [14] L. Castro, O. Agamennoni, and C. D'Attellis. Identification structure using rational wavelets: Examples of application. *Appl. Numerical Mathematics*, 47:345–363, 2003.
- [15] L. Castro, P. Julián, O. Agamennoni, and A. Desages. Wiener modelling using canonical piecewise linear functions. *Latin American Applied Research*, 29:265–272, 1999.
- [16] L.R. Castro, O.E. Agamennoni, and J.L. Figueroa. A control strategy using cpwlnoe structure. In *Proc. del IFAC International Symposium on Advanced Control of Chemical Process, ADCHEM 2006*, pages 335–340, Gramado, Brasil, 2 al 5 de abril de 2006, 2006. Ed. Elsevier, Oxford, UK.
- [17] L.R. Castro, O.E. Agamennoni, and M.P. Álvarez. From linear to nonlinear identification: One step at a time. *Mathematical and Computer Modelling*, 51(5):473–486, 2010.
- [18] L.R. Castro, J.L. Figueroa, and O.E. Agamennoni. BIBO stability for NOE model structure using HL CPWL functions. In *Proc. of the 24th IASTED Int. Conf., MIC 2005*, pages 91–96, Univ. California, Berkeley, 2005.
- [19] L.R. Castro, M.P. Álvarez, and O.E. Agamennoni. A NOE model identification algorithm using piecewise linear functions. In *In Proc. of EUROSIM 2007, Ljubljana, Slovenia, EUROSIM 2007*, 2007.
- [20] A.L. Cervantes, O.E. Agamennoni, and J.L. Figueroa. A nonlinear model predictive control system based on Wiener piecewise linear models. *Journal of Process Control*, 13(7):655–666, 2003.
- [21] S. Chen and S. Billings. Representation of non-linear systems: the NARMAX model. *Int. J. Control*, 49(3):1013–1032, 1989.
- [22] S. Chen and S. Billings. Neural networks for nonlinear dynamic system modelling and identification. *Int. J. Control*, 56:319–346, 1992.
- [23] S. Chen, S. Billings, and P. Grant. Non-linear system identification using neural networks. *Int. J. Control*, 51:1191–1214, 1990.
- [24] M. Chien and E. Kuh. Solving nonlinear resistive networks using piecewise-linear analysis and simplicial subdivision. *IEEE Trans. on Circ. and Syst.*, CAS-24:305–317, 1977.

- [25] L. Chisci, A. Garulli, A. Vicino, and G. Zappa. Block recursive parallelotopic bounding in set membership identification. *Automatica*, 34(1):15–22, 1998.
- [26] L.O. Chua and C. Kahlert. A generalized canonical piecewise-linear representation. *IEEE Trans. on Circ. and Syst.*, CAS-37:373–383, 1990.
- [27] L.O. Chua and S.M. Kang. Section-wise piecewise-linear functions: canonical representation, properties and applications. volume 65, pages 915–929, 1977.
- [28] J. Dieudonné. *Foundations of Modern Analysis*. Academic Press, 1960.
- [29] T. Oliveira e Silva. On the determination of the optimal pole position of Laguerre filters. *IEEE Trans. on Signal Process.*, 43(9):2079–2087, 1995.
- [30] P. Eykhoff. *System Identification, Parameter and State Estimation, 1st ed.* John Wiley and Sons, New York, 1974.
- [31] P. Eykhoff, P.M. Van der Grinten, H. Kwakernaak, and B.P. Veltman. Systems modelling and identification. In *Proc. 3rd IFAC Congress London*, London, 1966.
- [32] C. Fantuzzi, S. Simani, S. Beghelli, and R. Rovatti. Identification of piecewise affine models in noisy environment. *Int. J. of Control*, 75(18):1472–1485, 2002.
- [33] M. Di Federico, V. Jiménez-Fernández, P. Julián, and O. Agamennoni. An integrated circuit realization for a piecewise linear function. In *Proc. of the XII Reunión en Procesamiento, Información y Control (RPIC'07)*, Río Gallegos, Argentina, 2007.
- [34] Rui J.P. de Figueiredo and G. Chen. *Nonlinear Feedback Control Systems: An Operator Theory Approach*. Academic Press, USA, 1993.
- [35] E. Fogel and F. Huang. On the value of information in system identification-bounded noise case. *Automatica*, 1:229–238, 1982.
- [36] M. Fréchet. Sur les fonctionnels continus. *Ann. Ec. Norm. Sup*, 27:193–219, 1910.
- [37] Y. Fu and G. Dumont. An optimum time scale for discrete Laguerre network. *IEEE Trans. on Automatic Control*, 38(6):934–938, 1993.
- [38] D. George. Continuous nonlinear systems. *MIT RLE Tech. Report 355*, 1959.
- [39] G.J. Glowes. Choice of the time-scaling factor for linear system approximations using orthonormal Laguerre functions. *IEEE Transactions on Automatic Control*, 10:487–489, 1965.
- [40] R. Haber and H. Unbehauen. Structure identification of nonlinear dynamic systems - a survey on input/output approaches. *Automatica*, 26(4):651–677, 1990.

- [41] S. Haykin. *Neural Networks: a comprehensive Foundation*. Macmillan, New York, 1994.
- [42] X. He and H. Asada. A new method for identifying orders of input-output models for nonlinear dynamic systems. In *Proc. of the ACC*, pages 2520–2523, 1993.
- [43] M.A. Henson and D.E. Seborg. *Nonlinear Process Control, 1st ed.* Prentice Hall PTR, Upper Saddle River, N.J., 1997.
- [44] P.S.C. Heuberger, P.M.J. Van den Hof, and B. Wahlberg. *Modelling and identification with rational orthogonal basis functions*. Springer-Verlag, London, 2005.
- [45] V. Jiménez-Fernández, P. Julián, O. Agamennoni, M. Di Federico, L. Hernández-Martínez, and A. Sarmiento-Reyes. Finding operating points in networks containing mos transistors by a piecewise-linear approach. In *Proc. of the XII Reunión en Procesamiento, Información y Control (RPIC'07)*, Río Gallegos, Argentina, 2007.
- [46] T. Johansen and B. Foss. Constructing NARMAX models using ARMAX models. *Int. J. of Control*, 58:1125–1153, 1993.
- [47] P. Julián. *High Level Canonical Piecewise Linear Representation: Theory and Applications*. Tesis Doctoral en Control de Sistemas, DIEC-Universidad Nacional del Sur, Bahía Blanca, Buenos aires , Argentina, 1999.
- [48] P. Julián, A. Desages, and O. Agamennoni. High level canonical piecewise linear representation using a simplicial partition. *IEEE Transactions on Circuits and Systems - Part I*, 46:463–480, April 1999.
- [49] P. Julián, M. Jordán, and A. Desages. Canonical piecewise linear approximation of smooth functions. *IEEE Trans. on Circ. and Syst.*, 45:567–571, 1998.
- [50] P. M. Julián. *A High Level Canonical Piecewise Linear Representation: Theory and Applications*. PhD thesis, Universidad Nacional del Sur, Bahía Blanca, Argentina, UMI Dissertation Services, Michigan, USA, 1999.
- [51] P. M. Julián. *Una Representación Canónica Lineal a Tramos de alto Nivel: Teoría y Aplicaciones*. PhD thesis, Universidad Nacional del Sur, Bahía Blanca, Argentina, 1999.
- [52] S.M. Kang and L.O. Chua. A global representation of multidimensional piecewise-linear functions with linear partitions. *IEEE Trans. on Circ. and Syst.*, 25:938–940, 1978.
- [53] J.J. King and T. O'Canainn. Optimum pole positions for Laguerre functions models. *Electonic Letters*, 5:601–602, 1969.

- [54] M. Korenberg. Statistical identification of parallel cascades of linear and nonlinear systems. In *IFAC Proc. of Identification and System Parameter Estimation*, pages 669–674, 1982.
- [55] M.J. Korenberg and L.D. Paarmann. Orthogonal approaches to time-series analysis and system identification. *IEEE Signal Processing Magazine*, pages 29–43, 1991.
- [56] D.T. Lemma, M. Ramasamy, and M. Shuhaimi. System identification using orthonormal basis filters. *J. Applied Sci.*, 10:2516–2522, 2010.
- [57] J.N. Lin, H. Xu, and R. Unbehauen. A generalization of canonical piecewise-linear functions. *IEEE Trans. on Circ. and Syst.*, 41:345–347, 1994.
- [58] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, New Jersey, 2nd edition, 1999.
- [59] M.A. Masnadi-Shirazi and N. Ahmed. Optimum Laguerre networks for a class of discrete-time systems. *IEEE Trans. on Signal Process.*, 39:2104–2108, 1991.
- [60] M.B. Matthews and G.S. Moschytz. The identification of nonlinear discrete-time-fading memory systems using neural network models. *IEEE Trans. on Circ. and Syst. II*, 11:740–903, 1994.
- [61] M. Milanese. *Estimation theory and prediction in the presence of unknown but bounded uncertainty: A survey*. Robustness in identification and control. M. Milanese, R. Tempo and A. Vicino, Eds., Plenum Press, NY, 1989.
- [62] M. Milanese. Properties of least squares estimates in set membership identification. *Automatica*, 31(2):327–332, 1995.
- [63] M. Milanese and G. Belforte. Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors. linear families of models and estimators. volume AC-27, pages 408–414, 1982.
- [64] M. Milanese and C. Novara. Model quality in identification of nonlinear systems. *IEEE Transactions on Automatic Control*, 50(10):1606–1611, 2005.
- [65] M. Milanese and R. Tempo. Optimal algorithms theory for robust estimation and prediction. *IEEE Transactions on Automatic Control*, 30(8):730–738, 1985.
- [66] M. Milanese and A. Vicino. Optimal algorithms estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27:997–1009, 1991.
- [67] M. Milanese and A. Vicino. Information based complexity and non parametric worst-case system identification. *Journal of complexity*, 9:427–446, 1993.

- [68] K.S. Narendra and K. Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Trans. on Neural Networks*, 1:4–27, 1990.
- [69] O. Nelles. *Nonlinear System Identification*. Springer-Verlag, Berlin, Heidelberg, Germany, 2001.
- [70] J.P. Norton. Identification and application of bounded-parameter models. *Automatica*, 23:497–507, 1987.
- [71] G. Palm and T. Poggio. The Volterra representation and the Wiener expansion: Validity and pitfalls. *SIAM J. Appl. Math.*, 1:195–216, 1977.
- [72] T.W. Parks. Choice of time scale in Laguerre approximations using signal measurements. *IEEE Trans. on Automatic Control*, 16:511–513, 1971.
- [73] T. Poggio and F. Girosi. Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247:978–982, 1990.
- [74] K.S. Ray and D. Dutta Majumder. Simulation of a nonlinear steam generating unit. In *Proceedings of the international conference on systems, man and cybernetics*, volume 1, page 705, 1983.
- [75] J.A. Rodríguez, P. Julián, O. Lifschitz, O. Agamennoni, and V. Jiménez-Fernández. Vlsi microprocessor architecture for a simplicial pwl function evaluation core. In *EAM-TA 2008*, Buenos Aires, Argentina, 2008.
- [76] J.A. Rodríguez, O. Lifschitz, P. Julián, and O.E. Agamennoni. Verification and testing of a piecewise-linear asic microprocessor architecture. In *IBERCHIP Workshop, 2008*, 2008.
- [77] J.A. Rodríguez, O.D. Lifschitz, V.M. Jiménez-Fernández, P.M. Julián, and O.E. Agamennoni. Application-specific processor for piecewise linear functions computation. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(5):971–981, 2011.
- [78] W.L. Root. On system measure and identification. *System Theory*, pages 133–157.
- [79] M. Schetzen. Asymptotic optimum Laguerre series. *IEEE Trans. on Circ. Theory*, 18:493–500, 1971.
- [80] M. Schetzen. *The Volterra and Wiener theories of nonlinear systems*. John Wiley & Sons, 1980.
- [81] F.C. Scheppe. Recursive state estimation: unknown but bounded errors and system inputs. *IEEE Trans. on Automatic Control*, 13:22–288, 1968.
- [82] F.C. Scheppe. *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1973.

- [83] G. Sentoni, O. Agamennoni, A. Desages, and J. Romagnoli. Approximate models for nonlinear process control. *AIChE Journal*, 8:2240–2250, 1996.
- [84] G. B. Sentoni. *LNNet: Una Estructura de Aproximación para Sistemas No Lineales Discretos*. PhD thesis, Depto. de Ing. Eléctrica, Universidad Nacional del Sur, Bahía Blanca, Argentina, 1997.
- [85] J. Sjöberg and L.S.H. Ngia. *Neural Nets and Related Model Structures for Nonlinear System Identification*, pages 1–28. Nonlinear Modeling: Advanced Black-Box Techniques. J.A.K. Suykens and J. Vandevallé, Eds., Kluwer Academic Publishers, 1998.
- [86] J. Sjöberg, Q. Zhang, L. Ljung, A. Beneviste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 35(12):1691–1724, 1995.
- [87] I. Skrjanc, S. Blazic, and O.E. Agamennoni. Interval fuzzy modeling applied to pwl-Wiener models with uncertainties. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 35(5):1092–1095, 2005.
- [88] T. Söderström and P.G. Stoica. *System Identification*. Hemel Hempstead, UK: Prentice Hall, 1989.
- [89] E. Sontag. Nonlinear regulation: The piecewise linear approach. *IEEE Trans. on Automat. Control*, AC-26:346–357, 1981.
- [90] E. Sontag. Worst-case identification of nonlinear fading memory systems. *Automatica*, 31, 1995.
- [91] P. Van den Hof, P. Heuberger, and J. Bokor. Identification with generalized orthonormal basis functions-statistical analysis and error bounds. In *10th IFAC Symp. on System Identification*, volume 3, pages 3207–3212, Copenhagen, Denmark, 1994.
- [92] M. A. Abrahantes Vázquez. *Dominio de Generalización y Error de Aproximación en Estructuras Neuronales usadas en el Modelado de Sistemas*. PhD thesis, Depto. de Ing. Eléctrica, Universidad Nacional del Sur, Bahía Blanca, Argentina, 2000.
- [93] M. Abrahantes Vázquez and O. Agamennoni. Approximate models for nonlinear dynamical systems and its generalization properties. *Mathematical and Computer Modeling*, 33, 8-9:965–986, 2001.
- [94] A. Vicino and G. Zappa. Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Transactions on Automatic Control*, 41:774–785, 1996.
- [95] V. Volterra. *Theory of Functionals and of Integral and Integro-Differential Equations*. New York: Dover, 1959.

- [96] B. Wahlberg. System identification using Laguerre models. *IEEE Trans. on Automat. Control*, AC-36:551–562, 1991.
- [97] B. Wahlberg. System identification using Kautz models. *IEEE Trans. on Automat. Control*, AC-39(6):1276–1282, 1994.
- [98] B. Wahlberg and P. M. Mäkilä. On approximation of stable linear dynamic systems using Laguerre and Kautz functions. *Automatica*, 32(5):693–708, 1996.
- [99] L. Wang and W.R. Cluett. Optimal choice of time scaling factor for linear system approximations using Laguerre models. *IEEE Trans. on Automat. Control*, 39:1463–1467, 1994.
- [100] G.A. Watson. The numerical solution of total  $l_p$  approximation problems. In David Griffiths, editor, *Numerical Analysis*, Lecture Notes in Mathematics. Springer Berlin, Heidelberg.
- [101] N. Wiener. *The Theory of Prediction, Modern Math for Engineers*. Bechenbach: Mc. Graw Hill, 1956.
- [102] N. Wiener. *Nonlinear Problems in Random Theory*. MIT Press, Cambridge, MA, 1958.
- [103] H.S. Witsenhausen. Sets of possible states of linear systems given perturbed observations. *IEEE Trans. on Automatic Control*, 13:556–558, 1968.
- [104] A. Yazdizadeh and K. Khorasani. Adaptive time delay neural network structures for nonlinear system identification. *Neurocomputing*, 47:207–240, 2002.
- [105] G. Yen and Seok beom Lee. Multiple model approach by orthonormal bases for controller design. In *Proc. of the American Control Conference*, pages 2321–2325, Chicago, Illinois, 2000.
- [106] P. Young. Parameter estimation for continuous-time systems - a survey. *Automatica*, 17(1):23–39, 1981.
- [107] L.A. Zadeh. From circuit theory to system theory. In *Proc. IRE 50*, pages 856–865, 1962.

# Apéndice A

## Algoritmo de Actualización para el Modelo NOE-CLAT

En este Apéndice formulamos el algoritmo para hallar el vector de parámetros  $\mathbf{c}^{d,*}$  a partir del vector  $\mathbf{c}^{d-1,*}$ . Este algoritmo podría modificarse para computar  $\mathbf{c}^{d-1,*}$  a partir de  $\mathbf{c}^{d,*}$ . Observemos que calcular  $\mathbf{c}^{d,*}$  utilizando  $\mathbf{c}^{d-1,*}$ , es equivalente a obtener la representación CLATAN de una función definida sobre una partición simplicial  $H$  de un conjunto compacto  $\mathbf{S}$  con  $n_{div} = 2^d$  divisiones, a partir de su representación sobre la partición simplicial  $H$  del conjunto compacto  $\mathbf{S}$  con  $n_{div} = 2^{d-1}$  divisiones. Matemáticamente, eso significa resolver el siguiente sistema de ecuaciones lineales

$$\mathbf{c}^{d,*} \Lambda^d (V^d) = \mathbf{c}^{d-1,*} \Lambda^{d-1} (V^d). \quad (\text{A.1})$$

Para obtener la solución de la nueva representación dada por la ecuación (A.1), es necesario obtenerla para diferentes niveles de anidamiento (ver [48, 50]).

### Notación

$d$ : nivel de anidamiento correspondiendo a la base  $\Lambda^d$ .

$\mathbf{c}^{d,*}$ : vector final de parámetros asociado con aproximaciones CLATAN usando las bases  $\Lambda^d$ .

$n_{sec} = 2^{2d}$ : número de sectores correspondientes a  $n_{div} = 2^d$  divisiones de la región  $\mathbf{S}$  (se asume el mismo número de divisiones en cada dimensión).

$n_v$ : número de variables involucradas.

**Algoritmo**

$d = 0$  : Nivel de anidamiento 0.

$\mathbf{c}^{d,*}$  es la solución de la aproximación lineal como se describió en la Sección 3.2.

$d = 1$  : Nivel de anidamiento 1.

$$n_{com} = \binom{n_v}{1} = n_v,$$

$$i_v = i_n = 0,$$

$$n^{(1)} = 2n_{sec}.$$

```
{for  $i = 1$  to  $n_{com}$ 
  {for  $k_1 = 1$  to  $n_{sec}$ 
     $c_{i_n+2(k_1-1)+1}^{d,*} = c_{i_v+k_1}^{d-1,*}$ ,
  }
end}
 $i_v = i_v + n_{sec}$ ,
 $i_n = i_n + n^{(1)}$ .
end}
```

$d \geq 2$  : Nivel de anidamiento  $d$ .

$$n_{com} = \binom{n_v}{d},$$

$$i_v = i_n = 0,$$

$$n^{(2)} = 2^2 n_{sec},$$

$$n^{(d)} = 2n_{sec}^{d-1} \text{ for } d \geq 3.$$

```
{for  $i = 1$  to  $n_{com}$ 
  {for  $k_d = d$  to  $n_{sec}$ 
    {for  $k_{d-1} = d - 1$  to  $n_{sec}$ 
      :
      {for  $k_1 = 1$  to  $n_{sec}$ 
         $c_{i_n+2(k_1-1)+1}^{d,*} = c_{i_v+k_1}^{d-1,*}$ ,
      }
    }
     $i_v = i_v + n_{sec}$ ,
     $i_n = i_n + n^{(2)}$ ,
  }
  end}
  :
 $i_n = i_n + n^{(d-1)}$ .
end}
end}
```

## Apéndice B

# Método de los Coeficientes de Lipschitz

Las estructuras Laguerre-CLATAN y Kautz-CLATAN representan a operadores con memoria evanescente que, como vimos en la Subsección 4.4.1, son una generalización de los operadores continuos. Estos operadores transforman de manera continua las señales de entrada en señales de salida mediante la combinación no lineal de los sistemas de Laguerre o de Kautz mediante funciones CLATAN. La función CLATAN produce una transformación multivariable y continua sobre una región dada, de las señales provistas por ambos sistemas. Entonces

$$\tilde{y}(k+1) = f(z_0[u(k)], z_1[u(k)], \dots, z_m[u(k)]), \quad (\text{B.1})$$

donde  $z_i[u(k)]$ ,  $i = 0, \dots, m$  son las salidas de las funciones de transferencia discretizadas Laguerre o Kautz y  $f = f_{lat}$  representa una función en  $PWL_H[\mathbf{S}]$ .

La técnica para determinar la cantidad de sistemas se basa en la posibilidad de evaluar como varía la continuidad del operador Laguerre-CLATAN o Kautz-CLATAN de acuerdo al número  $m$  de sistemas utilizados. Por simplicidad notamos a las estructuras Laguerre-CLATAN y Kautz-CLATAN como

$$\tilde{y}(k+1) = g(\mathbf{z}_k) = g(z_0(k), z_1(k), \dots, z_m(k)), \quad (\text{B.2})$$

donde  $g(\cdot)$  representa a la función CLATAN y  $\mathbf{z}_k$  es un vector fila de  $[1 \times m]$

$$\mathbf{z}_k = [z_0(k), z_1(k), \dots, z_m(k)], \quad (\text{B.3})$$

donde cada componente  $z_j(k) = L_j u(k)$  o  $z_j(k) = K_j u(k)$  representa a la señal de entrada  $u(k)$  filtrada por el  $j$ -ésimo sistema de Laguerre o de Kautz. Sea  $\mathbf{Z}$  la matriz de orden  $[np \times m]$  en la cual son agrupados los  $np$  datos disponibles

$$\mathbf{Z}_k = [\mathbf{z}_0(k), \mathbf{z}_1(k), \dots, \mathbf{z}_{np}(k)]^T, \quad (\text{B.4})$$

donde cada  $\mathbf{z}_k$  está compuesto por el vector (B.3) para  $k = 1, \dots, np$ . Una vez determinados los valores de los polos del conjunto de sistemas elegido, es posible calcular a las señales  $z_j(k)$ . La cantidad de sistemas utilizados se refleja en la dimensión del dominio de la función  $g(\cdot)$ ,  $g: \mathbb{R}^m \rightarrow \mathbb{R}$ . Entonces el problema de estimar la cantidad de sistemas a utilizar puede ser visto como el de estimar

la dimensión para la cual la función desconocida  $g$  es continua. Es decir, qué puntos “cercaños” del espacio de entradas que toma valores de la matriz  $\mathbf{Z}$ , se transforman por la función  $g(\cdot)$  en puntos “cercaños” en el espacio de salida. Para ello utilizamos un método que se basa en los denominados coeficientes de Lipschitz.

Una función continua definida sobre una región cerrada es Lipschitz acotada. Basados en este resultado Xiangdong y Asada [42] presentaron el método de los coeficientes de Lipschitz para evaluar la dimensión del dominio de la función multivariable  $g$  en modelos no lineales E/S.

**Definición B.1** Se dice que que la función  $g : \mathbb{R}^m \rightarrow \mathbb{R}$  es Lipschitz acotada en una bola centrada en  $\mathbf{z}_0$ ,  $\mathbf{B} = \{\mathbf{z} \in \mathbb{R}^m : \|\mathbf{z} - \mathbf{z}_0\| \leq r, r > 0\}$  si existe una constante finita  $k_{lip}$ , tal que

$$|g(\mathbf{z}_i) - g(\mathbf{z}_j)| \leq k_{lip} \|\mathbf{z}_i - \mathbf{z}_j\|, \forall \mathbf{z}_i, \mathbf{z}_j \in \mathbf{B}, \quad (\text{B.5})$$

donde  $k_{lip}$  es la constante de Lipschitz y  $\|\cdot\|$  es la norma dada en el espacio de entradas.

Las funciones CLATAN son funciones Lipschitz acotadas, entonces la función  $g$  tiene sus derivadas parciales acotadas

$$|g_j| \leq |\partial g / \partial z_j| \leq k_{lip}, j = 0, \dots, m. \quad (\text{B.6})$$

Luego, los coeficientes de Lipschitz están definidos como

$$\lambda_{ij} = \frac{|y(i) - y(j)|}{\|\mathbf{z}_i - \mathbf{z}_j\|}, i \neq j,$$

donde  $y(i) = g(\mathbf{z}_i)$ , para  $i, j = 1, \dots, np$  por (B.2). Como la función  $g$  es continua, la condición de Lipschitz establece que los coeficientes de Lipschitz son acotados

$$0 \leq \lambda_{ij} < \infty.$$

Con los resultados anteriores y haciendo un análisis de sensibilidad a la relación E/S de la ecuación (B.2) se obtiene que

$$\delta y = g_1 \delta z_1 + g_2 \delta z_2 + \dots + g_m \delta z_m,$$

donde  $\delta y = g(\mathbf{z}_i) - g(\mathbf{z}_j)$  y las  $g_i$  son las derivadas parciales de  $g(\mathbf{z}_i)$  de (B.2) respecto de cada componente  $i$ . Luego

$$\lambda_{ij} = \frac{|\delta y|}{\sqrt{(\delta z_1)^2 + (\delta z_2)^2 + \dots + (\delta z_m)^2}} \quad (\text{B.7})$$

$$= \frac{|\delta z_1 + \delta z_2 + \dots + \delta z_m|}{\sqrt{(\delta z_1)^2 + (\delta z_2)^2 + \dots + (\delta z_m)^2}}. \quad (\text{B.8})$$

De acuerdo con las definiciones previas, tomando  $\delta = \max_{1 \leq i \leq n} |\delta_i|$  y  $|g_i| \leq k_{lip}$ , la cota para los coeficientes de Lipschitz puede obtenerse utilizando la desigualdad de Schwartz

$$\lambda_{ij}^m = k_{lip} \frac{|\delta z_1| + |\delta z_2| + \dots + |\delta z_m|}{\sqrt{(\delta z_1)^2 + (\delta z_2)^2 + \dots + (\delta z_m)^2}} \leq k_{lip} \sqrt{m}, \quad (\text{B.9})$$

donde  $m$  representa la cantidad de sistemas a ser usados. Sea  $m_0$  la cantidad deseada de esos sistemas, entonces si todas las variables  $z_i$ ,  $i = 1, \dots, m_0$  se encuentran en la reconstrucción de la

función  $g(\mathbf{z})$  los cocientes de Lipschitz para todos los pares de datos  $(\mathbf{z}_i, y_i)$  estarán acotados. Se pueden dar dos situaciones posibles bajo estos supuestos, si alguna de las variables no se encuentra presente en la reconstrucción los cocientes de Lipschitz serán arbitrariamente grandes; mientras que si existen más variables que las necesarias, entonces dichos coeficientes serán acotados. Analizamos estos dos casos a continuación.

*Si  $m < m_0$ , es decir el número utilizado de sistemas  $m$  es menos al número deseado  $m_0$ .*

Veamos que si alguna de las variables no se encuentra presenta en la reconstrucción, entonces los coeficientes de Lipschitz son arbitrariamente grandes.

Supongamos, sin perder generalidad, que la variable  $z_{m_0}$  no está incluida en la reconstrucción de la función desconocida  $g$ . Supongamos además que para algún par  $i, j$  de los datos disponibles se verifica que  $\delta z_1 = \dots = \delta z_{m_0-1}$  y que  $\delta z_{m_0} = \epsilon$  es una cantidad infinitesimal. Evaluamos ahora los cocientes de Lipschitz que con  $z_{m_0}$  ausente son arbitrariamente grandes y con  $z_{m_0}$  presente serán acotado. La distancia en el espacio de salida es, en ambos casos, la misma

$$\delta y = y_i - y_j,$$

ya que es un dato. La distancia en el espacio de entradas con  $z_{m_0}$  ausente será

$$\sqrt{(\delta z_1)^2 + (\delta z_2)^2 + \dots + (\delta z_m)^2}.$$

En estas condiciones el coeficiente de Lipschitz  $\lambda_{ij}^{m_0-1}$  es

$$\lambda_{ij}^{m_0-1} = \frac{|\delta y|}{\sqrt{(\delta z_1)^2 + (\delta z_2)^2 + \dots + (\delta z_m)^2}} \quad (\text{B.10})$$

$$= \frac{|g_{m_0}| \epsilon}{\sqrt{(\delta z_1)^2 + (\delta z_2)^2 + \dots + (\delta z_m)^2}} \rightarrow \infty. \quad (\text{B.11})$$

Sin embargo, si incluimos a la variable  $z_{m_0}$  en la reconstrucción, el coeficiente  $\lambda_{ij}^{m_0-1}$  es

$$\lambda_{ij}^{m_0-1} = |g_{m_0}|.$$

Resulta fácil ver que  $\lambda_{ij}^{m_0-1}$ , si  $z_{m_0}$  no está incluida, será mucho mayor que  $\lambda_{ij}^{m_0}$ , en el que todas las variables han sido incluidas.

*Si  $m > m_0$ , es decir el número utilizado de sistemas  $m$  es mayor al número deseado  $m_0$ .*

Cuando una variable redundante  $z_{m_0+1}$  ha sido incluida, del análisis anterior resulta fácil ver que  $\lambda_{ij}^{m_0+1}$  será sólo ligeramente menor que  $\lambda_{ij}^{m_0}$ , pero no significativamente.

La discusión anterior sienta las bases para que el número de entradas pueda ser identificado usando la información provista por los cocientes de Lipschitz. Procediendo de acuerdo a este objetivo definimos el siguiente índice

$$\lambda^m = \left( \prod_{k=1}^{m_v} \sqrt{m} \lambda^m(k) \right)^{1/m_v}, \quad (\text{B.12})$$

donde  $m_v = Mnp$ , siendo  $M$  una constante en el intervalo  $[0,01, 0,02]$ ,  $\lambda^m(k)$  el  $k$ -ésimo mayor coeficiente de Lipschitz calculado con las variables  $(z_1, z_2, \dots, z_m)$  entre todos los  $\lambda_{ij}^{m_0}$ ,  $i \neq j$ ,

$i, j = 1, 2, \dots, np$ , de la ecuación (B.9). El índice en (B.12) se define como el promedio geométrico de los  $m_v$  mayores valores de los  $\lambda_{ij}^m$  para disminuir el efecto producido el ruido que pudiera estar presente. Del análisis previo se puede apreciar que si  $\lambda^{m_0}$  es el número deseado de variables, entonces  $\lambda^{m_0+1}$  es muy próximo a  $\lambda_{ij}^{m_0}$ . Sin embargo,  $\lambda_{ij}^{m_0-1}$  es mucho mayor que  $\lambda_{ij}^{m_0}$ , y  $\lambda_{ij}^{m_0-2}$  es mucho mayor que  $\lambda_{ij}^{m_0-1}$ . Se puede ver que si se selecciona un  $m$  menor que  $m_0$ ,  $\lambda^m$  se incrementará significativamente. En cambio, si se selecciona un  $m > m_0$  no se obtiene una disminución significativa en el valor de  $\lambda^m$ . Luego, en el proceso de selección es conveniente tomar una constante de corte  $\gamma$ , tal que  $m_0$  pueda seleccionarse como

$$m_0 = \min_m [\lambda^m / \lambda^{m+1} < \gamma].$$

Lo expuesto puede resumirse en el siguiente algoritmo.

1. Comenzar con  $m = 1$ .
2. Verificar la continuidad de la función  $g(\cdot)$  para la totalidad del conjunto de  $np$  datos disponibles mediante los coeficientes de Lipschitz. Es decir, dada una constante de corte  $\gamma$ , verificar que los coeficientes entraron en la zona indicada por  $m_0 = \min_m [\lambda^m / \lambda^{m+1} < \gamma]$ .
3. Si la función  $g$  es continua, entonces  $m$  es el valor deseado y TERMINAR.

En caso contrario, incrementar  $m$  y volver al paso 1.

Hemos presentado una manera de estimar la cantidad de sistemas a utilizar en base a un concepto relacionado con la continuidad de los operadores, Laguerre-CLATAN y Kautz-CLATAN.

# Apéndice C

## Demostraciones Complementarias del Capítulo 4

Para sistemas de Kautz con dos parámetros vale la siguiente propiedad.

**Proposición C.1** El conjunto de sistemas de Kautz con dos parámetros es exponencialmente estable.

**Demostración:** Los sistemas de Kautz con dos parámetros se pueden expresar en el campo complejo en términos de su transformada  $\mathcal{Z}$  como  $K_j(z) = \sum_{i=0}^{\infty} k_j(i)z^{-i}$ , donde  $\beta, \bar{\beta}$  son los polos complejos conjugados en la región  $|z| < 1$ . Esto implica que la serie es absolutamente convergente para  $|z| > |\beta|$ . En particular, si elegimos  $z = z_1$  tal que  $|z_1| = r < 1$ ,  $|\beta| < r$ , obtenemos que  $\sum_{i=0}^{\infty} |k_j(i)z_1^{-i}| < \infty$ . Entonces, existe una constante finita  $M$  tal que  $|k_j(i)z_1^{-i}| < M$  y

$$|k_j(i)| < Mr^i. \quad (\text{C.1})$$

Con esta elección de  $r$ , existe  $q \in \mathbb{R}$ ,  $q > 0$  tal que  $|\beta| < r < e^{-q} < 1$ . En consecuencia, de (C.1),  $|k_j(i)| \rightarrow 0$  cuando  $i \rightarrow \infty$ , así como  $(e^{-qi})$ .  $\square$

**Lema C.2** El conjunto  $\mathcal{G}$  de funcionales  $G_j$  asociadas a los operadores discretos de Kautz  $K_j(z) = \sum_{i=0}^{\infty} k_j(i)z^{-i}$ , definidas como  $G_j u = \sum_{i=0}^{\infty} k_j(i)u(-i)$ , son continuas con la  $w$ -norma  $\|\cdot\|_w$ .

**Demostración:** Consideremos el conjunto de funcionales  $\mathcal{G} \doteq \{G_1, G_2, \dots\}$ , donde

$$G_j u = \sum_{i=0}^{\infty} k_j(i)u(-i)$$

es la funcional asociada al operador de Kautz  $K_j(z)$ .

Diremos que una función de ponderación o de pesos  $w'(i)$  domina a una función  $w(i)$  si  $w'(i) \geq w(i), \forall i$ . Renombramos la función de ponderación  $w'(i)$  como  $w(i)$ .

Sabemos que si un sistema tiene memoria evanescente con respecto a una sucesión de pesos, entonces también tiene memoria evanescente con respecto a cualquier otra sucesión de pesos que la domine [60]. Sin pérdida de generalidad, podemos tomar la sucesión de pesos  $w(i) = 1/(1+i)$ , entonces

$$\begin{aligned} |G_j u - G_j v| &= \left| \sum_{i=0}^{\infty} k_j(i) u(-i) - \sum_{i=0}^{\infty} k_j(i) v(-i) \right| \\ &\leq \sup_{i \geq 0} (|u(-i) - v(-i)| w(i)) \sum_{i=0}^{\infty} |k_j(i)| w(i)^{-1} \\ &\leq \|u - v\|_w \sum_{i=0}^{\infty} |k_j(i)| w(i)^{-1}. \end{aligned} \quad (\text{C.2})$$

Sabemos que los sistemas de Kautz son exponencialmente estables, luego por (C.1)

$$\sum_{i=0}^{\infty} |k_j(i)| w(i)^{-1} \leq M \sum_{i=0}^{\infty} r^i (i+1), \quad r < 1. \quad (\text{C.3})$$

Ahora, para cualquier  $N \in \mathbb{N}$ ,  $N < \infty$  tenemos que

$$\begin{aligned} \sum_{i=0}^N r^i (i+1) &= \sum_{i=0}^N r^i i + \frac{1 - r^{N+1}}{1 - r} \\ &= r \frac{d}{dr} \sum_{i=0}^N r^i + \frac{1 - r^{N+1}}{1 - r} \\ &= \frac{1 - r^{N+1}(N+2) + r^{N+2}(N+1)}{(1-r)^2} \\ &\leq \frac{1}{(1-r)^2} + \frac{r^{N+2}(N+1)}{(1-r)^2} \end{aligned}$$

Tomando límite para  $N \rightarrow \infty$  en ambas desigualdades obtenemos

$$\sum_{i=0}^{\infty} r^i (i+1) \leq \frac{1}{(1-r)^2}. \quad (\text{C.4})$$

Por lo tanto, por (C.3) y (C.4)

$$|G_j u - G_j v| \leq M \|u - v\|_w \frac{1}{(1-r)^2}.$$

Entonces dado  $\varepsilon > 0$ , teniendo  $\|u - v\|_w < \delta$  con  $\delta = \varepsilon(1-r)^2/M$  concluimos que

$$|G_j u - G_j v| < \varepsilon,$$

lo que completa la demostración.  $\square$

**Lema C.3** El conjunto  $\mathcal{G}$  de funcionales definidas en Lema 4.4.9 separa puntos en  $l^\infty(\mathbb{Z}_-)$ .

**Demostración:** Supongamos que  $u_1, u_2 \in l^\infty(\mathbb{Z}_-)$  y  $G_j u_1 = G_j u_2, \forall j$ . Sea  $u = u_1 - u_2$ , de modo que  $G_j u = 0, \forall j$ . Si probamos que  $u = 0$ , entonces el conjunto  $\mathcal{G}$  separa puntos en  $l^\infty(\mathbb{Z}_-)$ .

La funcional  $G_j$  puede escribirse como

$$G_j u = \sum_{i=0}^{\infty} k_j(i) |\beta|^{-i/2} |\beta|^{i/2} u(-i) = 0, \quad (\text{C.5})$$

donde  $\beta$  es uno de los polos generadores de los filtros de Kautz con dos parámetros.

las sucesiones  $(k_j(i) |\beta|^{-i/2})$  y  $(|\beta|^{i/2} u(-i))$  pertenecen al espacio  $l^2(\mathbb{Z}_+)$ . En efecto, como  $|\beta| < 1$  tenemos

$$\sum_{i=0}^{\infty} |\beta|^{i/2} u(-i)|^2 \leq \|u\|_\infty^2 \sum_{i=0}^{\infty} |\beta|^i < \infty. \quad (\text{C.6})$$

Además,

$$\sum_{i=0}^{\infty} |k_j(i) \beta^{-i/2}|^2 < \infty \quad (\text{C.7})$$

pues, usando el criterio de D'Alambert y teniendo en cuenta que  $|k_j(i)| < M e^{-qi}$  y  $|\beta| < 1$ , obtenemos

$$\lim_{i \rightarrow \infty} \frac{|\beta|^{-(i+1)} e^{-2(i+1)q}}{|\beta|^{-i} e^{-2iq}} = \frac{e^{-2q}}{|\beta|} < 1. \quad (\text{C.8})$$

El conjunto de funciones de Kautz es completo en  $l^2(\mathbb{Z}_+)$ , entonces  $\{|\beta|^{-i/2} k_j(i)\}$  es completo en  $l^2(\mathbb{Z}_+)$ . Por lo tanto, de (C.5) resulta que  $|\beta|^{i/2} u(-i) = 0, \forall i$ , y en consecuencia  $u = 0$ . Esto prueba que el conjunto de funcionales discretas de Kautz separa puntos en  $l^\infty(\mathbb{Z}_-)$ .  $\square$

