Identificación de Sistemas No Lineales

Juan Carlos Gómez
Identificación de Sistemas
Departamento de Electrónica, FCEIA
Universidad Nacional de Rosario

Contenido

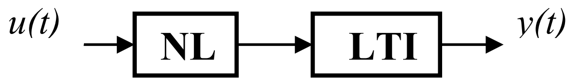
Introducción

- La mayoría de los sistemas tienen un comportamiento no lineal, excepto en un determinado rango de operación donde pueden ser considerados lineales.
- Modelos Lineales aproximan al sistema no lineal alrededor de un punto de operación.
- La performance del modelo lineal (i.e., sus características predictivas) se ven deterioradas al variar el punto de operación del sistema no lineal.
- Para describir globalmente el comportamiento del sistema se debe recurrir a Modelos No Lineales.

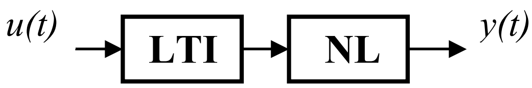
- La gran variedad de modelos no lineales hace que no sea posible obtener métodos generales de identificación, sino sólo para determinadas clases de modelos no lineales.
- Muchos sistemas no lineales pueden ser representados por la interconexión de sistemas lineales estacionarios y no linealidades estáticas. Estos modelos se denominan **orientados a bloques** (block-oriented nonlinear models). Las no linealidades estáticas aparecen por ejemplo debido a saturación de actuadores, sensores con características no lineales, etc.
- De entre los modelos orientados a bloques, los que han sido más estudiados son los **Modelos Hammerstein** y los **Modelos Wiener**.

Modelos No Lineales

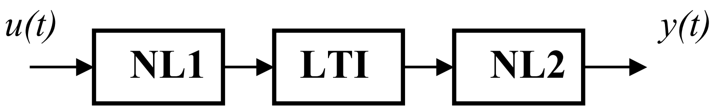
- **Modelos Hammerstein**



- **Modelo Wiener**



- **Modelo Hammerstein-Wiener**



- **Modelos orientados a bloques:** Incluyen los anteriores más otras posibles conexiones (serie, paralelo y en retroalimentación) de bloques LTI y no linealidades estáticas.

- **Modelo regresivo lineal, con regresor como función no lineal de los datos pasados**

$$\hat{y}(t,\theta) = \theta_1 \varphi_1\left(u^t, y^{t-1}\right) + \theta_2 \varphi_2\left(u^t, y^{t-1}\right) + \cdots + \theta_d \varphi_d\left(u^t, y^{t-1}\right) = \varphi^T(t)\theta$$

φ_i funciones no lineales arbitrarias de los datos pasados

Problema → Cómo elegir las funciones φ_i

Posibles soluciones

- Expansión tipo **caja negra.** Por ejemplo: $\varphi_i\left(u^t, y^{t-1}\right)$ polinomios en las entradas y salidas pasadas.

- Uso de leyes fundamentales (física, química, etc.) para detectar las no linealidades en el sistema.

- ## Modelos No Lineales en Espacio de Estados

$$x(t+1) = f(t, x(t), u(t), \omega(t), \theta)$$
$$y(t) = h(t, x(t), u(t), \upsilon(t), \theta)$$

donde
$\omega(t), \upsilon(t)$ son perturbaciones (procesos aleatorios)
θ vector de parámetros (desconocido)

NB: El problema de hallar un predictor óptimo para este modelo no lineal estocástico es muy dificil (no existe solución finito dimensional).

- ## Modelos No Lineales Tipo Caja Negra

Predictores de la forma:

$$\hat{y}(t \mid \theta) = g(Z^{t-1}, \theta) \qquad (1)$$

donde
Z^{t-1}: datos pasados
$Z^t = (y^t, u^t) = (y(1), u(1), \cdots, y(t), u(t))$

$\boxed{\text{Problema}} \longrightarrow$ Cómo elegir $g(\bullet, \bullet)$

$\boxed{\text{Posibles Soluciones}}$

- Uso de regresores

$$g(Z^{t-1}, \theta) = g(\varphi(t), \theta)$$

con

$$\varphi(t) = \varphi(Z^{t-1}) \qquad \text{vector de regresión}$$

- Un modelo más general sería permitiendo que el regresor dependa del parámetro

$$\varphi(t, \theta) = \varphi(Z^{t-1}, \theta)$$

- Con este enfoque, el problema de elegir $g(Z^{t-1},\theta)$ en (1) se descompone en dos problemas:
  1. Cómo elegir el vector de regresión $\varphi(t)$ como función de los datos pasados.
  2. Cómo elegir la función no lineal $g(\varphi,\theta)$.

- **Ejemplos de regresores**

  Por analogía con los modelos lineales se pueden obtener modelos NARX, NARMAX, NOE, NFIR.

  Los más comunes son modelos NARX y NFIR, en los cuales el regresor depende sólo de valores medidos (no estimados).

- **Ejemplos de funciones** $g(\varphi,\theta)$

  Típicamente se utiliza **expansión en funciones base**

$$g(\varphi,\theta)=\sum_{k=1}^{n}\alpha_k g_k(\varphi) \qquad (2)$$

$\theta=[\alpha_1 \quad \cdots \quad \alpha_n]^T$ vector de parámetros

$g_k(\bullet)$: funciones bases

Problema Clave → Cómo elegir las funciones base

Posibles Soluciones

- **Series de Volterra**

$$g_k(\varphi)=\varphi^k$$

donde $\varphi^k=\{\varphi_1^{k_1}\varphi_2^{k_2}\cdots\varphi_d^{k_d}\}$, con

$$k_1+k_2+\cdots k_d=k$$

- **Enfoques más modernos**

  Tienen las siguientes características:
  - Todas las g_k están formadas a partir de una **función base madre** $\kappa(x)$.
  - Esta función $\kappa(x)$ es función de una variable escalar x.
  - Típicamente las g_k son versiones escaladas (dilatadas) y trasladadas de $\kappa(x)$.

  Por ejemplo para $d=1$, se tendría

$$g_k(\varphi) = g_k(\varphi, \beta_k, \gamma_k) = \kappa(\beta_k(\varphi - \gamma_k))$$

  donde β_k son los parámetros de escalado, y γ_k son los parámetros de traslación.

<u>Ejemplos de 'funciones base madre'</u>

- Serie de Fourier (caso escalar)

$$\kappa(x) = \cos(x)$$

- Funciones seccionalmente constantes

$$\kappa(x) = \begin{cases} 1 & \text{si} \quad 0 \le x \le 1 \\ 0 & \quad \text{coc} \end{cases}$$

con $\gamma_k = k\Delta$, $\beta_k = 1/\Delta$, y $\alpha_k = f(k\Delta)$. Esto da una aproximación seccionalmente constante de cualquier función en un intervalo Δ.

- **Clasificación de Funciones Base**
  - **Bases Locales:** las variaciones significativas de la función se dan en un entorno.
  - **Bases Globales:** tienen una variación significativa en todo el eje real.

- **Funciones Base Multivariables**

  - **Producto Tensorial**

$$g_k(\varphi) = g_k(\varphi, \beta_k, \gamma_k) = \prod_{j=1}^{d} \kappa\left(\beta_k^j\left(\varphi_j - \gamma_k^j\right)\right)$$

  - **Construcción radial**

Las funciones dependen sólo de la distancia de φ a un dado punto central (centro).

$$g_k(\varphi) = g_k(\varphi, \beta_k, \gamma_k) = \kappa\left(\left\|\varphi - \gamma_k\right\|_{\beta_k}\right)$$

donde $\left\|\bullet\right\|_{\beta_k}$ denota una dada norma en el espacio de los regresores φ. Típicamente

$$\left\|\varphi\right\|_{\beta_k}^2 = \varphi^T \beta_k \varphi$$

donde β_k es una matriz de escalado definida positiva.

- **Construcción ridge**

Las funciones dependen sólo de la distancia de φ a un dado hiperplano

$$g_k(\varphi) = g_k(\varphi, \beta_k, \gamma_k) = \kappa\left(\beta_k^T \varphi + \gamma_k\right) \quad , \qquad \varphi \in \Re^{\mathrm{d}}$$

La función es constante para todos los φ que pertenecen al hiperplano

$$\left\{\varphi \in \Re^d : \beta_k^T \varphi = \text{const}\right\}$$

- **Calidad de la aproximación**

  El modelo usando expansión en funciones base resulta

$$g(\varphi,\theta)=\sum_{k=1}^{n}\alpha_k\kappa(\beta_k(\varphi-\gamma_k))$$

♦ Surge el interrogante de cuán bien esta expansión puede representar a cualquier posible sistema real '$g_0(\varphi)$'.

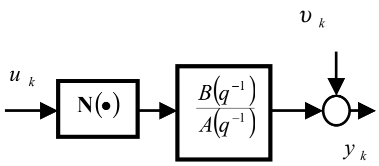
La respuesta es que (excepto para polinomios) cualquier función madre $\kappa(x)$ permite aproximar cualquier *razonable* $g_0(\varphi)$ arbitrariamente bien para un n suficientemente grande.

♦ Otro interrogante que surge es la *eficiencia* de la expansión, es decir cuán grande debe ser n para lograr un determinado grado de aproximación.

No hay una respuesta **general** para esto.

Identificación de Modelos Hammerstein

- **Modelos Hammerstein**
  - **Método no iterativo (Bai, 1998)**



$$\begin{aligned} y_k &= \frac{B(q^{-1})}{A(q^{-1})}\sum_{i=1}^{r}c_i g_i(u_k)+\upsilon_k \\ &= \frac{B(q^{-1})}{A(q^{-1})}\mathbf{N}(u_k)+\upsilon_k \end{aligned} \tag{1}$$

donde y_k, u_k y υ_k son la salida, entrada y perturbación en el instante k, respectivamente, $g_i(\bullet)$ son funciones no lineales usadas para describir la no linealidad estática $\mathbf{N}(\bullet)$, y

$$
\begin{aligned}
A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\
B(q^{-1}) &= b_1 q^{-1} + \cdots + b_m q^{-m}
\end{aligned} \quad ,
$$

con q^{-1} denotando el operador desplazamiento directo.

Se asume que los órdenes n, m, s, y las funciones no lineales $g_i(\bullet)$, son conocidas, y que el **objetivo** es estimar los parámetros de la parte lineal: $(a_1 \cdots a_n b_1 \cdots b_m)$, y de la parte no lineal: $(c_1 \cdots c_r)$, a partir de datos de entrada-salida.

Una elección típica de las funciones $g_i(\bullet)$ es $g_i(u_k) = u^i{}_k$, en cuyo caso, la parte no lineal queda representada por una expansión polinomial.

La ecuación (1) puede escribirse

$$
y_k = -\sum_{j=1}^{n} a_j y_{k-j} + \sum_{i=1}^{r} \sum_{j=1}^{m} b_j c_i u^i{}_{k-j} + \upsilon_k \tag{2}
$$

Definiendo los vectores

$$
\begin{aligned}
\theta &= (a_1, \cdots, a_n, b_1 c_1, \cdots, b_m c_1, \cdots, b_1 c_r, \cdots, b_m c_r)^T \\
\phi_k &= (-y_{k-1}, \cdots, -y_{k-n}, u_{k-1}, \cdots, u_{k-m}, \\
&\quad u^2{}_{k-1}, \cdots, u^2{}_{k-m}, \cdots, \\
&\quad u^r{}_{k-1}, \cdots, u^r{}_{k-m})^T
\end{aligned}
$$

la ecuación (2) puede escribirse

$$
y_k = \phi_k{}^T \theta + \upsilon_k \tag{3}
$$

Finalmente, considerando el conjunto de N datos $\{y_k, u_k\}_{k=1}^{N}$, la ecuación (3) puede escribirse

$$
Y_N = \Phi_N{}^T \theta + V_N \tag{4}
$$

donde

$$
\begin{aligned}
Y_N &= (y_1, \cdots, y_N)^T \quad , \qquad V_N = (\upsilon_1, \cdots, \upsilon_N)^T , \\
\Phi_N &= (\phi_1, \cdots, \phi_N)
\end{aligned}
$$

Luego, la estima de mínimos cuadrados $\hat{\theta}$, del vector de parámetros θ viene dada por

$$\hat{\theta} = \left(\Phi_N \Phi_N{}^T\right)^{-1} \Phi_N Y_N , \tag{5}$$

siempre que la inversa exista. Definiendo ahora los vectores

$$a = (a_1, \cdots, a_n)^T, b = (b_1, \cdots, b_m)^T, c = (c_1, \cdots, c_r)^T$$

y la matriz

$$\Theta_{bc} = \begin{pmatrix} b_1 c_1 & b_1 c_2 & \cdots & b_1 c_r \\ b_2 c_1 & b_2 c_2 & \cdots & b_2 c_r \\ \vdots & \vdots & \ddots & \vdots \\ b_m c_1 & b_m c_2 & \cdots & b_m c_r \end{pmatrix} = bc^T ,$$

el vector de parámetros θ puede escribirse

$$\theta = \left(a^T, \mathrm{vec}(\Theta_{\mathrm{bc}})^T\right)^T ,$$

donde $\mathrm{vec}(\Theta_{bc})$ es el vector columna obtenido apilando las columns de Θ_{bc}. Estimas de los vectores a y $\mathrm{vec}(\Theta_{bc})$ pueden entonces obtenerse de la estima $\hat{\theta}$ en (5). Denotemos $\hat{a}$ y $\widehat{\mathrm{vec}}(\Theta_{bc})$ a esas estimas.

El problema es como computar estimas de los vectores b y c, a partir de $\widehat{\mathrm{vec}}(\Theta_{bc})$.

Es claro que las más cercanas, en el sentido de la norma-2, estimas $\hat{b}$ y $\hat{c}$, son tales que minimizan la norma

$$\left\| \mathrm{vec}\left(\hat{b}\hat{c}^T\right) - \widehat{\mathrm{vec}}(\Theta_{bc}) \right\|_2^2 = \left\| \hat{b}\hat{c}^T - \hat{\Theta}_{bc} \right\|_F^2 ,$$

es decir

$$\left(\hat{b}, \hat{c}\right) = \underset{\mathrm{b,c}}{\mathrm{argmin}} \left\| \hat{\Theta}_{bc} - bc^T \right\|_F^2 ,$$

donde $\| \bullet \|_F$ es la norma de Frobenius.

La solución de este problema de minimización viene dada por la descomposición SVD de la matriz $\hat{\Theta}_{bc}$. El resultado está resumido en el siguiente Lema (Bai, 1998).

______ Sea $\hat{\Theta}_{bc} \in \Re^{m\times r}$ una matriz no nula, y sea $\hat{\Theta}_{bc} = USV^T$ su descomposición SVD, donde U y V son matrices ortogonales cuyas columnas son los vectores singulares izquierdos y derechos respectivamente,

$$U = (\mu_1, \mu_2, \cdots, \mu_m),\ V = (v_1, v_2, \cdots, v_r),$$

y donde S es una matriz diagonal con los valores singulares ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,r)} \geq 0$) en la diagonal. Entonces

$$\min_{b,c} \left\| \hat{\Theta}_{bc} - bc^T \right\|_F^2 = \sum_{i=2}^{\min(m,r)} \sigma_i^2,$$

$$(\mu_1, \sigma_1 v_1) = \underset{\mathrm{b,c}}{\operatorname{argmin}} \left\| \hat{\Theta}_{bc} - bc^T \right\|_F^2.$$

Prueba: Ver (Bai, 1998).

■

El **algoritmo** de identificación no lineal puede resumirse:

Paso 1. Computar la estima de mínimos cuadrados $\hat{\theta}$ en (5). Una estima $\hat{a}$ del vector de parametros a viene dada por las primeras n componentes de $\hat{\theta}$, y una estima $\hat{\Theta}_{bc}$ de la matriz Θ_{bc} puede construirse a partir de las últimas $m \times r$ componentes de $\hat{\theta}$.

Paso 2. Computar la SVD de $\hat{\Theta}_{bc}$ como en el Lema, de donde estimas de los parametros b y c pueden calcularse como

$$\hat{b} = \mu_1,$$

y

$$\hat{c} = \sigma_1 v_1,$$

respectivamente.