

Identificación usando Bases Ortonormales

Juan Carlos Gómez

Laboratory for System Dynamics and Signal Processing

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Universidad Nacional de Rosario, Argentina

E-mail: jcgomez@fceia.unr.edu.ar

Contenido

- Introducción
- Bases Ortonormales
- Identificación usando bases ortonormales
- Análisis asintótico de las estimas
- Ejemplos de Simulación
- Comentarios Finales

Introducción

- Las estructuras de tipo *caja negra* en general no permiten incorporar información 'a priori' sobre la dinámica del sistema.

Excepción → **Uso de Bases ortonormales**

- El uso de Bases Ortonormales permite disminuir el orden de los modelos manteniendo el grado de aproximación de los mismos.
- Permiten cuantificar los errores por dinámica no modelada, y debido a las perturbaciones. Es posible realizar un análisis asintótico de las estimas similar al existente para modelos FIR (Ljung & Yuan).
- Los modelos con Bases Ortonormales son aptos para estimación con mínimos cuadrados (algoritmos robustos, reducida carga computacional).

- Es posible extender los resultados a sistemas multivariados (Gómez, 1998).
- Es posible obtener realizaciones mínimas en espacio de estados de los sistemas representados con Bases Ortonormales.
- Es posible extender los métodos de identificación para algunas clases de sistemas no lineales (Gómez & Baeyens, 2000).
- Se ha desarrollado una teoría rica, con muchas aplicaciones.

Bases Ortonormales

- Espacios funcionales

Consideraremos espacios funcionales con estructura de **espacio de Hilbert** (espacios con producto interno que son completos con respecto a la métrica inducida por la norma inducida por el producto interno), asociados a las funciones transferencia racionales de sistemas en tiempo discreto, causales y estables.

$L_2(\mathbf{T})$: espacio de funciones cuadrado (Lebesgue) integrables en la circunferencia unitaria en el plano complejo $\mathbf{T} = \{z : |z| = 1\}$. Es un espacio de Hilbert cuando se lo equipa con el producto interno

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

producto interno

O equivalentemente

$$\langle F, G \rangle = \frac{1}{2\pi j} \oint_{\mathbf{T}} F(z) \overline{G\left(\frac{1}{z}\right)} \frac{dz}{z}$$

producto interno

$\mathbf{H}_2(\mathbf{T})$: el **espacio de Hardy** (con estructura de Hilbert) de funciones cuadrado (Lebesgue) integrables en la circunferencia unitaria, que son analíticas en el exterior del disco unitario $\mathbf{D} = \{z : |z| < 1\}$.

Es un subespacio propio de $L_2(\mathbf{T})$ con la misma definición de producto interno.

Este espacio puede pensarse como el *'espacio de las funciones transferencia racionales en tiempo discreto, causales y estables'*.

- **Bases Ortonormales**

$\{\mathbf{B}_k(q)\}_{k=0}^{\infty}$ es una **base ortonormal** del espacio de Hilbert $\mathbf{H}_2(\mathbf{T})$

si es un conjunto ortonormal **completo**, i.e. si verifica

$$\langle \mathbf{B}_k, \mathbf{B}_\ell \rangle = \delta_{k\ell}$$

con $\delta_{k\ell}$ la delta de Kronecker, y además

$$\text{span} \left(\{ \mathbf{B}_k(q) \}_{k=0}^{\infty} \right) = \mathbf{H}_2(\mathbf{T}) \quad (\text{genera todo el espacio})$$

Esto implica que cualquier función del espacio

$$G(q) \in \mathbf{H}_2(\mathbf{T})$$

tiene una representación única de la forma

$$G(q) = \sum_{k=0}^{\infty} \theta_k \mathbf{B}_k(q) \quad (1)$$

donde $\theta_k = \langle G, \mathbf{B}_k \rangle$.

Identificación usando Bases Ortonormales

La estructura de modelo es

$$y(n) = G(q)u(n) + v(n) \quad (2)$$

donde $v(n)$ es una perturbación, que se considerará un proceso aleatorio.

$G(q)$ es representada como una expansión en bases ortonormales de la forma (1). Obviamente, como $G(q)$ es desconocida, los coeficientes de la expansión no puede calcularse según $\theta_k = \langle G, \mathbf{B}_k \rangle$.

El objetivo es entonces estimar los parámetros de un modelo de dimensión finita

$$G(q) = \sum_{k=0}^{p-1} \theta_k \mathbf{B}_k(q) \quad (3)$$

de manera que una estima de $G(q)$ vendrá dada por

$$G(q, \hat{\theta}) = \sum_{k=0}^{p-1} \hat{\theta}_k \mathbf{B}_k(q) \quad (4)$$

donde $\hat{\theta}_k$ es una estima de θ_k .

La estructura de modelo (2),(4) puede escribirse en la forma de **regresor lineal**

$$y(n) = \varphi^T(n) \theta + v(n) \quad (5)$$

definiendo

$$\varphi^T(n) = [B_0(q)u(n), B_1(q)u(n), \dots, B_{p-1}(q)u(n)]$$

$$\theta = [\theta_0 \quad \theta_1 \quad \dots \quad \theta_{p-1}]^T$$

La estimación puede hacerse usando el **método de mínimos cuadrados**, que en este caso tiene una solución analítica. La estima de mínimos cuadrados (LSE) resulta

$$\hat{\theta}_N = \underset{\theta}{\operatorname{argmin}} \{V_N(\theta)\} = \left[\sum_{k=1}^N \varphi(k)\varphi^T(k) \right]^{-1} \left[\sum_{k=1}^N \varphi(k)y(k) \right] \quad (6)$$

donde $V_N(\theta)$ es un **criterio cuadrático** en los errores de predicción

$$V_N(\theta) = \frac{1}{N} \sum_{k=1}^N \operatorname{Tr} \left\{ [y(k) - \varphi^T(k)\theta] [y(k) - \varphi^T(k)\theta]^T \right\} \quad (7)$$

Algunas Bases Ortonormales en $H_2(\mathbb{T})$

- **Modelo FIR**

En este caso es

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

Es decir, tienen polos en el origen del plano Z .

Nota: 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.

- **Bases de Laguerre**

Están definidas como:

$$B_k(q) = \left(\frac{\sqrt{1-\xi^2}}{q-\xi} \right) \left(\frac{1-\xi q}{q-\xi} \right)^k \quad k \geq 0$$

donde $\xi \in \mathfrak{R}, |\xi| < 1$ es un parámetro real denominado coeficiente de Laguerre. Las bases tienen un único polo real con multiplicidad.

- **Bases de Kautz**

$$B_k(z) = \begin{cases} \frac{\sqrt{(1-a^2)(1-c^2)}}{z^2 - a(c+1)z + c} \left(\frac{cz^2 - a(c+1)z + 1}{z^2 - a(c+1)z + c} \right)^{\frac{k-1}{2}} & k \text{ impar} \\ \frac{\sqrt{(1-c^2)}(z-a)}{z^2 - a(c+1)z + c} \left(\frac{cz^2 - a(c+1)z + 1}{z^2 - a(c+1)z + c} \right)^{\frac{k}{2}} & k \text{ par} \end{cases}$$

donde $-1 < a < 1$, $-1 < c < 1$, $k \geq 0$. Notar que los polos son complejos conjugados, con multiplicidad. Las bases permiten incluir información *a priori* que uno pueda tener de la dinámica del sistema seleccionando adecuadamente los parámetros a y c .

- **Bases Generalizadas**

$$B_k(q) = \left(\frac{\sqrt{1 - |\xi_k|^2}}{q - \xi_k} \right) \prod_{i=0}^{k-1} \left(\frac{1 - \bar{\xi}_i q}{q - \xi_i} \right) \quad k \geq 0$$

Permiten la incorporación de polos arbitrarios sin la necesidad de repetición periódica. Las bases FIR, Laguerre, y Kautz son casos particulares de estas bases.

- **Bases Generadas por “inner functions”**

Definición: Una función transferencia racional $G(z)$ se denomina **inner** si es estable y satisface

$$G(z^{-1})G(z) = 1$$

Es decir, si es estable y **all pass**.

Sea $G_x(z)$ una inner function con grado de McMillan n_x , y con una realización mínima balanceada (A, B, C, D) . Denotemos

$$V_k(z) = z(zI - A)^{-1} B G_x^k(z)$$

Entonces la secuencia de funciones transferencia racionales

$$\{\mathbf{B}_{i,k}(z)\} = \{e_i^T V_k(z)\} \quad \begin{array}{l} i = 1, \dots, n_x \\ k = 0, \dots, \infty \end{array}$$

constituye una Base Ortonormal en $\mathbf{H}_2(\mathbf{T})$.

Análisis Asintótico de las Estimaciones

El error de estimación (en el dominio frecuencial) puede escribirse como

$$G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}) = \underbrace{G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta_*)}_{\text{variance}} + \underbrace{G(e^{j\omega}, \theta_*) - G(e^{j\omega})}_{\text{bias}}$$

donde θ_* es el valor al cual converge $\hat{\theta}$ cuando el número de datos $N \rightarrow \infty$. Pueden reconocerse dos términos:

- Error debido al ruido (**variance error**)

$$G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta_*)$$

que usualmente se cuantifica con la media (cuadrática) muestral

$$\mathbf{E} \left\{ \left| G(e^{j\omega}, \hat{\theta}) - G(e^{j\omega}, \theta_*) \right|^2 \right\}$$

Para evitar la influencia de error por dinámica no modelada se suele hacer un análisis asintótico donde ambos, el orden del modelo y el número de datos, se hacen tender a infinito.

- Error por dinámica no modelada (**bias error**, o **undermodelling error**), debido a que la estructura de modelo es muy simple para representar el sistema real

$$G(e^{j\omega}, \theta_*) - G(e^{j\omega})$$

Se suele cuantificar a través de una cota superior de su módulo.

- Error debido al ruido

Bases FIR

Un "**famoso**" resultado de Ljung establece que para N (número de datos) y p (orden del modelo) altos, entonces

$$\text{var} \left\{ G(e^{j\omega}, \hat{\theta}) \right\} \approx \frac{p}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

Para el caso multivariable (Ljung & Yuan)

$$\text{cov} \left\{ g(e^{j\omega}, \hat{\theta}) \right\} \approx \frac{p}{N} \Phi_u^{-1}(\omega) \otimes \Phi_v(\omega)$$

Bases Generalizadas

$$B_k(q) = \left(\frac{\sqrt{1 - |\xi_k|^2}}{q - \xi_k} \right) \prod_{i=0}^{k-1} \left(\frac{1 - \bar{\xi}_i q}{q - \xi_i} \right) \quad k \geq 0$$

Resultado "**no tan famoso**" de Ninness *et al.*, establece

$$\text{var} \left\{ G(e^{j\omega}, \hat{\theta}) \right\} \approx \frac{\gamma_p(\omega)}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

donde $\gamma_p(\omega) = \sum_{k=0}^{p-1} |B_k(e^{j\omega})|^2$ es el "**reproducing kernel**" de las bases.

Para el caso multivariable, un resultado "**aún menos famoso**" de Gómez (1998) establece

$$\text{cov} \left\{ g(e^{j\omega}, \hat{\theta}) \right\} \approx \frac{\gamma_p(\omega)}{N} \Phi_u^{-1}(\omega) \otimes \Phi_v(\omega)$$

Comentarios

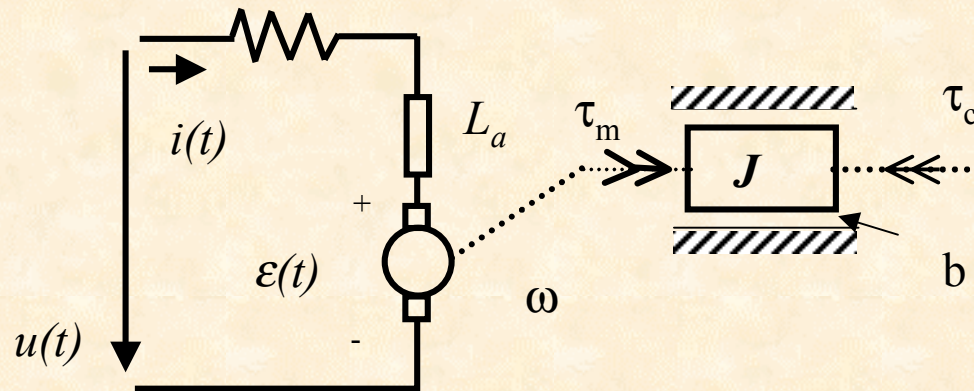
- Hay aquí un nuevo fenómeno (no presente en el caso FIR) que indica que el error depende de la ubicación de los polos de las bases, a través del factor $\gamma_p(\omega) = \sum_{k=0}^{p-1} |\mathbf{B}_k(e^{j\omega})|^2$, que es también función de la frecuencia.
- Para estas bases, estos resultados no pueden probarse a partir de los resultados de Ljung para el caso FIR, ya que las bases no presentan la propiedad de grupo

$$\mathbf{B}_k \mathbf{B}_\ell = \mathbf{B}_{k+\ell}$$

que presentan las bases FIR.

Ejemplo de Simulación

- **Proceso: Motor de Corriente Continua con excitación independiente constante (revisitado)**



Conversión Electromecánica

$$\varepsilon = k\omega$$

$$\tau_m = ki$$

- **Estructura de Modelo**

$$G(s) = \frac{\omega(s)}{U(s)} = \frac{K\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \longrightarrow G(z) = \text{ZOH}\{G(s)\} = \frac{b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}$$

□ El sistema “real”

$$L_a = 2.5 \text{ mHy} \quad R_a = 0.062 \ \Omega \quad k = 6.6 \text{ Nm/A}$$

$$b = 1.275 \text{ Nms} \quad J = 30 \text{ Nms}^2$$

$$G(s) = \frac{87.99}{s^2 + 1.3370s + 580.821}$$

Sistema Real (FT)



Equivalente ZOH ($T_s = 10^{-6}$ seg)

$$G(z) = \text{ZOH}\{G(s)\} = \frac{0.4399 \times 10^{-6} (z + 1)}{z^2 - 1.9999z + 0.9999}$$

□ Datos de Entrada-Salida

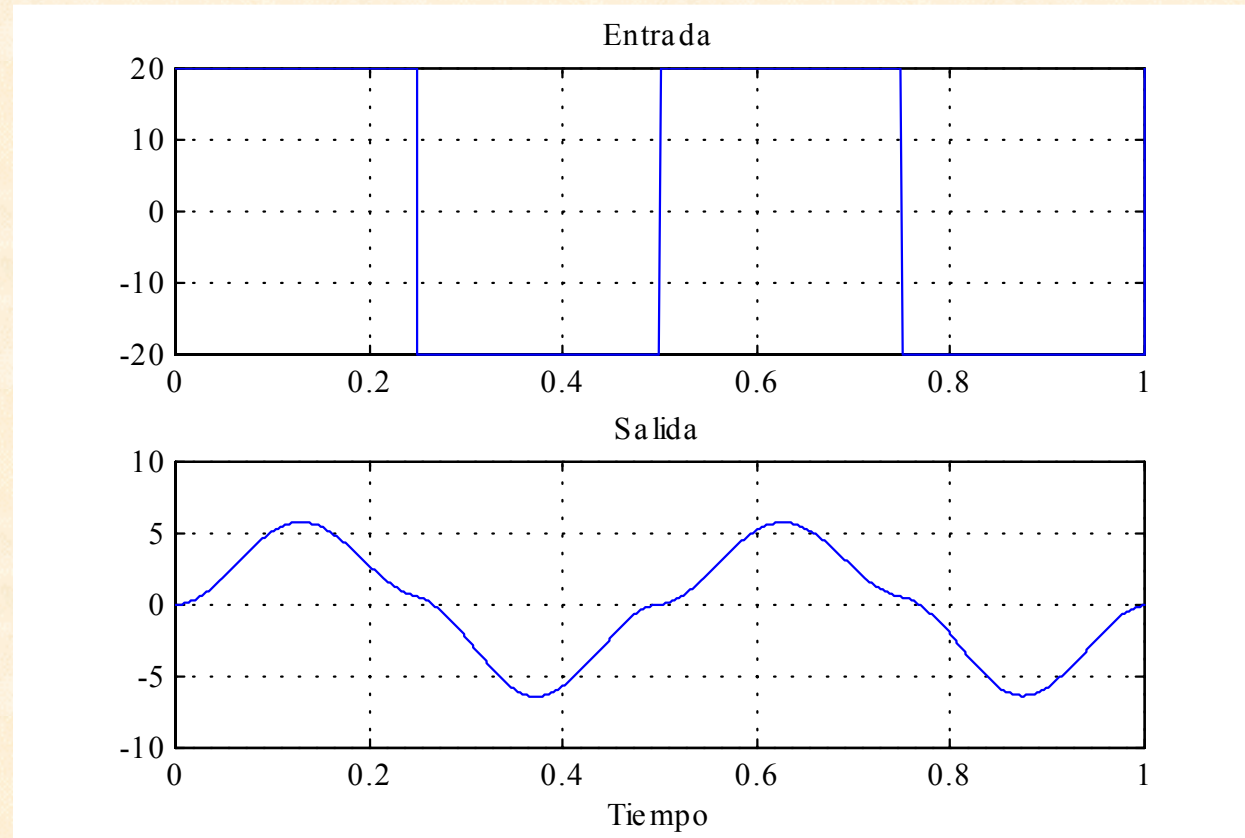


Figura 1: Datos de entrada-salida

□ Elección de las Bases

Se adopta una estructura de modelo con bases ortonormales de Kautz, y se consideran solamente dos términos en la expansión en series (2) (esto significa que identificaremos un modelo de segundo orden, con polos complejos conjugados). Los dos primeros elementos de las bases de Kautz resultan:

$$B_0(z) = \frac{\sqrt{1-c^2}(z-a)}{z^2 - a(c+1)z + c}$$

$$B_1(z) = \frac{\sqrt{(1-c^2)(1-a^2)}}{z^2 - a(c+1)z + c}$$

Se adoptan los siguientes valores de los parámetros a y c de las bases

$$a = 0.9997$$

$$c = 0.99997$$

Con estos valores, la estima de mínimos cuadrados del vector de parámetros resulta:

$$\hat{\theta}_N = \begin{bmatrix} -0.0275 \\ 0.4821 \end{bmatrix}$$

La figura (2) muestra la salida medida (en línea llena roja) y la correspondiente salida estimada (en línea de trazos azul).

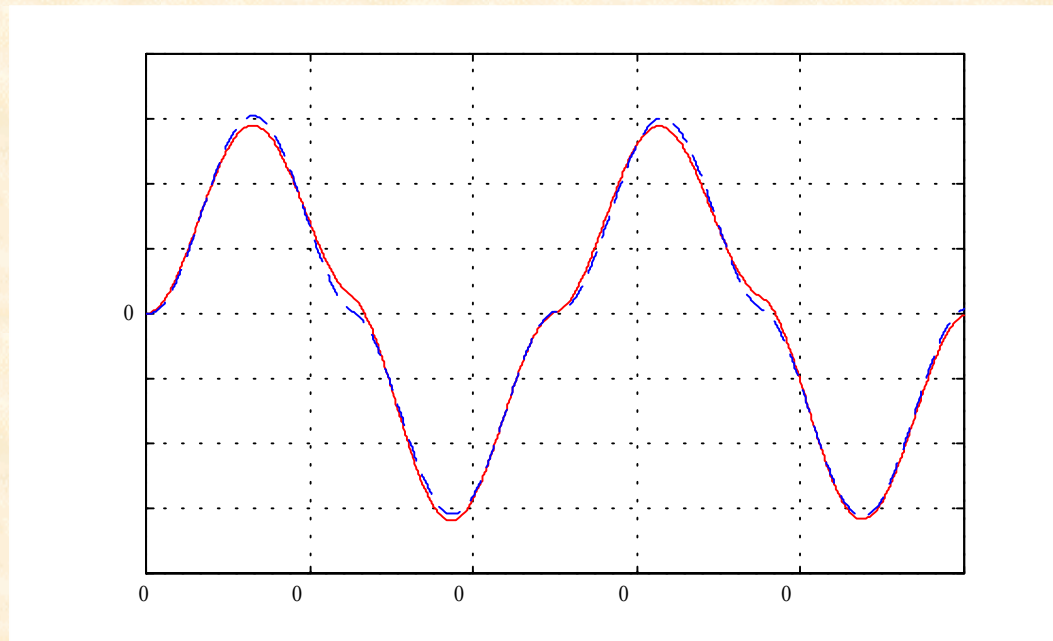


Figura 2: Salida medida (línea llena roja) y salida estimada (línea de trazos azul).

Comentarios Finales

Consideremos el caso de una estructura con polos fijos, de la forma

$$y(n) = G(q)u(n) + v(n) = \frac{C(q, \hat{\theta})}{D(q)}u(n) + v(n)$$

donde los polos de $D(q)$ son fijos y conocidos. Podemos escribir

$$y(n) = C(q, \hat{\theta}) \left(\frac{1}{D(q)}u(n) \right) + v(n) = C(q, \hat{\theta})\tilde{u}(n) + v(n)$$

y entonces interpretarlo como un problema de estimación de un modelo FIR con la entrada pre-filtrada por $1/D(q)$
Luego, podríamos concluir que

$$\text{var} \left\{ C(e^{j\omega}, \hat{\theta}) \right\} \approx \frac{p}{N} \frac{\Phi_v(\omega)}{\Phi_{\tilde{u}}(\omega)}$$

con $\Phi_{\tilde{u}}(\omega) = \frac{\Phi_u(\omega)}{|D(e^{j\omega})|^2}$. Es decir

$$\text{var} \{C(e^{j\omega}, \hat{\theta})\} \approx \frac{p |D(e^{j\omega})|^2}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

Por lo que podríamos concluir que

$$\text{var} \{G(e^{j\omega}, \hat{\theta})\} \approx \frac{1}{|D(e^{j\omega})|^2} \text{var} \{C(e^{j\omega}, \hat{\theta})\} \approx \frac{p}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

y entonces la varianza no dependería de la ubicación de los polos. Este razonamiento es sin embargo **erróneo**, ya que el filtro $D(e^{j\omega})$ está cambiando a medida que el orden p se incrementa, y por lo tanto el espectro $\Phi_{\tilde{u}}(\omega)$ también está cambiando con p , y estaríamos aplicando el resultado de Ljung para FIR que fue derivado asumiendo un espectro fijo de la entrada.