

Tema 2

Técnicas de identificación no paramétricas

2.1 Introducción

Como se ha comentado en la introducción, la identificación es un método de construcción de modelos basado en experimentos. El caso particular de un sistema lineal invariante en el tiempo puede ser descrito por su función de transferencia (*modelos paramétricos*) o por su correspondiente respuesta impulso o frecuencial (*modelos no paramétricos*). Se consideran modelos no paramétricos a aquellos en que no es posible definir un vector de parámetros finito para representarlo. Estas técnicas, la mayoría de ellas más fáciles o directas de aplicar o utilizar en un proceso real, aportan información muy útil para afrontar las decisiones que debe tomar el usuario ante un problema de *identificación* de un proceso, véase la figura 1.2. Como se expondrá en este tema, utilizando previamente algunas de dichas técnicas, será posible: considerar unos datos válidos o no, diseñar adecuadamente el experimento y, también, conocer algunas de las características de la estructura del modelo.

Son tres las técnicas de identificación no paramétricas presentadas, concretamente se evalúan las técnicas de análisis de la respuesta transitoria, el análisis de correlación entre señales y las técnicas frecuenciales. Previamente a la descripción de dichas técnicas (apartado 2.2), se realiza una introducción de características, propiedades y tipos de señales. Como apartado final, se describen herramientas para el diseño adecuado del experimento.

Comentar finalmente que hay métodos en la literatura [Shoukens91] que permiten parametrizar un modelo a partir de los resultados obtenidos con técnicas no paramétricas. Algunas de ellas se expondrán en el tema 3.

2.2 Introducción a las señales

2.2.1 Características generales de las señales. Definiciones

Previo a describir los tipos de señales de excitación, es necesario definir algunas características generales y propiedades de las señales.

Se define como media, \bar{u} , y covarianza, $R_u(\tau)$, de una señal estacionaria determinista, $u(t)$, a:

$$\begin{aligned}\bar{u} = \bar{m}_u &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t) \\ R_u(\tau) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N [u(t+\tau) - \bar{m}][u(t) - \bar{m}]^T\end{aligned}\tag{2.1}$$

donde N es el número de datos y τ es un entero entre 0 y N .

Al valor de la covarianza en $t=0$ se la denomina varianza, \mathbf{s}_u^2 , y se caracteriza por ser una función par:

$$R(-t) = R(t)$$

La energía total o RMS de una señal se calcula a partir de la ecuación (2.2).

$$RMS_u = \sqrt{\frac{1}{N} \sum_{t=1}^N u(t)^2} \quad (2.2)$$

Se define la transformada de Fourier discreta (DFT), $U_N(\mathbf{w})$, de esta señal, como:

$$U_N(\mathbf{w}) = \frac{1}{\sqrt{N}} \sum_{t=1}^N u(t) e^{-j\mathbf{w}t} \quad (2.3)$$

para valores de $\mathbf{w}=2\mathbf{p}k/N$, $k=1,\dots,N$. $U_N(\mathbf{w})$ se caracteriza por ser una función de periodo $2\mathbf{p}$, por lo tanto: $U_N(\mathbf{w}+2\mathbf{p})=U_N(\mathbf{w})$ y si $u(t)$ es real $U_N(-\mathbf{w})=conj(U_N(\mathbf{w}))$ (*conj* equivale al conjugado complejo). La señal $u(t)$ se puede representar por la inversa de DFT:

$$u(t) = \frac{1}{\sqrt{N}} \sum_{k=1}^N U_N(2\mathbf{p}k/N) e^{j2\mathbf{p}kt/N} \quad (2.4)$$

El valor de $U_N(2\mathbf{p}k/N)$ nos informa sobre el 'peso' que origina la descomposición de la señal $u(t)$ a la frecuencia $2\mathbf{p}k/N$ por los distintos valores de k . Su valor absoluto cuadrático $|U_N(2\mathbf{p}k/N)|^2$ es una medida de la contribución energética de esta frecuencia en la señal. La relación de Parseval:

$$\sum_{t=1}^N |U_N(2\pi k / N)|^2 = \sum_{t=1}^N u^2(t) \quad (2.5)$$

reafirma la interpretación que la energía de una señal puede descomponerse en la contribución energética de las diferentes frecuencias.

La densidad espectral, $S(\mathbf{w})$, de una señal $u(t)$ se define como la transformada de Fourier de la función covarianza:

$$S(\mathbf{w}) = \frac{1}{2\mathbf{p}} \sum_{t=-\infty}^{\infty} R(t) e^{-it\mathbf{w}} \quad (2.6)$$

y por ser la covarianza una función par:

$$S(\mathbf{w}) = \frac{1}{\mathbf{p}} \sum_{t=0}^{\infty} R(t) \cos(\mathbf{w}t) \quad (2.7)$$

Se puede establecer una relación entre el espectro y la transformada de Fourier discreta de una señal; esta viene dada por la ecuación (2.8).

$$S(\mathbf{w}) = \frac{1}{2\mathbf{p}N} |U_N(\mathbf{w})|^2 \quad (2.8)$$

En general, para realizar una buena identificación de parámetros es necesario aplicar una señal de entrada frecuencialmente rica. Para garantizar que los algoritmos de estimación converjan a una solución única, es necesario imponer unos requerimientos mínimos en las

señales de test. Esto se denomina condiciones *persistentemente excitadas* (*persistent excitation*).

Se dice que una variable discreta es persistentemente excitada de orden n si existe el límite:

$$R_u(t) = \lim_{N \rightarrow \infty} \sum_{t=1}^N u(t)u(t-t) \quad (2.9)$$

y la matriz:

$$A_n = \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & \cdots & R_{uu}(n-1) \\ R_{uu}(-1) & R_{uu}(0) & & \vdots \\ \vdots & & \ddots & \\ R_{uu}(1-n) & \cdots & & R_{uu}(0) \end{bmatrix} \quad (2.10)$$

es definida positiva [Åström71] y en consecuencia invertible. De aquí se deduce que la condición necesaria para realizar una estimación consistente de un modelo lineal de orden n es que la señal de test sea persistentemente excitada de orden $2n$.

Otra definición de señal persistentemente excitada se encuentra en [Söderström89]. Una señal $u(t)$ es persistentemente excitada (*p.e.*) de orden n si su densidad espectral no es nula en n valores.

- Si $u(t)$ es un escalón de amplitud s : $R_u(t) = s^2$ para todo t y la matriz de $R_u(n)$ es no singular para $n=1$, por lo tanto es una función *p.e.* de orden 1
- Si $u(t)$ es un impulso: $R_u(t) = 0$ para todo t y la matriz de $R_{uu}(n)=0$, no es *p.e.*

2.2.2 Propiedades de las señales periódicas

Las señales de excitación pueden ser periódicas o no periódicas, las características de las primeras se describen a continuación.

Si se considera que $u(t)$ es una señal determinista periódica con un periodo M , $u(t) = u(t-M)$ para todo t , el valor medio m y la función de covarianza $R_u(t)$, definidos en (2.1), tiene por expresión:

$$m = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N u(t) = \frac{1}{M} \sum_{t=1}^M u(t) \quad (2.11)$$

$$R_u(t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N [u(t+t) - m][u(t) - m]^T = \frac{1}{M} \sum_{t=1}^M [u(t+t) - m][u(t) - m]^T$$

para una señal periódica la función de covarianza vale:

$$R(M+t) = R(t) \quad (2.12)$$

La antitransformada de Fourier discreta de una señal periódica puede ser escrita, de acuerdo con la ecuación (2.4), como:

$$u(t) = \frac{1}{\sqrt{M}} \sum_{k=1}^M A_k e^{j2\pi kt/M} \quad (2.13)$$

siendo:

$$A_k = \frac{1}{\sqrt{M}} \sum_{t=1}^M u(t) e^{-j2\pi k t / M}$$

Si $u(t)$ es una señal periódica y contiene N muestras, con $N = kM$ donde k es el número de periodos, se deduce que la contribución energética a las distintas frecuencias vale:

$$|U_N(\omega)|^2 = \begin{cases} k |A_k|^2, & \text{si } \omega = \frac{2\pi k}{N}, \quad k = 0, \pm k, \pm 2k, \pm \dots, \pm \frac{N}{2} \\ 0, & \text{si } \omega = \frac{2\pi k}{N}, \quad k \neq 0, \pm k, \pm 2k, \pm \dots, \pm \frac{N}{2} \end{cases} \quad (2.14)$$

La función espectral de señales periódicas discretas se puede evaluar como un conjunto de impulsos Dirac de cada una de las frecuencias presentes en la señal [Söderström89]:

$$S(\omega) = \sum_{k=0}^{M-1} C_k \delta\left(\omega - 2\pi \frac{k}{M}\right) \quad (2.15)$$

donde C_k es la matriz de coeficientes a determinar en cada caso y pueden ser calculados a partir de:

$$C_k = \frac{1}{M} \sum_{g=0}^{M-1} \mathbf{a}^{-kg} r(g) \quad \text{con } \mathbf{a} = e^{j2\pi/M}$$

2.2.3 Descripción de algunas de las señales de excitación

Hay distintos tipos de señales de entrada, por ejemplo la función impulso, función escalón, ruido blanco o ‘coloreado’, señales sinusoidales, ruido binario pseudo-random (PRBS), etc. En este apartado se van a comentar algunas de las características y propiedades de dichas señales.

2.2.3.1 Ruido blanco y ‘coloreado’

El ruido blanco es una secuencia de variables random uniformemente distribuidas de media cero y una varianza determinada. Dado una señal $e(t)$,

$$\begin{aligned} \bar{e} &= E\{e(t)\} = \frac{1}{N} \sum_{t=1}^N e(t) = 0 \\ \text{var}\{e(t)\} &= R_e(0) = \frac{1}{N} \sum_{t=1}^N [e(t) - \bar{e}]^2 = I^2 \end{aligned} \quad (2.16)$$

Se caracteriza porque:

$$1/N \sum_{t=1}^N e(t)e(t+\mathbf{t}) \rightarrow 0 \quad \text{cuando } N \rightarrow \infty$$

La matriz (2.9) $R_u(n) = I^2 I_n$ (I_n matriz identidad de orden n), es por tanto una matriz definida positiva para todo n , indicando que el ruido blanco es una señal persistentemente excitada.

El ancho de banda de un ruido blanco es hipotéticamente infinito, cuando se trata de una señal discreta el ancho de banda está acotado por el número de muestras y el periodo de muestreo de dicha señal (2.3).

El ruido blanco filtrado se caracteriza por tener un ancho de banda finito. Se obtiene filtrando adecuadamente el ruido blanco.

$$u(t) + c_1 u(t-1) + \dots + c_m u(t-m) = e(t) + d_1 e(t-1) + \dots + d_m e(t-m) \quad (2.17)$$

o

$$C(q^{-1}) u(t) = D(q^{-1}) e(t)$$

siendo:

$$\begin{aligned} C(q^{-1}) &= 1 + c_1 q^{-1} + \dots + c_m q^{-m} \\ D(q^{-1}) &= 1 + d_1 q^{-1} + \dots + d_m q^{-m} \end{aligned}$$

Definimos q^{-1} como el operador retardo, por lo tanto $q^{-1} u(t) = u(t-1)$.

A este tipo de filtro se le llama Auto- Regresivo de Media Móvil (ARMA):

si $C_i = 0 \Rightarrow$ MA (media móvil)

si $D_i = 0 \Rightarrow$ AR (auto- regresivo)

Los pasos a seguir para diseñar un filtro son:

1. definir el orden de los polinomios $C(q^{-1})$ y $D(q^{-1})$;
2. fijar el valor de los parámetros;
3. generar una secuencia de variables Random;
4. calcular: $u(t) = \frac{D(q^{-1})}{C(q^{-1})} e(t)$

Para garantizar la estabilidad, las raíces del polinomio $C(q^{-1})$ debe estar dentro del círculo unidad. El ancho de banda de dicho señal será función del filtro diseñado.

2.2.3.2 Secuencia binaria pseudo-random (PRBS)

Es una secuencia de pulsos rectangulares, modulados en amplitud, que se caracteriza por tener dos valores, correspondientes a los niveles lógicos 0 y 1, asignados generalmente a $+V$ y $-V$. Es una señal que en su contenido frecuencial, se aproxima a un ruido blanco discreto. Se denomina pseudo random por el hecho de que la longitud del pulso es una variable random. Puede ser generada directamente a partir de un ruido blanco, por ejemplo:

$$u(t) = \text{sign}(\text{randn}(500,1))$$

es persistentemente excitada en toda su longitud.

2.2.3.3 Secuencia binaria de longitud máxima, MLBS.

Es la señal binaria mas ampliamente utilizada. Se puede generar fácilmente utilizando el desplazamiento de registros mediante una realimentación apropiada, en la tabla 2.1 se muestra como generar un MLBS según el número de registros. Un ejemplo de la generación de este tipo de señales se muestre en la figura 2.1, [Söderström89].

La MLBS es una señal periódica de longitud:

$$M = 2^m - 1 \quad (2.18)$$

siendo m el número de registros a desplazar, y de período:

$$T_{MLBS} = M\Delta t \quad (2.19)$$

donde Δt es el periodo del reloj que desplaza los registros.

Tabla 2.1. Generación de la PRBS [Landau90]

Número de registros	2	3	4	5	6	7	8	9	10
Longitud de la secuencia	3	7	15	31	63	127	255	511	1023
Bits sumados	1,2	1,3	3,4	3,5	5,6	4,7	2,8	5,9	7,10

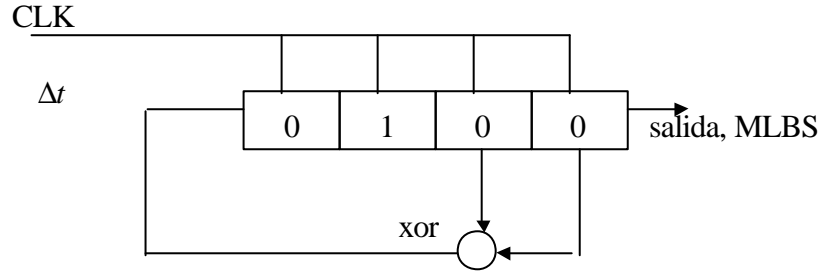


Figura 2.1. Generación de una señal MLBS de 4 registros

Ver que como mínimo uno de los N registros debe ser distinto a cero.

La función de covarianza de esta señal, cuando el periodo de muestreo es de T_s , y considerando el caso en que $T_s < \Delta t$, vale [Goodfrey80]:

$$R_u(\mathbf{t}) = \begin{cases} V^2 \left(1 - \frac{T_s}{\Delta t}\right) & T_s < \Delta t, \pm M\Delta t, \pm 2M\Delta t, \dots \\ \frac{-V^2}{M} & \text{para los otros valores} \end{cases} \quad (2.20)$$

La densidad espectral calculada a partir de la ecuación (2.6) tiene por expresión

$$\begin{aligned} S_u(\mathbf{w}) &= 2V^2 \int_0^1 \left(1 - \frac{M+1}{M}t\right) \cos(\mathbf{w}t) dt + 2 \frac{V^2}{M} \int_1^M \cos(\mathbf{w}t) dt \\ &= \frac{V^2}{M^2} (M+1) \frac{\sin^2(\mathbf{w}/2)}{(\mathbf{w}/2)^2} \\ &= \frac{V^2 (M+1)}{M^2} \frac{\sin^2(k\mathbf{p}/M)}{(k\mathbf{p}/M)^2} \quad \text{para } k = 1, 2, 3, \dots \end{aligned} \quad (2.21)$$

y la densidad espectral discreta [Söderström89], calculada con la fórmula (2.8) vale:

$$S_u(\mathbf{w}) = \frac{V^2}{M^2} \left[d(\mathbf{w}) + (M+1) \sum_{k=1}^{M-1} d\left(\mathbf{w} - 2\mathbf{p} \frac{k}{M}\right) \right] \quad (2.22)$$

donde k son los armónicos.

La frecuencia de los armónicos queda determinada por Dt i M :

$$f_k = \frac{k}{M Dt} \quad (2.23)$$

y la potencia media del ancho de banda se da aproximadamente $0.443/Dt$ Hz [Evans92]. Por lo tanto es posible, ajustando el valor de M y Dt , cubrir apropiadamente el rango de frecuencias del sistema. El espectro de esta señal decrece con la frecuencia.

El diseño de esta señal puede realizarse de la siguiente manera:

- 1- se selecciona un valor de Dt que origine la máxima frecuencia de interés en el punto -3dB
- 2- con Dt fijado, se ajusta M teniendo en cuenta dos aspectos, $M=2^m -1$ y f_{\min} (calculada a partir de la ecuación (2.23) para $k=1$) debe de estar situada en el rango de interés
- 3- finalmente se selecciona el número de datos a adquirir en cada cambio de registros, la frecuencia de reconstrucción o de muestreo de la señal será de: número de datos por registro/ Dt .

Esta forma de diseño, origina señales con unas características muy cerradas:

- conjuntos $Dt \approx 0.443/BW$, donde BW representa el valor de la frecuencia del proceso a -3dB
- y $M Dt > Ts$, el tiempo de estabilización del proceso.

En [Söderström89] se demuestra que este señal es persistentemente excitado de orden M .

2.2.3.4 Señales *multisinusoidales*

Esta consiste en una suma arbitraria de cosenos de armónicos conexos [Evans92]

$$u(t) = \sum_{i=1}^F A_i \cos(2\pi f_i t + \mathbf{f}_i) \quad (2.24)$$

siendo A_i la amplitud de cada armónico, F el número de armónicos, \mathbf{f} el vector de fases y f la frecuencia, el vector i puede ser cualquier secuencia de valores enteros. Esta señal tiene una gran ventaja respecto a las señales binarias y es que su potencia está concentrada en un intervalo definido de frecuencia. Son señales con un ancho de banda limitada.

La historia temporal de una señal, puede medirse cuantitativamente a partir del factor de cresta, CF , definido en (2.25) [Rees92]. Cuando tenemos procesos con ruido, es necesario que las señales tengan un factor de cresta bajo, de esta forma se consigue que la señal de excitación tenga la máxima energía con una amplitud acotada.

$$CF = \frac{\max(|u(t)|)}{u(t)_{rms}} \quad (2.25)$$

A escala comparativa, diremos que una señal sinusoidal puro tiene un factor de cresta igual a 1.414, mientras que en el caso de una señal binaria pseudo aleatoria (PRBS), $CF=1$. Respecto a las señales multisinusoidales, el valor del factor de cresta depende de la fase de los armónicos. En el caso de considerar fases nulas, el factor de cresta de la señal es alto, $CF > 1$.

Per minimizar el CF de estos tipos de señales, de debe seleccionar cuidadosamente la fase de cada uno de los armónicos [Rees93].

Dos posibles métodos para el diseño de la fase son:

- A-** El método propuesto por Schroeder. La fase de cada armónico se determina utilizando la formula:

$$f_i = \frac{pk^2}{F} \quad (2.26)$$

donde f_i es la fase del armónico i de la señal, k es el nombre del componente, y F es el número total de componentes del espectro; los valores de i y k pueden ser distintos si se excluye alguno de los armónicos de la señal multisinusoidal.

- B-** El método propuesto por Guillaume, L8. Consiste en minimizar el del CF, definido en este caso por la expresión:

$$CF = \frac{l_\infty(u)}{l_2(u)} \quad (2.27)$$

donde $l_\infty(u)$ es la norma de Chebyshev de $u(t)$ y la norma l_2 es el valor de RMS. Se utiliza el método de Gauss-Newton para minimizar el valor de la norma l_∞ . Este método tiene el inconveniente que su cálculo es muy largo.

En el diseño de las señales multisinusoidales se deben considerar dos aspectos. El primero es que la señal diseñada debe concentrar toda la potencia en el ancho de banda del proceso estudiado. Y el segundo, es que debe tener el suficiente número de armónicos para conseguir minimizar el CF, cubrir el espectro necesario y permitir la detección de errores en la modelización. Por otro lado, el hecho de incorporar muchos armónicos es perjudicial ja que reduce la energía de cada test de frecuencia y, en consecuencia, disminuye la precisión de la respuesta frecuencial.

La función de autocovarianza de una suma de sinusoides, $R_u(t)$, es una función coseno (2.28), donde el valor de los pesos C_i depende del valor de $w_i = 2pf_i$ y de A^2 , tal como se demuestra en [Söderström89].

$$r_u(t) = \sum_{i=1}^F C_i \cos(w_i t) \quad (2.28)$$

con:

$$\begin{aligned} C_i &= \frac{A_i^2}{2} & \text{si } w_i t \neq p, 2p, 3p, \dots \\ C_i &= A_i^2 \sin^2 f_i & \text{si } w_i t = p, 2p, 3p, \dots \end{aligned}$$

La densidad espectral discreta para $0 \leq w_i \leq p$, tiene por expresión [Söderström89]:

$$S_u(w) = \sum_{i=1}^F \frac{C_i}{2} (d(w - w_i) + d(w + w_i)) \quad (2.29)$$

A partir de la ecuación (2.29) y en el intervalo de valores $[-p, p]$ se deduce [Söderström89] que este tipos de señales son persistentemente excitados de orden:

$$\begin{cases} 2F & \text{si } 0 < w_1, w_F < p \\ 2F - 1 & \text{si } 0 = w_1 \text{ o } w_F = p \\ 2F - 2 & \text{si } 0 = w_1 \text{ i } w_F = p \end{cases} \quad (2.30)$$

2.3 Análisis de la respuesta transitoria

Las señales test no periódicas utilizadas para evaluar la respuesta transitoria son: impulsos de diferente amplitud y de corta duración, función escalón y función rampa. Para seleccionar la señal test adecuada, debe tenerse en consideración: las propiedades de la señal, la posibilidad de la señal de ser generada y aplicada y el tipo de información que se persigue del proceso.

2.3.1 Análisis de la respuesta impulso

La metodología de identificación denominada análisis de la respuesta impulso consiste en aplicar como entrada al proceso una señal impulso. La ecuación general utilizada en la representación de la función de transferencia impulso de un sistema es:

$$y(t) = G(q^{-1})u(t) + v(t) = \sum_{k=1}^{\infty} g(k)u(t - k) + v(t) \quad (2.31)$$

en donde $G(q^{-1})$ representa la función de transferencia y $g(k)$ la respuesta impulso.

Cuando este sistema está sujeto a una entrada impulso definida por:

$$u(t) = \begin{cases} a & , t = 0 \\ 0 & , t \neq 0 \end{cases} \quad (2.33)$$

la salida es:

$$y(t) = a g(t) + v(t) \quad (2.34)$$

El valor estimado de los coeficientes de la respuesta impulso se determina a partir de:

$$g(t) = \frac{y(t)}{a} \quad (2.35)$$

con un error de estimación de: $v(t)/a$.

Los problemas que se plantean en el análisis de la respuesta impulso son:

- queda restringido al estudio de sistemas estables,
- dificultades de generar una señal impulso,
- problemas de sincronización entre la entrada y el muestreo,
- dificultades con la amplitud de la señal,
- problemas cuando hay saturaciones y no linealidades,
- dificultades con las colas del impulso debido a su larga duración y bajas amplitudes,
- gran sensibilidad al ruido.

2.3.2 Análisis de la respuesta escalón

Un complemento de la respuesta a un impulso es la muy conocida respuesta a un escalón, que tienen como ventaja que es más fácil de generar. La señal de test se describe como:

$$u(t) = \begin{cases} \mathbf{a} & , t > 0 \\ 0 & , t \leq 0 \end{cases} \quad (2.36)$$

Aplicada esta al sistema descrito en (2.31) se obtiene que la salida puede determinarse mediante:

$$y(t) = \mathbf{a} \sum_{k=1}^t g(k) + v(t) \quad (2.37)$$

Partiendo de ella se puede estimar la respuesta impulso como:

$$g(t) = \frac{y(t) - y(t-1)}{\mathbf{a}} \quad (2.38)$$

siendo el error de estimación: $[v(t) - v(t-1)]/\mathbf{a}$. En muchas aplicaciones prácticas la utilización de la ecuación (2.38) da errores elevados.

Muy a menudo es interesante observar simplemente la respuesta a un escalón. El modelo obtenido se usa por ejemplo para el diseño de controladores. Este método permite de forma simple hacer una estimación de un modelo paramétrico. Los métodos para la evaluación de la respuesta escalón se basan en un pequeño nombre de características de la función respuesta. Uno de los métodos más antiguos y muy conocidos es el introducido por *Küpfmüller*, el cual dado un sistema de primer orden con retardo expresado por la ecuación diferencial:

$$T \frac{dy(t)}{dt} + y(t) = Ku(t - t) \quad (2.39)$$

o en forma de función de transferencia:

$$G(s) = \frac{K}{1 + sT} e^{-st} \quad (2.40)$$

calcula el valor de los parámetros K , T y t analizando la respuesta escalón de la forma indicada en la figura 2.2.

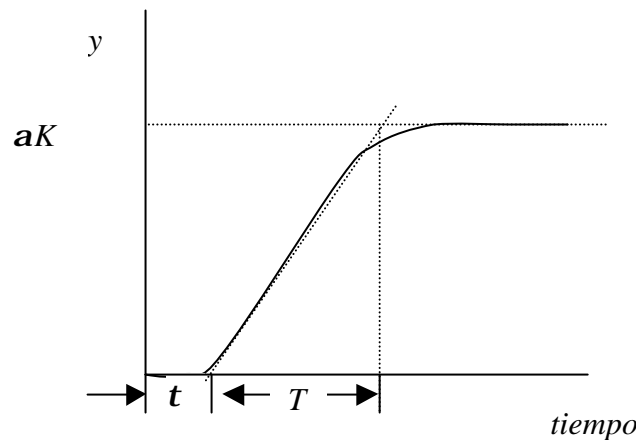


Figura 2.2. Evaluación de la respuesta a un escalón según *Küpfmüller*

Otro método, desarrollado por *Strejc*, realiza la estimación de los valores de los parámetros de un modelo de orden n con la misma constante de tiempo y retardo:

$$G(s) = \frac{K}{(1 + sT)^n} e^{-st} \quad (2.41)$$

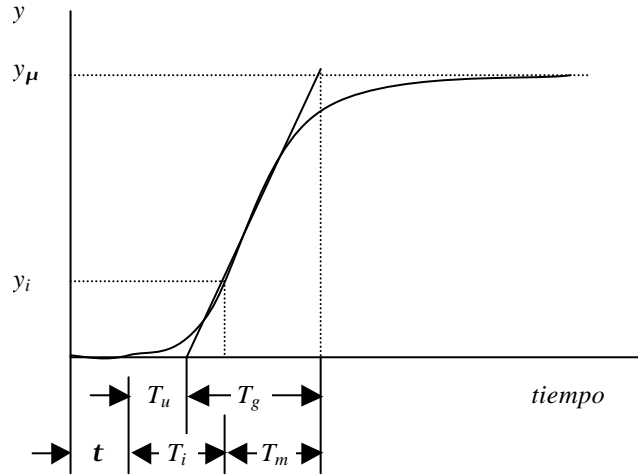


Figura 2.3. Evaluación de la respuesta a un escalón según *Strejc*

Las características de la respuesta escalón se muestran en la tabla 2.2 y la constante de tiempo se calcula mediante: $t_1 = T/n$.

Un método más complejo es el propuesto por *Schwarze* el cual permite evaluar las respuestas escalón y rampa para diferentes estructuras de modelos. Más información al respecto se encuentra en [Rake80].

Tabla 2.2. Valores característicos de la respuesta escalón

n	$\frac{T_g}{T}$	$\frac{T_u}{T}$	$\frac{T_u}{T_g}$	$\frac{T_i}{T}$	$\frac{y_i}{y_\infty}$	$\frac{T_m}{T}$	$\frac{T_m}{T_g}$
1	1.000	0.000	0.000	0	0.000	1.000	1.000
2	2.718	0.282	0.104	1	0.204	2.000	0.736
3	3.695	0.805	0.218	2	0.323	2.500	0.677
4	4.463	1.425	0.319	3	0.353	2.888	0.674
5	5.199	2.100	0.410	4	0.371	3.219	0.629
6	5.699	2.811	0.493	5	0.384	3.510	0.616
7	6.226	3.549	0.570	6	0.394	3.775	0.606
8	6.711	4.307	0.642	7	0.401	4.018	0.599
9	7.146	5.081	0.709	8	0.407	4.245	0.593
10	7.590	5.869	0.773	9	0.413	4.458	0.587

2.4 Análisis de correlación

El análisis de correlación es un método basado en un análisis estadístico y tiene como objetivo estimar la respuesta impulsional del proceso. Esta, se determina a partir del análisis de correlación entre la señal de entrada y la señal de salida. Las señales test más utilizadas en este ensayo son las secuencias de ruido blanco o, en la práctica, las PRBS.

2.4.1 Definiciones previas

La función de *covarianza cruzada* entre dos variables, $x(t)$ y $y(t)$ se determina:

$$\begin{aligned} \text{Cov}\{x(t_1), y(t_2)\} &= E\{(x(t_1) - \bar{x})(y(t_2) - \bar{y})\} \\ R_{xy}(\mathbf{t}) &= E\{(x(t) - \bar{x})(y(t + \mathbf{t}) - \bar{y})\} \cong \frac{1}{N} \sum_{t=0}^N ((x(t) - \bar{x})(y(t + \mathbf{t}) - \bar{y})) \end{aligned} \quad (2.42)$$

con $\mathbf{t}=0, \pm 1, \pm 2, \dots$

La función de *auto correlación* se determina:

$$r_y(y(t_1), y(t_2)) = \frac{R_y\{y(t_1), y(t_2)\}}{\mathbf{s}_y^2} = \frac{R_y(\mathbf{t})}{R_y(0)} \quad (2.43)$$

ya que: $R_y(0) = \mathbf{s}_y^2 = \text{Var}\{y(t)\}$

La función de *correlación cruzada* se calcula:

$$r_{xy}(x(t_1), y(t_2)) = \frac{R_{xy}\{x(t_1), y(t_2)\}}{\mathbf{s}_x \mathbf{s}_y} = \frac{R_{xy}(\mathbf{t})}{\sqrt{R_x(0)} \sqrt{R_y(0)}} \quad (2.44)$$

Definición: Se dice que dos señales $x(t)$ y $y(t)$ no están correlacionados (son independientes) con un 95% de confianza si:

$$\begin{aligned} P\left(|r_{xy}(\mathbf{t})| \leq \frac{1.96}{\sqrt{N}}\right) &= 0.95 \\ P\left(|R_{xy}(\mathbf{t})| \leq \frac{1.96 \sqrt{R_x(0)} \sqrt{R_y(0)}}{\sqrt{N}}\right) &= 0.95 \end{aligned} \quad (2.45)$$

2.4.2 Análisis de correlación. Descripción

El modelo utilizado en el análisis de correlación es la respuesta impulso o función de pesos definida en (2.31) o la función continua equivalente.

Si la entrada es una secuencia casi estacionaria de media nula con:

$$R_u(\mathbf{t}) = E\{(u(t))(u(t + \mathbf{t}))\} \quad (2.46)$$

y asumiendo que se trata de un proceso estocástico estacionario, la relación equivalente entre la función de covarianza cruzada de la entrada con la salida y la función de auto covarianza de la entrada es:

$$R_{yu}(\mathbf{t}) = E\{y(t)u(t-\mathbf{t})\} = \sum_{k=0}^{\infty} g(k)E\{u(t)u(t-\mathbf{t})\} + E\{v(t)u(t-\mathbf{t})\} \quad (2.47)$$

A esta ecuación de la denomina ecuación de *Wiener-Hopf*. En el caso en que la señal de entrada y las perturbaciones sean independientes:

$$R_{yu}(\mathbf{t}) = \sum_{k=0}^{\infty} g(k)E\{u(t)u(t-\mathbf{t})\} = \sum_{k=0}^{\infty} g(k)R_u(\mathbf{t}-k) \quad (2.48)$$

En la práctica, ya que las señales de entrada y salida no son infinitas, las funciones de covarianza se estiman a partir de los datos de la siguiente forma:

$$\begin{aligned} \hat{R}_{yu}(\mathbf{t}) &= \frac{1}{N} \sum_{t=1-\min(\mathbf{t},0)}^{N-\max(\mathbf{t},0)} y(t)y(t+\mathbf{t}), \quad \mathbf{t} = 0, \pm 1, \pm 2, \dots \\ \hat{R}_u(\mathbf{t}) &= \frac{1}{N} \sum_{t=1}^{N-\mathbf{t}} u(t)u(t+\mathbf{t}) \end{aligned} \quad (2.49)$$

En estas condiciones, el valor estimado de los pesos de la función $g(k)$ puede ser determinado solucionando la ecuación (2.50) que consiste en un sistema lineal de dimensión finita.

$$\hat{R}_{yu}(\mathbf{t}) = \sum_{k=0}^{\infty} \hat{g}(k)R_u(\mathbf{t}-k) \quad (2.50)$$

El problema queda muy simplificado en el caso en que se utilicen señales de test especiales con funciones de auto covarianza sencillas. Cuando se utiliza, por ejemplo, un ruido blanco como señal de test, sabiendo que $\hat{R}_u(\mathbf{t}) = 0$ para $\mathbf{t} > 0$, la ecuación 2.50 queda reducida a:

$$g(\mathbf{t}) = \hat{R}_{yu}(\mathbf{t}) / R_u(0) \quad (2.51)$$

siendo fácilmente calculable a partir de los datos experimentales.

En el caso en que la señal de excitación no sea un ruido blanco, el cálculo de la ecuación 2.50 no es trivial. A continuación se proponen dos métodos para resolverla:

1. Función de pesos truncada o respuesta impulsional finita (FIR). En este caso se considera que la función impulso es de orden finito: $g(k)=0$ para $k \geq M$. Para ser utilizado, es necesario que M sea mayor que la constante dominante del sistema. En este caso una buena aproximación para solucionar la ecuación 2.50 es:

$$\hat{R}_{yu}(\mathbf{t}) = \sum_{k=0}^{M-1} \hat{g}(k)R_u(\mathbf{t}-k) \quad (2.52)$$

Escribiendo esta ecuación de forma desarrollada para $\mathbf{t} = 0, 1, \dots, M-1$, se obtienen el siguiente conjunto de ecuaciones lineales:

$$\begin{pmatrix} \hat{R}_{yu}(0) \\ \hat{R}_{yu}(1) \\ \hat{R}_{yu}(2) \\ \vdots \\ \hat{R}_{yu}(M-1) \end{pmatrix} = \begin{pmatrix} \hat{R}_u(0) & \hat{R}_u(1) & \hat{R}_u(2) & \dots & \hat{R}_u(M-1) \\ \hat{R}_u(1) & \hat{R}_u(0) & \hat{R}_u(1) & \dots & \hat{R}_u(M-2) \\ \hat{R}_u(2) & \hat{R}_u(1) & \hat{R}_u(0) & \dots & \hat{R}_u(M-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{R}_u(M-1) & \hat{R}_u(M-2) & \hat{R}_u(M-3) & \dots & \hat{R}_u(0) \end{pmatrix} \begin{pmatrix} \hat{g}(0) \\ \hat{g}(1) \\ \hat{g}(2) \\ \vdots \\ \hat{g}(M-1) \end{pmatrix} \quad (2.53)$$

Las ecuaciones anteriores pueden ser aplicadas a más de M valores de t dando lugar a un sistema de ecuaciones lineales sobre determinado.

2. *Aplicando filtros.* Un segundo método propuesto sería el de aplicar un filtro para blanquear la señal de entrada. Cuando ambas señales, entrada y salida, son filtradas con el mismo filtro $L(q^{-1})$

$$\begin{aligned} y_F(t) &= L(q^{-1})y(t) \\ u_F(t) &= L(q^{-1})u(t) \end{aligned} \quad (2.54)$$

las señales filtradas presentan la misma relación que en la ecuación (2.31)

$$y_F(t) = \sum_{k=0}^{\infty} g(k)u_F(t-k) + v_F(t) \quad (2.55)$$

El filtro L debe ser seleccionado de tal manera que la señal de entrada $u_F(t)$ sea lo más parecida posible a un ruido blanco. Puede calcularse considerando que L es un filtro AR (auto regresivo): $A(q^{-1})u(t)=e(t)$, donde el polinomio $A(q)=L(q)$ puede calcularse utilizando el método de mínimos cuadrados (expuesto en el capítulo 3). El orden del polinomio A , na , suele estar comprendido entre 4 y 8. Se determina la función impulsional aplicando la ecuación (2.51) de la forma:

$$\hat{g}(k) = \frac{R_{y_F u_F}(k)}{R_{u_F}(0)} \quad (2.56)$$

Más detalles sobre el mismo se exponen en [Ljung94].

Las propiedades básicas del análisis de correlación son:

- Al igual que en el análisis de la respuesta transitoria, el análisis de correlación da una rápida información sobre la constante de tiempo dominante del sistema y del retardo puro.
- No es necesario ninguna entrada en especial.
- Dan como resultado una tabla de datos o una gráfica que no pueden ser utilizados directamente en simulación.
- Los resultados obtenidos parten del hecho de que la entrada es independiente de las perturbaciones y/o ruido, esto limita el uso de este método a sistemas sin realimentación (anillo abierto).

2.5 Técnicas frecuenciales

Los sistemas lineales pueden también definirse a partir de la respuesta frecuencial $G(j\omega)$. Mientras que las respuestas transitorias y el análisis de correlación tienen por objetivo estimar la respuesta impulsional, las técnicas frecuenciales tienen por objetivo la estimación directa de la respuesta frecuencial, el cómo a partir de ella se estiman los parámetros de la función de transferencia se abordará en el tema 3. Las técnicas que se presentan en este apartadp son el análisis de Fourier y el análisis espectral.

2.5.1 Análisis de la respuesta frecuencial

El análisis de la respuesta frecuencial es un método simple para obtener información acerca de un sistema lineal. Consiste simplemente en describir como se comporta el sistema cuando esta sometido a una entrada senoidal. Al aplicar una señal de entrada definida por:

$$u(t) = a \sin \omega t \quad (2.57)$$

la respuesta del sistema puede describirse como:

$$y(t) = |G(j\omega)|a \sin(\omega t + f(\omega)) + v(t) + \text{transitorio} \quad (2.58)$$

donde $f(\omega) = \arg G(j\omega)$

Asumiendo que se puede ignorar el transitorio y considerar solo la respuesta estacionaria, la salida del sistema puede utilizarse directamente para determinar $G(j\omega)$. Observando simplemente la señal de entrada y de salida, se puede calcular la ganancia y el desfase entre ambas señales. La figura 2.4 muestra como evaluar la respuesta a una entrada senoidal.

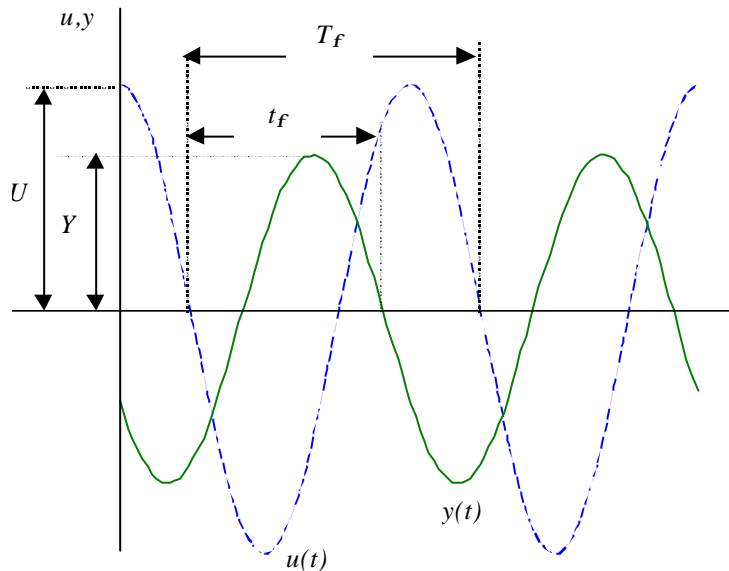


Figura 2.4. Evaluación de la respuesta sinusoidal.

La ganancia y la fase se obtienen:

$$|G(j\omega)| = Y/U$$

$$f(\omega) = -2\pi t_f / T_f$$

Cuando el experimento se repite con distintas frecuencias pertenecientes al ancho de banda del sistema, el resultado puede dibujarse en forma de *diagrama de Bode*.

Las ventajas de este método son:

- Es fácil de utilizar y no se requiere un procesamiento difícil de los datos.

- No es necesario hacer ninguna suposición sobre la estructura del modelo, solo considerar que se trata de un modelo lineal.
- Es fácil concentrar el estudio del proceso en un rango de frecuencias determinado.

Los inconvenientes que presenta este método son:

- Dan como resultado una tabla de datos o un gráfico que no puede ser utilizado directamente en simulación.
- Son necesario largos periodos de pruebas si se quiere evaluar el valor de $G(j\omega)$ para un gran número de frecuencias.
- Es muy sensible al ruido presente en el sistema y no da ningún valor medio estadístico.
- En la mayoría de las veces resulta difícil determinar el valor de desfase entre las señales.

Para evitar el problema de las perturbaciones, se puede utilizar el denominado método de correlación para el análisis de la respuesta frecuencial. Tal como se muestra en la figura 2.5, consiste en multiplicar la salida del sistema por una señal seno y una señal coseno de la misma frecuencia que la entrada del sistema y a continuación integrar las señales obtenidas considerando un intervalo específico de medida T_m :

$$y_s(\omega) = \int_0^{T_m} y(t) \sin \omega t \, dt = \frac{1}{2} T_m |G(j\omega)| a \cos f(\omega) \quad (2.59)$$

$$y_c(\omega) = \int_0^{T_m} y(t) \cos \omega t \, dt = \frac{1}{2} T_m |G(j\omega)| a \sin f(\omega) \quad (2.60)$$

La estimación de la amplitud y de la fase se calcula mediante:

$$|\hat{G}(j\omega)| = \frac{2}{aT_m} \sqrt{y_s^2(\omega) + y_c^2(\omega)} \quad (2.61)$$

$$\hat{f}(\omega) = \arctan \frac{y_c(\omega)}{y_s(\omega)} + kp \quad (2.62)$$

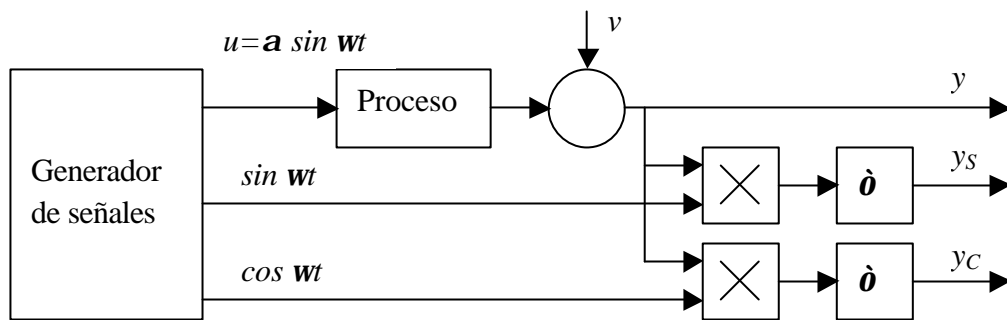


Figura 2.5. Análisis de la respuesta frecuencial con técnicas de correlación

Para minimizar el efecto de las perturbaciones, la duración de las medidas T_m debe ser siempre un múltiplo del periodo de la señal de test. En el mercado hay equipos que utilizan este método para realizar el análisis frecuencial. El principal inconveniente de este método radica en el hecho de que en la mayoría de procesos industriales no es posible aplicar una señal senos como entrada en operación normal y, al igual que el método anterior, requiere

repetir los experimentos para muchas frecuencias, hecho que en muchos casos lo hace inviable.

2.5.2 Análisis de Fourier

Se dice que una señal, $u(t)$, es de energía finita si:

$$\int_{-\infty}^{\infty} |u(t)| dt < \infty \quad \text{o} \quad T \sum_{k=-\infty}^{\infty} |u(kT)| < \infty \quad (2.63)$$

Si la entrada tiene energía finita, es posible determinar la función de transferencia frecuencial utilizando la transformada de Fourier de las señales de entrada y salida:

$$Y(\mathbf{w}) = G(i\mathbf{w})U(\mathbf{w}) \quad \Rightarrow \quad G(i\mathbf{w}) = \frac{Y(\mathbf{w})}{U(\mathbf{w})} \quad (2.64)$$

En el caso de tener acceso a datos en un intervalo de tiempo finito $0 < t < S$, la transformada de Fourier de las señales se determina mediante:

$$Y_s(\mathbf{w}) = \int_0^S y(t) e^{-j\mathbf{w}t} dt, \quad U_s(\mathbf{w}) = \int_0^S u(t) e^{-j\mathbf{w}t} dt \quad (2.65)$$

por lo tanto, la estimación de la función de transferencia empírica (ETFE) se determina con la ecuación:

$$\hat{G}_s(j\mathbf{w}) = \frac{Y_s(\mathbf{w})}{U_s(\mathbf{w})} \quad (2.66)$$

En el caso en que se disponga de datos muestreados: $u(kT)$, $y(kT)$ para $k=1, \dots, N$, se puede utilizar para el cálculo de la transformada de Fourier discreta:

$$Y_s(\mathbf{w}) = T \sum_{k=1}^N y(kT) e^{-j\mathbf{w}kT} \quad \text{y} \quad U_s(\mathbf{w}) = T \sum_{k=1}^N u(kT) e^{-j\mathbf{w}kT} \quad (2.67)$$

siendo T el periodo de muestreo y $N=S/T$.

Cuando la frecuencia tiende por valor $\mathbf{w}=r \, 2\mathbf{p}/N$, para $r = 0, \dots, N-1$, y N se ajusta a una potencia de dos, es posible utilizar la FFT para estimar los valores de $Y_s(\mathbf{w})$ y $U_s(\mathbf{w})$.

El siguiente teorema nos permite evaluar las propiedades de la ETFE.

Teorema 2.1: Considerando que un sistema tiene por expresión

$$y(t) = \int_0^{\infty} g(\mathbf{t}) u(t-\mathbf{t}) d\mathbf{t} + v(t) \quad (2.68)$$

y asumiendo que:

$$|u(t)| \leq c_u \quad \text{y} \quad \int_0^{\infty} \mathbf{t} |g(\mathbf{t})| d\mathbf{t} = c_g$$

se deduce que [Ljung94]

$$\left| \hat{G}_S(j\omega) - G(j\omega) \right| \leq \frac{2c_u \cdot c_g}{|U_S(\omega)|} + \frac{|V_S(\omega)|}{|U_S(\omega)|} \quad (2.69)$$

Para interpretar el resultado se deben hacer suposiciones respecto a la señal de entrada.

Caso 1. Entrada periódica y S es un múltiplo del período:

- ETFE está definida per un número fijo de frecuencias.
- A estas frecuencias la ETFE converge al valor real y su varianza decrece en función de $1/N$.

Caso 2. La entrada es una señal no periódica:

- La ETFE converge asintoticamente a la función de transferencia cuando se incrementa el número de datos, a mayor S mayor número de frecuencias evaluadas.
- La variancia de ETFE no decrece al incrementar S , esta se mantiene constante e igual a la relación señal/ruido para cada frecuencia.

Entre las ventajas e inconvenientes del análisis de Fourier destacan:

- Es un método fácil e eficiente, especialmente cuando se aplica la FFT.
- Permite una buena estimación de $G(j\omega)$ cuando la entrada es una señal periódica.
- Para señales no periódicas la función obtenida es muy fluctuante.

2.5.3 Análisis espectral

El espectro de una señal $u(t)$ o periodograma, $F_u(\omega)$, se puede interpretar como una media de su contenido frecuencial y se define como:

$$\Phi_u(\omega) = 1/N \dot{U}(\omega) \dot{U}^*(\omega) \quad (2.70)$$

Las propiedades de los periodogramas son:

1. Una señal puramente senoidal presenta picos en este diagrama.
2. Con otros tipos de señales el resultado obtenido es muy fluctuante.
3. Suavizando visualmente los resultados obtenidos puede obtenerse, a partir de este análisis, una idea del contenido frecuencial de la señal.

Dado una señal $u(t)$ obtenida de un proceso estocástico con una densidad espectral $\Phi_u(\omega)$, el valor estimado de $\hat{\Phi}_N$ será:

$$\hat{\Phi}_N(\omega) = \Phi_u(\omega) + R_N \quad (2.71)$$

siendo R_N el termino residual, responsable de las fluctuaciones. Este término se caracteriza por tener un valor medio nulo y una variancia que depende del tipo de señal y puede tender a cero a medida que se aumenta N .

Como resultado de aplicar la transformada de Fourier al modelo representado por:

$$y(t) = G(q)u(t) + v(t) \quad (2.72)$$

se obtiene:

$$Y(\omega) = G(j\omega)U(\omega) + V(\omega) \quad (2.73)$$

Considerando que las señales $u(t)$ y $v(t)$ son independientes, al multiplicar la ecuación (2.73) por el término $U(\mathbf{w})$ se obtiene la ecuación (2.74) con la cual es posible estimar la función frecuencial (2.75). Y, al mismo tiempo, como resultado de elevar al cuadrado la ecuación (2.73) se obtiene la ecuación (2.76) que hace posible estimar el espectro de la perturbación (2.77).

$$\Phi_{yu}(\mathbf{w}) = G(i\mathbf{w})\Phi_u(\mathbf{w}) \quad (2.74)$$

La estimación de la función espectral se realiza mediante:

$$\hat{G}_N(i\mathbf{w}) = \frac{\hat{\Phi}_{yu}^N(\mathbf{w})}{\hat{\Phi}_u^N(\mathbf{w})} \quad (2.75)$$

El análisis del espectro de las perturbaciones se obtiene:

$$\Phi_y(\mathbf{w}) = |G(i\mathbf{w})|^2 \Phi_u(\mathbf{w}) + \Phi_v(\mathbf{w}) \quad (2.76)$$

$$\hat{\Phi}_v^N(\mathbf{w}) = \hat{\Phi}_y^N(\mathbf{w}) - \frac{|\hat{\Phi}_{yu}^N(\mathbf{w})|^2}{\hat{\Phi}_u^N(\mathbf{w})} \quad (2.77)$$

Para aplicar la ecuación (2.75), debe determinarse en primer lugar el espectro de la señal de entrada $u(t)$ y el espectro cruzado entre las señales $u(t)$ y $y(t)$. Se ha visto en (2.71) que el espectro tiene asociado un término residual, por lo tanto para realizar una buena estimación de la función de transferencia del sistema será necesario reducir este término. En la bibliografía hay distintos métodos para reducir la variación o término residuo del espectro de una señal, aquí se describen dos de ellos por ser los utilizados en el programa MATLAB.

1. Método de Welch (*pwelch* de MATLAB)

Consiste en descomponer la señal en L segmentos de longitud M (siendo M una potencia de dos) y con un tanto por ciento de solapamiento, determinar el espectro de cada uno de estos segmentos y hacer un promedio entre los resultados obtenidos.

$$\hat{\Phi}_N(\mathbf{w}) = \frac{1}{L} \sum_{k=1}^L \hat{\Phi}_M^{(k)}(\mathbf{w}) \quad (2.78)$$

Con ello se reduce el factor R_N a costa de reducir la resolución espectral.

2. Método de Blackman-Tukey

Este método consiste en reducir las fluctuaciones haciendo un promedio entre un nombre determinado de frecuencias vecinas.

$$\hat{\Phi}_N(\mathbf{w}) = \int_{-p}^p W_g(\mathbf{w} - \mathbf{x}) \hat{\Phi}_N(\mathbf{x}) d\mathbf{x} \quad (2.79)$$

donde: $W_g(\mathbf{w})$ es una función ventana centrada y $1/g$ describe la longitud de la ventana.

Otra forma distinta a la ecuación (2.70) para calcular el espectro de una señal consiste en utilizar la función varianza:

$$\Phi_u(\mathbf{w}) = T \sum_{k=-\infty}^{\infty} R_u(kT) e^{-i\mathbf{w}kT} \quad (2.80)$$

Por lo tanto la ecuación (2.80) puede ser expresada también como:

$$\hat{\Phi}_N(\mathbf{w}) = T \sum_{k=-g}^g W_g(k) \hat{R}_u^N(kT) e^{-i\mathbf{w}kT} \quad (2.81)$$

Las etapas que proponen *Blackman-Tukey* para estimar el espectro de una función son:

- a) Definir el tipo de ventana. En general para el análisis espectral se suele utilizar la ventana *Hamming*:

$$w_g(k) = \begin{cases} \frac{1}{2} (1 + \cos(\pi k/g)) & |k| < g \\ 0 & |k| \geq g \end{cases}$$

- b) Determinar la longitud de la ventana. Este valor es un compromiso entre la resolución espectral (mayor g menor resolución) y la variancia (a mayor g mayor variancia).
- Cuando se desean determinar picos de resonancia son necesarios valores grandes de g , comportando un aumento de la variancia.
 - Cuando se desea evaluar un espectro llano se utilizan valores pequeños de longitud de ventana.
- c) Calcular $R_u(k)$ para $k=0, \dots, g$
- d) Determinar $\hat{\Phi}_N(\mathbf{w})$ a partir de la ecuación (2.81). El espectro cruzado entre las señales de entrada y salida se determinan a partir de:

$$\hat{\Phi}_{yu}^N(\mathbf{w}) = \sum_{\ell=-g}^g \hat{R}_{yu}^N(\ell) W_g(\ell) e^{-i\ell\mathbf{w}} \quad (2.82)$$

2.6 Diseño de los experimentos

Para utilizar debidamente los métodos de identificación es necesario, siempre que sea posible, diseñar adecuadamente el experimento. Para ello es necesario disponer de unos conocimientos básicos sobre el comportamiento del proceso, seleccionar las señales de entrada, ... entre otras. Esta información puede obtenerse realizando una serie de experimentos básicos y sencillos conocidos con el nombre de *experimentos preliminares*.

2.6.1 Experimentos preliminares

Los experimentos preliminares permiten obtener información básica del comportamiento del sistema. Se conocen con este nombre los siguientes tipos de ensayos:

- (a) Los experimentos de *funcionamiento libre* (*free run experiments*) consisten en analizar los datos correspondientes al funcionamiento normal del proceso. Por lo tanto, no se requiere ninguna activación específica de las entrada. Este tipo de experimentos permite determinar las características de las perturbaciones que afectan a la salida del proceso y el ruido inherente en las señales capturadas.
- (b) El experimento en *escalón* (*staircase experiment*) es útil para observar el grado de linealidad del proceso estudiado. Consiste en aplicar, a una de las entradas seleccionadas, un test con escalones sucesivos, el intervalo de tiempo de cada uno de los cambios del escalón debe ser suficiente para estabilizar el sistema. El test de linealidad, en estado

estacionario, consiste en: (1) seleccionar de los datos de entrada y salida los puntos de equilibrio, (2) ajustar a los mismos un polinomio de segundo o tercer orden y (3) observar el valor de los términos de mayor orden del polinomio ajustado; en el caso de que estos términos sean muy significativos implica que el proceso tiene un gran componente de no linealidad. Para eliminar dicha no linealidad, puede precompensarse los datos del proceso utilizando el modelo inverso del polinomio. En el caso de que no sea posible aplicar una señal escalonada al proceso (técnica no paramétrica expuesta en el apartado 2.3.1), por motivos técnicos o económicos, si que se acepta habitualmente la utilización de un escalón simple, en este caso solo será posible estimar la constante de tiempo más relevante del proceso estudiado.

- (c) Los experimentos con *ruido blanco* (*white noise experiment*) pueden ser útiles para determinar el ancho de banda y los retardos del proceso. Este tipo de experimento se caracterizan por excitar el proceso mediante un ruido blanco. Con los datos resultantes de dicho experimento puede realizarse el estudio espectral y/o el análisis de correlación [Zhu93], técnica también no paramétrica descritas anteriormente. Otros tipos de señales a utilizar en la realización de dicho experimento son señales con el ancho de banda limitado como la *PRBS* (*pseudo random binary sequence*) o senoides múltiples (estas señales se describen en el próximo apartado).
- (d) Finalmente, se menciona la técnica de la respuesta frecuencial del proceso realimentado con un relé con histéresis [Åström88]. Este método consiste en conectar el proceso de la forma indicada en la figura 2.6. Dando lugar, en la mayoría de los procesos, a una onda cuadrada a la salida del relé y una onda aproximadamente sinusoidal a la salida del proceso. Esta técnica es muy útil para determinar las características fundamentales del sistema en el caso en que, por los motivos que sea, no pueda operarse en lazo abierto.

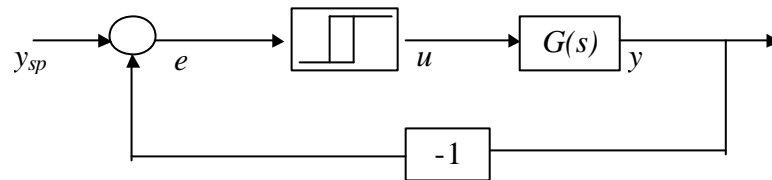


Figura 2.6. Diagrama de bloques de un proceso realimentado con relé

Como resultado de descomponer en serie de Fourier la señal de salida del relé y considerando solo el primer armónico, se obtiene que la ecuación descriptiva del relé con histéresis es [Cook86]:

$$N(A_y) = \frac{4A_u}{\pi A_y} \left(\sqrt{1 - \frac{h^2}{A_y^2}} - j \frac{h}{A_y} \right) \quad (2.83)$$

siendo, A_u la amplitud del relé; A_y la amplitud de oscilación de la salida; y h la histéresis del relé. Dado un proceso que tiene por función de $G(s)$, las condiciones de oscilación se dan cuando:

$$N(A_y)G(i\omega) = -1 \quad (2.84)$$

que corresponde al caso en que la función de transferencia evaluada en el campo frecuencial, $G(\mathbf{w}j)$, intercepta con $-1/N(A_y)$, figura 2.7. La ecuación (2.84) se obtiene a partir del requerimiento que una onda sinusoidal de frecuencia \mathbf{w} se propaga a través del anillo realimentado con la misma amplitud y fase.

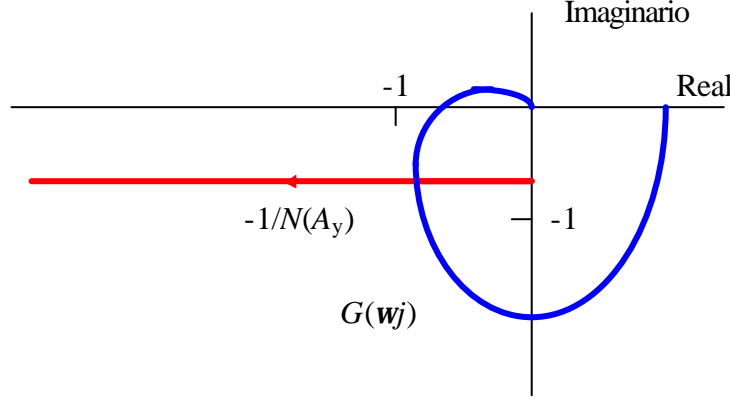


Figura 2.7. Diagrama de Nyquist.

Conocido el valor de $N(A_y)$, puede estimarse un punto de la respuesta frecuencial, que coincida con la pulsación de oscilación \mathbf{w}_h . El módulo y la fase de este punto se determinan a partir de la transformada de Fourier de las señales:

$$|G(i\mathbf{w}_h)| = \left| \frac{Y(\mathbf{w}_h)}{U(\mathbf{w}_h)} \right| = K_h \quad (2.85)$$

$$\arg(G(i\mathbf{w}_h)) = \text{atan}\left(\frac{Y(\mathbf{w}_h)}{U(\mathbf{w}_h)}\right) = \mathbf{f}_h$$

La información sobre la constante de tiempo del proceso estudiado se obtendrá considerando que sus características se ajusta a una determinada función de transferencia. La función de transferencia utilizada por la mayoría de los autores es:

$$G(s) = \frac{k}{(Ts + 1)} e^{-Ls} \quad (2.86)$$

siendo k la ganancia, T la constante de tiempo del proceso y L el retardo puro .

Igualando las ecuaciones (2.85) con el módulo y la fase de la función de transferencia (2.86) se obtienen dos ecuaciones con tres incógnitas: T , k y L .

$$K_h = \frac{k}{\sqrt{T^2 \mathbf{w}_h^2 + 1}} \quad (2.87)$$

$$\mathbf{f}_h = -\text{atan}(T\mathbf{w}_h) - L\mathbf{w}_h$$

La solución se alcanza realizando un segundo experimento con un valor h , histéresis del relé, distinto. Procesando adecuadamente las ecuaciones anteriores, se obtiene que la

constante de tiempo y la ganancia del sistema pueden calcularse a partir de la ecuación (2.88).

$$\begin{bmatrix} T^2 \\ k^2 \end{bmatrix} = \begin{bmatrix} -K_{h1}^2 w_{h1}^2 & 1 \\ -K_{h2}^2 w_{h2}^2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} K_{h1}^2 \\ K_{h2}^2 \end{bmatrix} \quad (2.88)$$

2.6.2 Diseño y selección de las señales de entrada

Las variables de salida del proceso suelen quedar definidas al formularse el problema de modelación, por ejemplo la variable a controlar o a predecir (simular). No sucede lo mismo con la variable de entrada, en general suele haber varias variables de entrada y por lo tanto se plantea el problema de seleccionar cual/es de ellas son las adecuadas para conseguir los objetivos propuestos. Además del problema de la selección, hay otro aspecto a tener en cuenta consiste en determinar la amplitud de la señal de excitación.

2.6.2.1 Selección de las señales de entrada

A la hora de seleccionar cual o cuales son las variables de entrada para el control o la predicción, hay algunas consideraciones que deben tenerse en cuenta:

- (1) La selección de la entrada esta muy influenciada por la amplitud y por como incide en el rango de la salida.
- (2) En el caso de tener como objetivo el control, la variable de entrada debe ser manipulable.
- (3) La función de transferencia salida/entrada debe ser lineal o linealizable (las no linealidades pueden ser compensadas por ejemplo usando un lazo cerrado sencillo).
- (4) Las entradas medidas pero no manipulables pueden utilizarse como señales de entrada en el proceso de identificación (esto implica que el número de entradas a utilizar en la identificación podrá ser mayor que en el control).
- (5) La utilización de la experiencia de los operadores de proceso y el conocimiento que se tenga sobre el mismo son esenciales para la elección de la entrada. Uno de los objetivos de los experimentos preliminares será el de ayudar a la selección de la señal/es de entrada.

2.6.2.2 Diseño de las señales de entrada

El diseño de las señales de entrada involucra dos aspectos: la determinación de la amplitud y la selección de la forma.

En el caso de la determinación de la amplitud de la señal de test se debe tener en cuenta:

- Debe estar restringida por el rango de variación permitido. Por razones de seguridad y económicas, no es posible introducir grandes fluctuaciones en el proceso.
- En la práctica muchos procesos son no lineales y los modelos lineales son solo una aproximación. Una linealización de una dinámica no lineal será válida solo en una cierta región. Para estimar los parámetros de un modelo lineal, una amplitud muy grande en la señal de entrada no será útil.
- Pero, por otro lado, con una amplitud mayor en la entrada se obtiene una mejora en la precisión del modelo estimado, por ejemplo la precisión de un parámetro en términos estadísticos es inversamente proporcional a la energía de la entrada. Esto es lógico ya que al aumentar $u(t)$ la relación señal/ruido incrementa y por lo tanto las perturbaciones o

ruidos tienen una influencia menor. Un criterio sencillo para conocer cual debe ser la amplitud mínima será el de buscar que el efecto de la entrada respecto a la salida sea perceptible visualmente. Otro aspecto que no debemos despreciar, es que la amplitud de la señal de entrada esta muy influenciada por la frecuencia de dicha señal, por lo tanto la medida del *factor de cresta*, $\max(|u(t)_{\max}|, |u(t)_{\min}|) / u(t)_{\text{rms}}$ es también importante. Para una amplitud dada, la *energía de la entrada (input power)* es proporcional a $1/(\text{factor de cresta})^2$.

- El rango de frecuencias contenido en la señal excitación es también un elemento importante a tener en cuenta. En función de las características del proceso, como puede ser la constante de tiempo, el rango de frecuencias más adecuadas puede determinarse mediante:

$$\begin{aligned} f_{\min} &\cong \frac{0.1}{2pT} \\ \frac{5}{2pT} &< f_{\max} \leq \frac{10}{2pT} \end{aligned} \quad (2.89)$$

2.6.3 Selección del período de muestreo y la duración del experimento

Para la identificación de modelos en tiempo discreto el periodo de muestreo deber ser seleccionado correctamente antes de empezar el experimento, ya que una vez almacenados los datos este no puede ser modificado. Al operar con datos muestreados, es inevitable que se pierda información sobre el proceso, por ello es necesario seleccionar la frecuencia de muestreo para que dichas pérdidas sean insignificantes. En la mayoría de los casos las muestras se adquieren a instantes equidistantes de tiempo.

Para seleccionar el periodo se deben tener en cuenta las siguientes consideraciones:

- En muchos casos el periodo de muestreo viene definido por la aplicación final. Por ejemplo el instrumento de control a utilizar, los instrumentos de medida, el tipo de algoritmo de control, ..., limitan en muchos casos el periodo de muestreo.
- La pérdida de información debida al periodo se explica mejor en el dominio frecuencial. Es bien conocido que una señal sinusoidal con una frecuencia mayor que la mitad de la frecuencia de muestreo (denominado frecuencia de *Nyquist*) no puede, al ser muestreada, distinguirse de otra por debajo de este periodo. Consecuentemente, la parte del espectro de la señal que corresponde a altas frecuencias será interpretada como una contribución a bajas frecuencias, denominado fenómeno *alias*. Esto también significa que el espectro de una señal muestreada será la superposición de diferentes partes del espectro inicial.

La información acerca de las frecuencias superiores a la frecuencia de *Nyquist* se pierde en el muestreo. La distorsión producida por el efecto alias, puede ser reducida utilizando un filtro *antialias*. El ancho de banda del filtro deberá ser mucho menor que la mitad de la frecuencia de muestreo. El diseño del filtro depende del orden y de la frecuencia de corte. Una manera de determinar la frecuencia de corte es aplicando la fórmula:

$$f_{\text{corte}} \cong f_{\max} + (f_s/2 - f_{\max}) / 2.$$

Es práctico aplicar el mismo filtro en las señales de salida y entrada. Al disminuir el efecto alias, se incrementa la relación señal/ruido.

- Asumiendo que la duración del experimento es fija, puede ser útil aplicar un periodo de muestreo lo más alto posible, de esta manera se dispone de mas datos.
- Si el numero total de datos a adquirir que es fijo, el periodo de muestreo no debe ser muy grande ni muy pequeño. El primer caso hace que los datos contengan muy poca información acerca de las dinámicas a alta frecuencia y el segundo caso hace que el ruido tendrá mucha influencia.
- Una regla a aplicar puede ser entre 5 o 10 – muy frecuentemente alrededor de 10 – el tiempo de estabilización del proceso. Es muchas veces peor seleccionar un periodo de muestreo demasiado grande que demasiado pequeño.
- Conocida la f_{\max} del sistema, en [Ogata87] se propone como fórmula para determinar el periodo de muestreo: $8f_{\max} \leq f_s \leq 10f_{\max}$.
- Tiempos de muestreo muy pequeños pueden provocar problemas prácticos: todos los polos estarán situados alrededor de 1 en el plano z y el modelo determinado pasará a ser numéricamente muy sensible. En muchos casos el modelo resultante de un muestreo rápido, será de fase no mínima a pesar de que el proceso continuo sea de fase mínima, y en el caso de un proceso con retardo, este, debe ser modelado con muchos datos de retardo causando dificultades, por ejemplo, en el diseño de controladores.

La duración del experimento afecta tanto en la precisión de los parámetros, el cual es inversamente proporcional a la duración del experimento, como a la resolución espectral. Una regla a tener en cuenta es que la duración del experimento debe ser mayor a 5 – 10 veces la mayor constante de tiempo del sistema que se desea identificar. Otro aspecto a tener en cuenta es el rango de frecuencias analizado, este debería ser de dos o tres décadas. Finalmente, la duración del experimento debe tener en cuenta el coste de la realización del experimento.

Un aspecto importante a considerar, cuando se usan señales periódicas, es que la duración del experimento debe cubrir un número completo de periodos o secuencias.

2.6.4 Pretratamiento de los datos

Una vez se dispone de los datos es necesario realizar un tratamiento previo de los mismos, antes de ser utilizados en los algoritmos de identificación. Hay distintas deficiencias en los datos adquiridos que deben ser corregidos:

- Perturbaciones de altas frecuencias en los datos;
- Ocasionales datos espurios y/o olvidados,
- Derivas, offset y perturbaciones a bajas frecuencias,
- Los valores numéricos de las distintas señales no tienen el mismo orden de magnitud,
- Presencia de tiempos de transporte o retardo importantes.

Después de adquirir los datos, lo que debe hacerse es visualizarlos con el objetivo de inspeccionar y analizar si hay algún tipo de deficiencias en ellos. Seguidamente con el objetivo de solucionar el problema, pueden aplicarse alguno de los tratamientos comentados a continuación.

1. *Perturbaciones a altas frecuencias.* Las perturbaciones a altas frecuencias de los datos adquiridos indican que el periodo de muestreo y/o el filtro previo al muestreo no han sido convenientemente seleccionados. Si el periodo de muestreo es innecesariamente pequeño, los datos pueden volverse a muestrear tomando una muestra de cada x muestras de los datos iniciales. En este caso, antes del nuevo muestreo, debe aplicarse un filtro antialias.
2. *Datos espurios y olvidos.* Los datos espurios y olvidos tienen una gran influencia en el resultado de la estimación. Es por lo tanto necesario eliminar los datos incorrectos e insertar datos en los tiempos en que ha habido errores de adquisición. Las etapas a seguir para determinar estos problemas son:
 - Fijar, utilizando el conocimiento previo, un valor de la amplitud de la señal que no sea propia del sistema real;
 - Calcular la desviación estándar de la señal;
 - Interpolan todas las muestras de la señal original que tengan una amplitud superior o inferior al rango fijado definido por los valores no posibles o como múltiplo de la desviación estándar.

Hay algoritmos que permiten detectar los fallos en la adquisición de datos. De todas formas, se debe ser muy prudente para no falsificar los datos originales.

3. *Perturbaciones lentas (offset, tendencias, derivas).* Este tipo de perturbación es muy frecuente en los datos. Ellas suelen ser debidas a causas externas y en la mayoría de los casos se prefiere no incluirlas en el modelo. Hay, básicamente, dos formas de eliminar el problema:
 - Eliminando las perturbaciones mediante un tratamiento de los datos;
 - Definiendo adecuadamente el ruido del modelo teniendo en cuenta las perturbaciones

La primera aproximación implica sustraer directamente las tendencias y offset. En la práctica las técnicas que se aplican para ello son:

- a) Linealizando el modelo alrededor de un punto de equilibrio $u(t)$, $y(t)$:

$$\begin{aligned} y(t) &= \hat{y}(t) - \bar{y} \\ u(t) &= \hat{u}(t) - \bar{u} \end{aligned} \quad (2.90)$$

siendo (\bar{u}, \bar{y}) un punto de equilibrio físico y $(\hat{u}(t), \hat{y}(t))$ las señales medidas.

- b) Parecido al caso anterior pero sustrayendo el valor medio de las variables medidas (2.1).
- c) El offset puede ser estimado de forma explícita extendiendo el modelo y utilizando las variables originales

$$A(q^{-1})\hat{y}(t) = B(q^{-1})\hat{u}(t) + \mathbf{a} + v(t) \quad (2.91)$$

donde \mathbf{a} indica el offset.

En el caso de tendencias o derivas una alternativa a la ecuación (2.90), es substituir el valor medio de las señales por una recta de pendiente definida por la tendencia de las señales o una curva en el caso en que la deriva no sea constante en el tiempo.

Una segunda aproximación es considerar \mathbf{a} como parte del ruido del modelo. En el caso más sencillo esto corresponde a diferenciar los datos (2.92) y luego utilizarlos en la identificación, pero el modelo resultante de la utilización de estos nuevos datos amplifica las altas frecuencias, por ello no puede ser utilizado en muchas aplicaciones.

$$\begin{aligned}\Delta y &= \hat{y}(t) - \hat{y}(t-1) \\ \Delta u &= \hat{u}(t) - \hat{u}(t-1)\end{aligned}\tag{2.92}$$

4. *Escalado de las variables.* En los procesos industriales no todas las variables de entrada y salida son de la misma magnitud. Este problema es importante cuando se evalúa la función de sistemas multivariantes. El problema puede ser eliminado corrigiendo adecuadamente el offset y el escalado.
5. *Tiempos de retardo.* En muchos procesos industriales, los tiempos de retardo no son despreciables y es muy importante describir correctamente el retardo del proceso en el modelo. En los métodos de estimación de parámetros, el retardo es tratado como un desplazamiento temporal de las entradas respecto a las salidas. Una de las señales es utilizada como referencia la otra es desplazada en el tiempo para compensar el retardo. Usualmente, es la entrada la señal desplazada, por ejemplo en el caso de un retardo d :

$$u(t) = q^{-d} \hat{u}(t) = \hat{u}(t-d)\tag{2.93}$$

Utilizando la variable desplazada, el modelo puede ser estimado mediante los métodos estándar.

Bibliografía adicional sobre el tema

Bibliografía adicional sobre el tema

- [Åström71] K.J. Åström and P. Eykhoff. System Identification – A Survey. *Automatica*, Vol. 7 pp. 123-162. Pergamon Press, 1971. Great Britain.
- [Evans92] D.C. Evans, D. Rees, D.L. Jones. Design of test signals for identification of linear systems with nonlinear distortion. *IEEE Trans. on Instrumentation and Measurement*, 16, 6, p. 768-774, 1992.
- [Fasol80] K.H. Fasol, H.P. Jorgl. Principles of models building and identification, *Automatica*, 16, p.501-518, 1980
- [Godfrey80] K.R. Godfrey. Correlation methods. *Automatic*, 16, p. 527-534, 1980

- [Iserman80] R. Iserman. Practical aspects of process identification, *Automatica*, 16, p.575-587, 1980
- [Johansson93] R. Johansson. System modeling and identification. Prentice Hall, 1993
- [Landau90] I.D. Landau. System identification and control design, Prentice Hall, 1990.
- [Ljung87] L.Ljung. System identification – Theory for the user, Prentice Hall, 1987.
- [Ljung94] L.Ljung and T.Glad. Modelling of dynamic system, Prentice Hall, 1994.
- [Ogata87]. K. Ogata. Discrete-time Control Systems. Prentice-Hall International Editions, United States.
- [Söderström89] T. Söderström, P. Stoica. System identification, Prentice Hall, 1989.
- [Rake80] H.Rake. Step response and frequency response methods, Academic Press, 1994.
- [Zhu93] Y. Zhu, T. Backx. Identification of multivariable industrial processes – for simulation, diagnosis and control, Springer-Verlag, 1993.