

OBSERVACIONES SOBRE EL PROBLEMA IDENTIFICACIÓN CON UN NÚMERO REDUCIDO DE DATOS

Miguel E. Peña; Eduardo F. Camacho; Sandra Piñón
 mpena@inaut.unsj.edu.ar; eduardo@esi.us.es; spinon@cartuja.us.es
 Departamento de Sistemas y Automática (ESI)
 Universidad de Sevilla

Resumen

En este trabajo se analizan algunos aspectos que surgen cuando se realiza la identificación de un sistema con un número reducido de datos.

Palabras clave: Identificación, modelado, Sistemas lineales afín a tramos.

1 Introducción

En este trabajo se realizan algunas observaciones sobre distintos aspectos del problema de identificación con un número reducido de datos para un sistema no lineal. La mayoría de las propuestas de identificación, ([9] [10], [12], [11], [3]) se basan en disponer de todos los datos que se necesiten para realizar una buena identificación. Cuando se realiza la identificación de sistemas no lineales, esta suposición no siempre es válida y pueden surgir algunos inconvenientes.

El trabajo se ha organizado como sigue: primero se analiza el muestreo de los datos y se establece una condición necesaria para realizar una buena identificación. A continuación se analiza la influencia del índice de desempeño y de la distribución de los datos en el proceso de identificación con un número limitado de datos. Posteriormente se analiza el problema de la presencia de ruido en la identificación, aclarando algunos conceptos. Por último se define la zona de validez del modelo y se presentan conclusiones.

2 Muestreo de los datos

En esta sección se analiza cuál debe ser la distribución de los datos para una buena identificación. En la figura 1 se observan las muestras que se pueden obtener de una función (saturación) que presenta una transición entre dos valores. En la figura 1 (a) y (b) el muestreo se realiza en forma uniforme en el espacio de entrada. Se puede apreciar que cuando la transición se realiza en forma abrupta (figura 1 (b)) solo un dato se encuentra en la zona de transición. En la figura 1 (c) y (d) los datos se encuentran a la misma distancia mínima

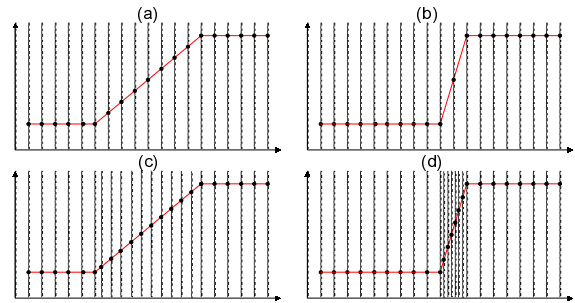


Figura 1: Distinta distribución de los datos para la identificación.

unos de otros en el espacio entrada - salida. En este caso se puede apreciar que existen más datos en el espacio de entrada donde la pendiente es mayor. En base a esta gráfica, intuitivamente se puede decir que tienen que haber más datos en donde la función presenta mayores cambios.

Una pregunta importante en la identificación de sistemas es: ¿qué tan próximos unos de otros tienen que estar los datos en el espacio de entrada del sistema para que un modelo pueda aproximar la ecuación del sistema con un error acotado? Si las muestras no se han obtenido correctamente, cualquier índice de desempeño que se utilice puede producir resultados erróneos en la identificación. A continuación se da una proposición para dar una respuesta a esta pregunta para sistemas de una entrada y una salida. Para ello se parte del teorema de Taylor que dice:

Teorema: De aproximación de Taylor.

Dada una función $f(x)$ continua y diferenciable hasta el orden $n + 1$ en un intervalo $I = (x_{inf}, x_{sup})$ que contiene a x_0 , entonces para todo x en I , $f(x)$ será “bien” aproximada por un polinomio de Taylor de grado n definido como

$$T_n(x, x_0) = \sum_{j=0}^n \frac{1}{j!} \left. \frac{\partial^j f(x)}{\partial x^j} \right|_{x=x_0} (x - x_0)^j$$

en el sentido que

$$f(x) = T_n(x, x_0) + R_n(x, x_0)$$

$$|R_n(x, x_0)| \leq \frac{M|x - x_0|^{n+1}}{(n+1)!}$$

$$M = \max_{z \in I} \left| \frac{\partial^{n+1} f(x)}{\partial x^{n+1}} \right|$$

Este teorema muestra que si en el intervalo (x_{inf}, x_{sup}) existen todas las derivadas de la función hasta la derivada $\partial^{n+1} f(x) / \partial x^{n+1}$, entonces se puede aproximar la función por un polinomio de Taylor de grado n y el error de aproximación está acotado por una cota del módulo de la derivada $n+1$ multiplicada por la distancia desde el valor que se quiere encontrar la función y el punto x_0 .

Proposición: Muestreo de los datos.

Sea $y(x) = f(x)$ la función del sistema a identificar, $\hat{y}(x) = \hat{f}(x)$ la función del modelo que lo identifica, ambas continuas y diferenciables en un intervalo $I = (x_{inf}, x_{sup})$, $e(x) = y(x) - \hat{y}(x)$ el error de identificación, x_i para $i = 1, \dots, k$ las k muestras uniformemente distribuidas en el espacio de entrada, d_x la distancia entre las muestras en el espacio de entrada,

Si $y(x_i) = \hat{y}(x_i)$ y por lo tanto $e(x_i) = 0$ para $i = 1, \dots, k$

Entonces

$$|e(x)| \leq \max_{x \in I} \left| \frac{\partial y(x)}{\partial x} - \frac{\partial \hat{y}(x)}{\partial x} \right| (d_x/2) \text{ para } x \in I \quad (1)$$

$$|e(x)| \leq \left[\max_{x \in I} \left| \frac{\partial y(x)}{\partial x} \right| + \max_{x \in I} \left| \frac{\partial \hat{y}(x)}{\partial x} \right| \right] (d_x/2) \quad (2)$$

Demostración:

Debido a que $y(x)$ y $\hat{y}(x)$ son continuamente diferenciables entonces $e(x)$ también lo es, por lo tanto se puede utilizar un polinomio de Taylor con centro en el dato de entrada x_i para aproximar la función en el intervalo $I_i = [x_i - (d_x/2), x_i + (d_x/2)]$ como

$$e(x) = T_n^e(x, x_i) + R_n^e(x, x_i)$$

Considerando que $e(x_i) = 0$ se puede escribir

$$e(x) = R_n^e(x, x_i)$$

Utilizando una aproximación de primer orden y la cota de los residuos del teorema de Taylor

$$|e(x)| \leq M_i |x - x_i| \text{ para } x \in I_i$$

donde

$$M_i = \max_{x \in I_i} \left| \frac{\partial e(x)}{\partial x} \right|$$

y como $|x - x_i| < (d_x/2)$ resulta

$$|e(x)| < \max_{x \in I_i} \left| \frac{\partial e(x)}{\partial x} \right| (d_x/2)$$

Considerando todos los intervalos para $i = 1, \dots, k$

$$|e(x)| \leq \max_{x \in I} \left| \frac{\partial e(x)}{\partial x} \right| (d_x/2) \quad (3)$$

Como $e(x) = y(x) - \hat{y}(x)$ entonces $\frac{\partial e(x)}{\partial x} = \frac{\partial y(x)}{\partial x} - \frac{\partial \hat{y}(x)}{\partial x}$ y la ecuación (3)

$$|e(x)| \leq \max_{x \in I} \left| \frac{\partial y(x)}{\partial x} - \frac{\partial \hat{y}(x)}{\partial x} \right| (d_x/2)$$

comprobando la ecuación (1).

Como $\left| \frac{\partial y(x)}{\partial x} - \frac{\partial \hat{y}(x)}{\partial x} \right| \leq \left| \frac{\partial y(x)}{\partial x} \right| + \left| \frac{\partial \hat{y}(x)}{\partial x} \right|$ resulta

$$\max_{x \in I} \left| \frac{\partial y(x)}{\partial x} - \frac{\partial \hat{y}(x)}{\partial x} \right| \leq \max_{x \in I} \left[\left| \frac{\partial y(x)}{\partial x} \right| + \left| \frac{\partial \hat{y}(x)}{\partial x} \right| \right]$$

además

$$\max_{x \in I} \left[\left| \frac{\partial y(x)}{\partial x} \right| + \left| \frac{\partial \hat{y}(x)}{\partial x} \right| \right] \dots$$

$$\dots \leq \max_{x \in I} \left| \frac{\partial y(x)}{\partial x} \right| + \max_{x \in I} \left| \frac{\partial \hat{y}(x)}{\partial x} \right|$$

Resulta

$$|e(x)| \leq \left[\max_{x \in I} \left| \frac{\partial y(x)}{\partial x} \right| + \max_{x \in I} \left| \frac{\partial \hat{y}(x)}{\partial x} \right| \right] (d_x/2)$$

comprobándose la ecuación (2). ▼▼▼

La ecuación (1) de esta proposición muestra que si la salida del modelo solo puede igualar a la salida del sistema en los puntos de muestra (que son los únicos que se conocen del sistema), el error de salida en los sectores entre muestras estará acotado por la distancia entre las muestras multiplicada por el máximo error entre las derivadas.

En general no se conoce el error entre las derivadas de antemano, pero sí se pueden conocer cotas de las funciones tanto del sistema como del modelo. En este caso se puede utilizar la ecuación (2).

Si se supone que el método de identificación es lo suficientemente bueno de manera que

$$\max_{x \in I} \left| \frac{\partial y(x)}{\partial x} - \frac{\partial \hat{y}(x)}{\partial x} \right| \leq \max_{x \in I} \left| \frac{\partial y(x)}{\partial x} \right|$$

lo que quiere decir que la cota del error de las derivadas es menor que la máxima derivada de la

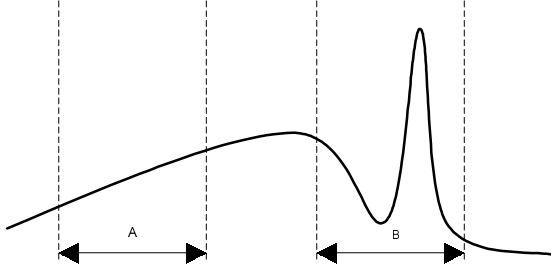


Figura 2: Distinto muestreo para distintas zonas de la función.

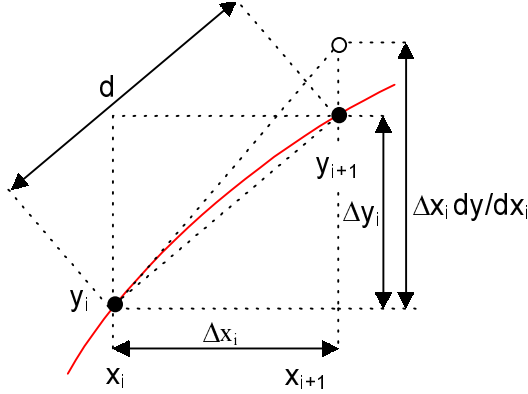


Figura 3: Aproximación de la distancia entre muestras en el espacio de entrada y salida.

función del sistema, por lo que se puede establecer que

$$|e(x)| \leq \max_{x \in I} \left| \frac{\partial y(x)}{\partial x} \right| (d_x/2)$$

y si se fija un error máximo $|e(x)|_{max}$ entre la función del sistema y la del modelo se puede determinar la distancia máxima d_{xmax} a la que deben estar los datos en el espacio de entrada como

$$d_{xmax} \leq \frac{2|e(x)|_{max}}{\max_{x \in I} \left| \frac{\partial y(x)}{\partial x} \right|} \quad (4)$$

Esta ecuación (4) expresa qué tan próximos deben estar los datos en el espacio de entrada para tener una buena aproximación de la función.

Se debe notar que el establecimiento de la distancia entre las muestras es una dependencia local. Para la figura 2 se pueden establecer distintas distancias entre muestras para el intervalo A que para el intervalo B.

Si se considera que las funciones son suaves y por lo tanto no cambia la derivada considerablemente de una muestra a otra resulta

$$d_{x_i} \leq \frac{2|e(x)|_{max}}{\left| \frac{\partial y(x)}{\partial x} \right|_{x_i}} \quad (5)$$

por lo tanto la distancia entre muestras en el espacio de entrada es inversamente proporcional a la derivada de la función. Por otro lado, para que la distancia entre dos puntos en el espacio de entrada - salida sea aproximadamente constante para todas las muestras definidas como

$$d^2 = \Delta^2 x_i + \Delta^2 y_i$$

lo que se muestra en la figura 3, se debe cumplir que

$$d_{x_i} = \frac{d}{\sqrt{1 + \left(\left| \frac{\partial y(x)}{\partial x} \right|_{x_i} \right)^2}} \quad (6)$$

Comparando las ecuaciones (5) y (6) se puede apreciar que para valores $dy/dx \gg 1$ las dos funciones se aproximan. Se puede entender que para que las muestras tengan una buena distribución en el espacio de entrada, estas tienen que estar uniformemente distribuidas en el espacio de entrada - salida y las distancias entre las muestras es función tanto de la derivada máxima de la función como del error máximo admisible.

Si en el muestreo de los datos se pudieran obtener no solo los valores de la función sino también los valores de las n derivadas en los puntos de muestra, si se garantizara que en estos puntos de muestra la función y las n derivadas del sistema y del modelo se igualan, y que además, el método de identificación es lo suficientemente bueno, entonces se puede encontrar la cota para la distancia entre muestras utilizando el teorema de Taylor con más términos en forma similar a la proposición 2. Bajo estas circunstancias resulta la distancia entre muestras

$$d_{xmax} \leq 2 \left(\frac{(n+1)! |e(x)|_{max}}{\max_{x \in I} \left| \frac{\partial^{n+1} y(x)}{\partial x^{n+1}} \right|} \right)^{1/(n+1)}$$

y el error máximo dado una distancia de muestreo como

$$|e(x)|_{max} \leq \frac{\max_{x \in I} \left| \frac{\partial^{n+1} y(x)}{\partial x^{n+1}} \right|}{(n+1)!} (d_{xmax}/2)^{(n+1)} \quad (7)$$

Para el caso multidimensional, el establecer una distancia máxima entre muestras se complica. Se analiza el caso en que los datos están en el espacio de entrada en forma de rejilla uniformemente espaciada como se muestra en la figura 4, por lo tanto para m entradas al sistema

$$d_{x1} = d_{x2} = \dots = d_{xm} = d_x$$

Se puede considerar una aproximación como se muestra en la figura 4. El peor caso ocurre cuando

el gradiente de la función (este es la máxima derivada direccional) de salida respecto a las entradas coincide con la dirección de los vértices del hipercubo que rodea a cada dato del espacio de entrada. Por lo que se puede escribir la ecuación (4) como

$$d_{xmax} \leq \frac{2 |e(x)|_{max}}{\sqrt{m} max_{x \in I} \left\| \frac{\partial y(x)}{\partial x} \right\|}$$

donde $max_{x \in I} \left\| \frac{\partial y(x)}{\partial x} \right\|$ es la norma o módulo máximo del gradiente de la función en todo el espacio de trabajo.

Ejemplo: Muestreo de los datos.

Dada como función del sistema a identificar el polinomio

$$y = 150x^4 - 322.5x^3 + 212.235x^2 - 39.9577x + 0.37868$$

en el intervalo $I = [0, 1]$ y dada una distancia entre muestras $d_x = 0.1$, se pueden calcular las distintas cotas para el error suponiendo que se cumplen las condiciones para aplicar la ecuación (7).

Considerando que solo coincide el valor de la función del sistema y la del modelo en los puntos de muestra, entonces la derivada primera del sistema será

$$dy/dx = 600x^3 - 967.5x^2 + 424.47x - 39.9577$$

donde para el intervalo $I = [0, 1]$ resulta $max_{x \in [0, 1]} |dy/dx| = 39.9577$ y por lo tanto $|e(x)| \leq 1.9979$. Considerando que además coinciden en los puntos de muestra la derivada primera de la función, entonces la derivada segunda del sistema será

$$d^2y/dx^2 = 1800x^2 - 1935x + 424.47$$

donde para el intervalo $I = [0, 1]$ resulta $max_{x \in [0, 1]} |d^2y/dx^2| = 424.47$ y por lo tanto $|e(x)| \leq 0.53059$. Considerando que esto se cumple también para la derivada segunda de la función, entonces la derivada tercera del sistema será

$$d^3y/dx^3 = 3600x - 1935$$

donde para el intervalo $I = [0, 1]$ resulta $max_{x \in [0, 1]} |d^3y/dx^3| = 1935$ y por lo tanto $|e(x)| \leq 0.040313$. Considerando que además coinciden en los puntos de muestra la derivada tercera de la función, entonces la derivada cuarta del sistema será

$$d^4y/dx^4 = 3600$$

donde para el intervalo $I = [0, 1]$ resulta $max_{x \in [0, 1]} |d^4y/dx^4| = 3600$ y por lo tanto $|e(x)| \leq$

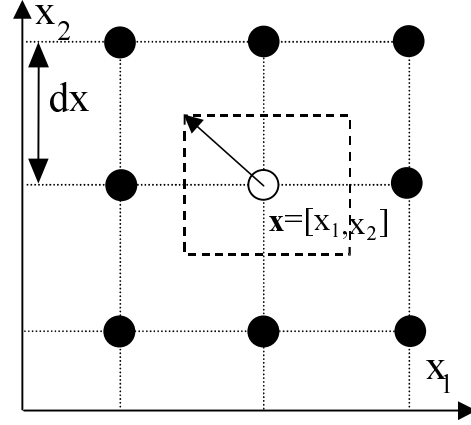


Figura 4: Datos en el espacio de entrada en forma de grilla

0.0009375. Se puede observar que a medida que se considera que en los puntos de muestra se igualan más derivadas, el error máximo de la aproximación disminuye. Esto se puede apreciar en la figura 5.

Si se considera que coinciden en los puntos de muestra hasta la derivada cuarta de la función del sistema entonces el error máximo estaría acotado por la derivada quinta. Como $d^5y/dx^5 = 0$ la cota del error es nula $|e(x)| \leq 0$. Esto es congruente ya que si para una función se anula la derivada $n + 1$ se puede aproximar exactamente (sin error) por medio de un polinomio de Taylor de grado n . Cualquier función que iguale el valor de la función y sus n derivadas en el punto de muestra que no sea el polinomio de Taylor, necesariamente deberá tener derivadas de orden superior a n no nulas y por lo tanto no cumplirá la condición para poder aplicar la ecuación (7) que es que el método de identificación sea lo suficientemente bueno, es decir

$$max_{x \in I} \left| \frac{\partial^n y(x)}{\partial x^n} - \frac{\partial^n \hat{y}(x)}{\partial x^n} \right| \leq max_{x \in I} \left| \frac{\partial^n y(x)}{\partial x^n} \right|$$

3 Influencia del índice de desempeño y de la distribución de los datos.

3.1 Inconvenientes observados

En esta sección se analizan otros inconvenientes en la utilización de los índices de desempeño cuando se trabaja con un número reducido de datos. En la identificación con propuestas por aproximadores universales de funciones (Redes Neuronales, Modelos Borrosos, Funciones de Base Radial, etc.[1],

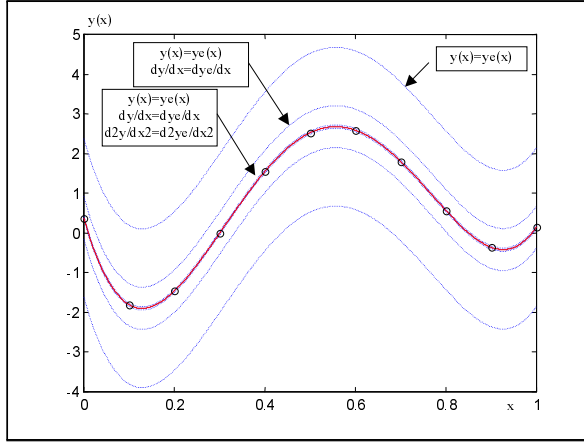


Figura 5: Error de modelado al aumentar el número de derivadas conocidas de la función en los puntos de muestra.

[2], [5], [6], [8]) la estructura del sistema que se pretende identificar no es la misma que la estructura del modelo usado. Si bien estos modelos pueden aproximar cualquier función con el grado de precisión que se desee (son aproximadores universales de funciones), una vez fijada una estructura (número de neuronas, número de reglas, etc.) para el modelo, siempre aparecerá un error de modelado debido a la diferencia entre la estructura del modelo y del sistema.

La diferente distribución de los datos de entrada y salida y la utilización de diferentes índices de desempeño puede provocar una aproximación diferente. A continuación se analiza un ejemplo ilustrativo de estos conceptos.

Si se quiere identificar una función

$$y(i) = x^2(i) \quad (8)$$

Dado un conjunto de datos de entrada $x = [x(i), \dots, x(N)]$ se obtiene los datos de salida $y = [y(i), \dots, y(N)]$. Se aproxima esta función por una recta de la forma

$$y(i) = ax(i) + b \quad (9)$$

Esto implicaría un error en la estructura del modelo y por lo tanto naturalmente aparece un error de modelado.

En la figura 6, se observan los datos de entrada y salida (puntos) con los que se ha realizado la identificación de la función (8), la aproximación por una recta (9) (línea continua) con mínimo *error cuadrático medio*

$$V_{MSE} = \frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i))^2 \quad (10)$$

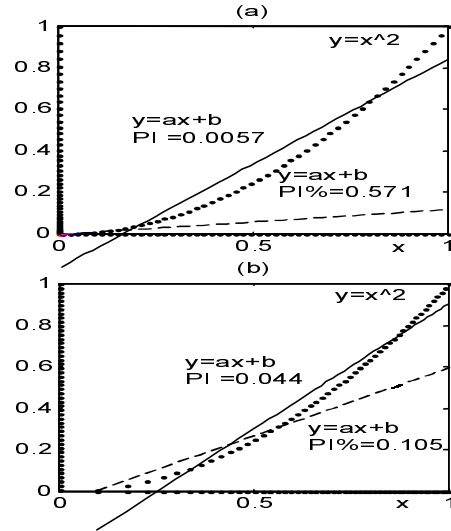


Figura 6: Aproximación de la función $y = x^2$ por la función $y = ax + b$; línea continua, aproximación con mínimo *error cuadrático medio* (10); línea discontinua, aproximación con un mínimo *error cuadrático medio porcentual* (11) (a) datos uniformemente distribuidos en la entrada. (b) datos uniformemente distribuidos en la salida.

($y(i)$ e $\hat{y}(i)$ son la salida medida y la estimada respectivamente) y la aproximación con una recta (9) (línea discontinua) con un mínimo *error cuadrático medio porcentual*

$$V_{PMSE} = \frac{100}{N} \sum_{i=1}^N \left(\frac{y(i) - \hat{y}(i)}{y(i)} \right)^2 \quad (11)$$

En la parte (a) de esta figura se muestran los resultados cuando los datos están uniformemente distribuidos en la entrada y en la figura (b) cuando los datos están uniformemente distribuidos en la salida.

En este ejemplo se puede apreciar que la recta con la que se aproxima la curva no solo depende de la función a identificar. Tanto la distribución de los datos como el índice de desempeño son factores importantes en la “mejor aproximación” de la función y se debe analizar la relación entre la distribución de los datos y el índice de desempeño usado.

A continuación se realiza una propuesta para modificar los índices de desempeño para que el modelo obtenido tenga un mejor comportamiento global. Supóngase que se tiene la función $y = f(x)$ y se quiere estimar por la función $\hat{y} = \hat{f}(x, \theta)$. Se puede definir como índice de desempeño

$$V_{gen} = \|\hat{y}(x, \theta) - y(x)\|_x^2$$

donde $\|\cdot\|$ es una norma en el espacio de funciones continuas. Eligiendo específicamente como norma

$$V_{Int} = \int_x (\hat{y}(x, \theta) - y(x))^2 dx$$

Primero se analiza el caso de entrada - salida simple. Si se discretiza la integral con una aproximación rectangular (orden 0) (ver figura 7 (a)), se obtiene la ecuación

$$V_{Int}^{D0} = \sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x_i))^2 \Delta x_i \quad (12)$$

donde $\Delta x_i = \Delta^{(+)}x_i/2 + \Delta^{(-)}x_i/2$, $\Delta^{(+)}x_i = (x_{i+1} - x_i)$ y $\Delta^{(-)}x_i = (x_i - x_{i-1})$. Suponiendo que los datos tienen una distribución uniforme en el espacio de entrada, entonces $\Delta x_i = \Delta x$, por lo tanto se puede escribir

$$V_{Int}^{D0} = \Delta x \sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x))^2$$

Resulta que el argumento que hace mínimo una función es el mismo que si la misma está multiplicada por una constante positiva, es decir

$$\min_{\theta} \left(\sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x))^2 \right) = \dots$$

$$\dots \min_{\theta} \left(\Delta x \sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x))^2 \right)$$

Se puede proponer un índice genérico como

$$V_G = \sum_{i=1}^N \lambda(i) (y(i) - \hat{y}(i))^2 \quad (13)$$

El parámetro $\lambda(i)$ puede tomar en cuenta distintos factores como la amplitud de la señal de salida, el grado de confianza del dato, etc.

Si se define

$$\lambda(i) = 1/N \quad (14)$$

se obtiene V_{MES} (10). Si se define

$$\lambda(i) = \frac{100}{Ny(i)^2} \quad (15)$$

se obtiene $V_{MES\%}$. (11). Minimizar (12) es equivalente a minimizar (13) con el peso (14). En un proceso de identificación normalmente los datos no están uniformemente espaciados en el espacio de entrada, por lo tanto no es lo mismo la minimización de (12) que de (13, 14).

Para el caso de sistemas con entradas múltiples y salida simple la situación se complica ya que no es tan sencillo realizar la integral en forma numérica. Se analiza el caso en el que los puntos están en el espacio de entrada en forma de rejilla con diferentes longitudes como se muestra en la figura (8). En este caso el índice propuesto con una aproximación de orden 0 es

$$V_{Int}^{D0} = \int_{x_1} \dots \int_{x_m} (\hat{y}(x, \theta) - y(x))^2 dx_1 \dots dx_m$$

$$\approx \sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x_i))^2 \Delta V_{x_i}$$

Donde $x = [x_1, \dots, x_m]^T$, $x_i = [x_{1,i}, \dots, x_{m,i}]^T$, $\Delta V_{x_i} = \Delta x_{1,i} \dots \Delta x_{m,i}$, $\Delta x_{j,i} = \Delta^{(+)}x_{j,i}/2 + \Delta^{(-)}x_{j,i}/2$

En esta ecuación ΔV_{x_i} es una medida del hipervolumen en el espacio de entrada. Se puede observar en esta ecuación que los datos que están más aislados en el espacio de entrada tienen mayor peso al momento de calcular el índice.

Se puede hacer una aproximación de la integral de primer orden (figura (7) (b)); en este caso para una función $f(x)$ su integral aproximada considerando el área de los trapecios resulta

$$V_{Int}^{D1} = \sum_{i=1}^N [f(x_i) \Delta x_i + \dots$$

$$\dots \left(\frac{\partial f(x)}{\partial x} \Big|_{x=x_i} \left(\frac{\Delta^{(+)}x_i}{2} - \frac{\Delta^{(-)}x_i}{2} \right) \right) \frac{\Delta x_i}{2}]$$

considerando que $\Delta x_i = (\Delta^{(+)}x_i + \Delta^{(-)}x_i)/2$ y que $(\Delta^{(+)}x_i)^2 - (\Delta^{(-)}x_i)^2 = (\Delta^{(+)}x_i - \Delta^{(-)}x_i)(\Delta^{(+)}x_i + \Delta^{(-)}x_i)$, entonces

$$V_{Int}^{D1} = \sum_{i=1}^N f(x_i) \Delta x_i + \dots$$

$$\dots \frac{1}{8} \sum_{i=1}^N \frac{\partial f(x)}{\partial x} \Big|_{x=x_i} \left((\Delta^{(+)}x_i)^2 - (\Delta^{(-)}x_i)^2 \right)$$

Para el caso del error de aproximación de funciones resulta

$$f(x) = (\hat{y}(x, \theta) - y(x))^2$$

$$\frac{\partial f(x)}{\partial x} = 2(\hat{y}(x, \theta) - y(x)) \left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x, \theta)}{\partial x} \right)$$

por lo tanto

$$V_{Int}^{D1} = \sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x_i))^2 \Delta x_i$$

$$+ \frac{1}{4} \sum_{i=1}^N \left\{ (\hat{y}(x_i, \theta) - y(x_i)) \left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x)}{\partial x} \right) \Big|_{x=x_i} \right. \\ \left. \left((\Delta^{(+)} x_i)^2 - (\Delta^{(-)} x_i)^2 \right) \right\}$$

En esta ecuación se puede observar que si los datos están uniformemente espaciados entonces $\Delta^{(+)} x_i = \Delta^{(-)} x_i$ y $\Delta x_i = \Delta x$ regresando al índice (13, 14) donde no se tiene en cuenta la distribución de los datos en el espacio de entrada.

Se puede considerar una aproximación de la integral con una serie de Taylor de segundo orden. Para este caso

$$f(x) \approx f(x_i) +$$

$$\frac{\partial f(x)}{\partial x} \Big|_{x=x_i} (x - x_i) + \frac{1}{2} \frac{\partial^2 f(x)}{\partial x^2} \Big|_{x=x_i} (x - x_i)^2$$

Integrando para una muestra resulta

$$\int_{x_i - \Delta^{(-)} x_i}^{x_i + \Delta^{(+)} x_i} f(x) dx = \\ \int_{x_i - \Delta^{(-)} x_i}^{x_i + \Delta^{(+)} x_i} \left[f(x_i) + \frac{\partial f(x)}{\partial x} \Big|_{x=x_i} (x - x_i) + \frac{\partial^2 f(x)}{2 \partial x^2} \Big|_{x=x_i} (x - x_i)^2 \right] dx$$

Para el primer término

$$\int_{x_i - \Delta^{(-)} x_i / 2}^{x_i + \Delta^{(+)} x_i / 2} f(x_i) dx = [f(x_i) x]_{x_i - \Delta^{(-)} x_i / 2}^{x_i + \Delta^{(+)} x_i / 2} \\ = f(x_i) \Delta x_i$$

El segundo término

$$\int_{x_i - \Delta^{(-)} x_i / 2}^{x_i + \Delta^{(+)} x_i / 2} \frac{\partial f(x)}{\partial x} \Big|_{x=x_i} (x - x_i) dx \\ = \left[\frac{1}{2} \frac{\partial f(x)}{\partial x} \Big|_{x=x_i} (x - x_i)^2 \right]_{x_i - \Delta^{(-)} x_i / 2}^{x_i + \Delta^{(+)} x_i / 2} \\ = \frac{1}{8} \frac{\partial f(x)}{\partial x} \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^2 - (\Delta^{(-)} x_i)^2 \right)$$

Lo que es congruente con la regla de los trapecios vista anteriormente. El tercer término es

$$\int_{x_i - \Delta^{(-)} x_i / 2}^{x_i + \Delta^{(+)} x_i / 2} \frac{1}{2} \frac{\partial^2 f(x)}{\partial x^2} \Big|_{x=x_i} (x - x_i)^2 dx = \\ \left[\frac{1}{6} \frac{\partial^2 f(x)}{\partial x^2} \Big|_{x=x_i} (x - x_i)^3 \right]_{x_i - \Delta^{(-)} x_i / 2}^{x_i + \Delta^{(+)} x_i / 2} = \\ \frac{1}{48} \frac{\partial^2 f(x)}{\partial x^2} \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^3 + (\Delta^{(-)} x_i)^3 \right)$$

Sumando para todas las muestras

$$\int f(x) dx = \sum_{i=1}^N f(x_i) \Delta x_i$$

$$+ \sum_{i=1}^N \frac{1}{8} \frac{\partial f(x)}{\partial x} \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^2 - (\Delta^{(-)} x_i)^2 \right) \\ + \frac{1}{48} \frac{\partial^2 f(x)}{\partial x^2} \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^3 + (\Delta^{(-)} x_i)^3 \right)$$

para el error de aproximación resulta

$$f(x) = (\hat{y}(x, \theta) - y(x))^2 \\ \frac{\partial f(x)}{\partial x} = 2 (\hat{y}(x, \theta) - y(x)) \left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x, \theta)}{\partial x} \right) \\ \frac{\partial^2 f(x)}{\partial x^2} = 2 \left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x, \theta)}{\partial x} \right)^2 \\ + 2 \left(\frac{\partial^2 \hat{y}(x, \theta)}{\partial x^2} - \frac{\partial^2 y(x, \theta)}{\partial x^2} \right) \left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x, \theta)}{\partial x} \right)$$

por lo tanto

$$\int f(x) dx = \sum_{i=1}^N (\hat{y} - y)^2 dx_i \\ + \frac{1}{4} \sum_{i=1}^N \left[(\hat{y} - y) \left(\frac{\partial \hat{y}}{\partial x} - \frac{\partial y}{\partial x} \right) \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^2 - (\Delta^{(-)} x_i)^2 \right) \right. \\ \left. + \frac{1}{24} \sum_{i=1}^N \left[\left(\frac{\partial \hat{y}}{\partial x} - \frac{\partial y}{\partial x} \right)^2 \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^3 + (\Delta^{(-)} x_i)^3 \right) \right. \right. \\ \left. \left. + \frac{1}{24} \sum_{i=1}^N \left[\left(\frac{\partial^2 \hat{y}}{\partial x^2} - \frac{\partial^2 y}{\partial x^2} \right) \left(\frac{\partial \hat{y}}{\partial x} - \frac{\partial y}{\partial x} \right) \Big|_{x=x_i} \left((\Delta^{(+)} x_i)^3 + (\Delta^{(-)} x_i)^3 \right) \right] \right]$$

En esta ecuación se han realizado algunas simplificaciones de la notación ($\hat{y}(x, \theta) = \hat{y}$, $y(x, \theta) = y$). Si los datos están uniformemente espaciados resulta

$$\int f(x) dx = \sum_{i=1}^N (\hat{y}(x_i, \theta) - y(x_i))^2 \Delta x \\ + \frac{1}{96} \Delta x^2 \sum_{i=1}^N \left[\left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x, \theta)}{\partial x} \right)^2 \Big|_{x=x_i} \right] \Delta x \\ + \frac{1}{96} \Delta x^3 \sum_{i=1}^N \left[\left(\frac{\partial^2 \hat{y}(x, \theta)}{\partial x^2} - \frac{\partial^2 y(x, \theta)}{\partial x^2} \right) \left(\frac{\partial \hat{y}(x, \theta)}{\partial x} - \frac{\partial y(x, \theta)}{\partial x} \right) \Big|_{x=x_i} \right]$$

En esta ecuación se aprecia que el primer término es la suma de los errores entre la salida estimada y la medida al cuadrado (similar al índice de error tradicional); el segundo término es el error entre las derivadas de la salida del modelo y la medida al cuadrado; y el tercero es el producto entre el error de la derivadas primera y segunda. Se lo puede interpretar como el índice tradicional con el agregado de distintos términos de regularización del modelo [4], [7].

Estos índices de desempeño no pueden ser directamente usados en la identificación. Primero los datos no necesariamente están distribuidos en el espacio de entrada en forma de rejilla y segundo no se posee a priori la información de las derivadas sucesivas de la salida respecto de la entrada. Para salvar estos aspectos se puede realizar un preprocesamiento de los datos para obtener una medida lo suficientemente buena de estos valores. En

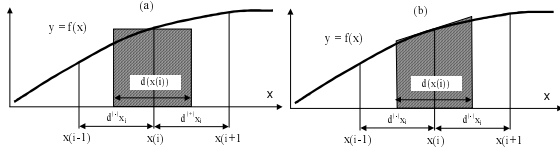


Figura 7: Aproximación discreta de la integral.

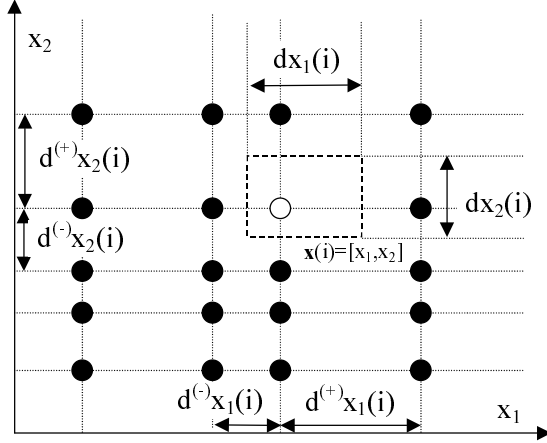


Figura 8: Distribución de los datos en forma de rejilla no uniformemente espaciada.

la sección siguiente (sección 3.2) se describen algunos algoritmos para obtener una aproximación de estos valores.

3.2 Propiedades locales de los datos

En esta sección se analizan algunas propiedades e índices que caracterizan en forma local a cada dato. Algunas de estas propiedades se pueden utilizar como una aproximación en los índices desarrollados en la secciones anteriores.

Se detalla cómo obtener en forma numérica estimaciones locales para cada dato de los valores de las derivadas parciales y de la distancia entre los datos para cada coordenada.

Dado un dato en el espacio de entrada - salida $[\mathbf{x}(i), y(i)]$, donde $\mathbf{x}(i) = [x_1(i), \dots, x_m(i)]$ es el vector de entradas, m es el número de entradas e $y(i)$ es la salida correspondiente, se pueden obtener propiedades locales a dicho dato.

Definición 1: Datos δ próximos en el espacio de entrada: $\mathbf{x}^{i,\delta}(j)$

Son los datos de muestra en el espacio de entrada de un entorno δ del dato de muestra $\mathbf{x}(i)$. Por lo tanto, $\mathbf{x}(j)$ es un dato δ próximo de $\mathbf{x}(i)$, deno-

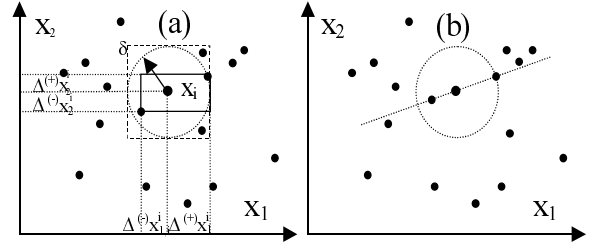


Figura 9: Propiedades locales de los datos

tado por $\mathbf{x}^{i,\delta}(j)$, si $\|\mathbf{x}(i) - \mathbf{x}(j)\|_A \leq \delta$, donde $\mathbf{x}^{i,\delta}(j) = [\mathbf{x}_1^{i,\delta}(j), \dots, \mathbf{x}_m^{i,\delta}(j)]^T$ y $\|\cdot\|_A$ es una norma en el espacio de entrada.

En la figura 9 (a) existen dos datos δ próximos del dato $\mathbf{x}(i)$ si se considera la norma 2 y cuatro si se considera la norma infinito.

Definición 2: Número de datos δ próximos en el espacio de entrada: $k^{i,\delta}$

Definido un entorno δ al dato $\mathbf{x}(i)$ es el número total de datos en el espacio de entrada δ próximos a dicho dato.

Definición 3: Datos k próximos en el espacio de entrada: $\mathbf{x}^{i,k}(j)$

Son los k datos de muestra más próximos en el espacio de entrada del dato de muestra $\mathbf{x}(i)$. La distancia se mide por $\Delta_{i,j} = \|\mathbf{x}(i) - \mathbf{x}(j)\|_A$.

Definición 4: Radio del entorno de los datos k próximos en el espacio de entrada: $\delta^{i,k}$

Es la máxima distancia entre los datos k próximos en el espacio de entrada al dato $\mathbf{x}(i)$, por lo tanto

$$\delta^{i,k} = \max_{j=1,\dots,k} \Delta_{i,j}$$

Se debe notar que un entorno alrededor de cada dato de entrada $\mathbf{x}(i)$ se puede determinar tomando los k datos más próximos o definiendo la distancia máxima de los datos que pertenecen al entorno δ . Por lo tanto si se define δ y se obtiene $k^{i,\delta}$ y es equivalente a definir k y obtener $\delta^{i,k}$.

Definición 5: Distancia por coordenada de los datos k próximos en el espacio de entrada $\Delta x_{l,i}^{i,k}$

Es la mínima distancia por coordenada que incluye a los datos k próximos ($\mathbf{x}^{i,k}(j)$) en el espacio de entrada, por lo tanto

$$\Delta x_{l,i}^{i,k} = \max \left(x_l(i), x_l^{i,k}(1), x_l^{i,\delta}(2), \dots, x_l^{i,\delta}(k) \right) - \min \left(x_l(i), x_l^{i,k}(1), x_l^{i,\delta}(2), \dots, x_l^{i,\delta}(k) \right)$$

para $l = 1, \dots, m$

Definición 6: Distancia positiva por coordenada de los datos k próximos en el espacio de entrada: $\Delta^{(+)}x_l^{i,k}$

Es la mínima distancia que incluye a los datos k próximos por coordenada para valores mayores a $x(i)$, por lo tanto

$$\Delta^{(+)}x_l^{i,k} = \max\left(x_l(i), x_l^{i,k}(1), x_l^{i,k}(2), \dots, x_l^{i,k}(k)\right) - x_l(i)$$

para $l = 1, \dots, m$

Definición 7: Distancia negativa por coordenada de los datos próximos en el espacio de entrada: $\Delta^{(-)}x_l^{i,k}$

Es la mínima distancia que incluye a los δ datos próximos por coordenada para valores mayores a x_i , por lo tanto

$$\Delta^{(-)}x_l^{i,k} = x_l(i) - \min\left(x_l(i), x_l^{i,k}(1), x_l^{i,k}(2), \dots, x_l^{i,k}(k)\right)$$

para $l = 1, \dots, m$

La figura 9 (a) muestra estos conceptos cuando $m = 2$.

Definición 8: Hipervolumen de los k datos próximos en el espacio de entrada: $V^{i,k}$

Es el volumen encerrado por los datos k próximos del dato $x(i)$ (incluido el mismo) en el espacio de entrada, por lo tanto

$$V^{i,k} = \prod_{l=1}^m \Delta x_l^{i,k}$$

En la figura 9 (a) con línea continua de datos se encierra el hipervolumen en el espacio de entrada para $k = 2$ con la norma 2.

Definición 9: Hipervolumen medio de los datos próximos en el espacio de entrada: $\bar{V}^{i,k}$

Es el hipervolumen medio que corresponde a cada dato k próximo del dato $x(i)$. Se puede obtener como

$$\bar{V}^{i,k} = V^{i,k} / (k + 1)$$

Definición 10: Distancia promedio por coordenada de los datos k próximos en el espacio de entrada

Es la distancia que correspondería a cada dato si los mismos estuvieran en forma de grilla en el espacio de entrada. Se puede definir como

$$\bar{\Delta}x_l^{i,k} = \Delta x_l^{i,k} / \sqrt[m]{k + 1}$$

para $l = 1, \dots, m$.

Se debe notar que

$$\prod_{l=1}^m \bar{\Delta}x_l^{i,k} = \prod_{l=1}^m \frac{\Delta x_l^{i,k}}{\sqrt[m]{k + 1}} = \frac{1}{k + 1} \prod_{l=1}^m \Delta x_l^{i,k} = V^{i,k}$$

Definición 11: Modelo lineal local de los datos k próximos en el espacio de entrada: $y_{int}^i = (\mathbf{a}_{int}^i)^T \mathbf{x} + b_{int}^i$

Es el modelo lineal (hiperplano)

$$y_{int}^i = (\mathbf{a}_{int}^i)^T \mathbf{x} + b_{int}^i$$

para el cual es mínimo el error cuadrático medio entre la salida de los datos k próximos y la salida de dicho modelo para los datos k próximos. Por lo tanto se tienen que encontrar los parámetros \mathbf{a}_{int}^i , b_{int}^i que minimizan el funcional

$$J = \sum_{j=1}^k (y(\mathbf{x}(i)) - y_{int}(\mathbf{x}(i)))^2 + \sum_{j=1}^k (y(\mathbf{x}^{i,k}(j)) - y_{int}(\mathbf{x}^{i,k}(j)))^2$$

Para encontrar esta función lineal se puede definir

$$\mathbf{y}^i = [y(\mathbf{x}(i)) \ y(\mathbf{x}^{i,k}(1)) \ \dots \ y(\mathbf{x}^{i,k}(k))]^T$$

$$\mathbf{X}^i = \left[\begin{bmatrix} \mathbf{x}(i) \\ 1 \end{bmatrix} \ \begin{bmatrix} \mathbf{x}^{i,k}(1) \\ 1 \end{bmatrix} \ \dots \ \begin{bmatrix} \mathbf{x}^{i,k}(k) \\ 1 \end{bmatrix} \right]^T$$

$$\mathbf{p}^i = \begin{bmatrix} \mathbf{a}_{int}^i \\ b_{int}^i \end{bmatrix}$$

por lo tanto $\mathbf{y}^i = \mathbf{X}^i \mathbf{p}^i$ y los parámetros se encuentran por mínimos cuadrados como

$$\mathbf{p}^i = (\mathbf{X}^{iT} \mathbf{X}^i)^{-1} \mathbf{X}^{iT} \mathbf{e}_i$$

Definición 12: Derivada de la salida respecto de la entrada del modelo lineal local de los datos k próximos.

Si se deriva la salida del modelo lineal local respecto de las entradas se obtiene que

$$\frac{\partial y_{int}^i(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}_{int}^i$$

Definición 13: Número mínimo de datos k próximos con información de la derivada de la salida respecto de las entradas: $k_{\zeta, \min}^i$.

Cuando se obtiene la derivada de la salida respecto de la entrada del modelo local lineal, los datos en el espacio de entrada tienen que poseer información en todas las direcciones de dicho espacio. El número mínimo necesario de puntos próximos es igual a la dimensión del espacio de entrada, es decir $k_{\min}^{iNecesario} = m$.

No siempre los $k_{\min}^{iNecesario}$ puntos más próximos en el espacio de entrada tienen información de todas las derivadas parciales de la salida respecto de la

entrada. Por ejemplo para 2 entradas $k_{min}^{iNecesario}$ es 2. En la figura 9 (b) se observa que tanto el dato como los 2 más próximos se encuentran sobre una recta. En este caso solo se puede determinar la derivada parcial de la salida sobre el vector que corresponde a la dirección de dicha recta. No se puede determinar la derivada parcial de la salida respecto de las entradas.

Este problema ocurre cuando hay una dependencia lineal entre los datos de entrada. Cuando existe dependencia lineal la matriz $\mathbf{X}^{iT}\mathbf{X}^i$ no tiene inversa y esta dependencia lineal se puede medir con el rango de dicha matriz.

Dado un $\zeta > 0$, un número pequeño se dice que el número mínimo de datos k próximos con información de todas las derivadas de la salida respecto de las entradas, $k_{\zeta, min}^i$ es el mínimo número de datos k próximos que hacen que el mínimo autovalor de la matriz $\mathbf{X}^{iT}\mathbf{X}^i$ sea mayor que ζ , lo que quiere decir que el rango de la matriz $\mathbf{X}^{iT}\mathbf{X}^i$ es $m + 1$.

El procedimiento a seguir para estimar las propiedades locales de cada dato $\mathbf{x}(i)$ es fijado $\zeta > 0$ y un k (o un δ)

- 1) Encontrar los datos k (o los datos δ) próximos de $\mathbf{x}(i)$
- 2) Formar la matriz $\mathbf{X}^{iT}\mathbf{X}^i$
- 3) Si el mínimo autovalor de $\mathbf{X}^{iT}\mathbf{X}^i$ es menor que ζ aumentar en uno k (o incrementar el valor de δ) y regresar al paso 1.
- 4) Si el mínimo autovalor de $\mathbf{X}^{iT}\mathbf{X}^i$ es mayor que ζ obtener los valores de los parámetros locales al dato $\mathbf{x}(i)$.

Basado en la propiedad local de los datos se puede plantear utilizar un índice de desempeño en donde se tenga en cuenta el grado de aislamiento de cada dato.

3.3 Ejemplo demostrativo

En la figura 11 se muestra el desempeño de la identificación de la función por (8) por la función (9) para 5 distribuciones distintas de probabilidad de los datos de la entrada (mostradas en la figura 12). Al comparar estos resultados con los mostrados en la figura 10 que surgen de usar el índice tradicional (10), se puede observar que la recta de aproximación permanece casi invariable para las distintas funciones de probabilidad.

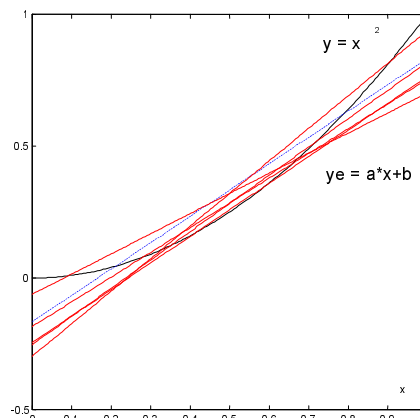


Figura 10: Aproximación de la función $y = x^2$ por la función $y = ax + b$ con el índice (10) para distintas distribuciones de probabilidad de los datos de entrada.

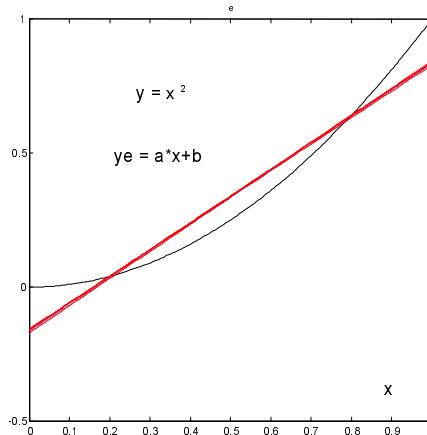


Figura 11: Aproximación de la función $y = x^2$ por la función $y = ax + b$ con el índice (12) para distintas distribuciones de probabilidad de los datos de entrada.

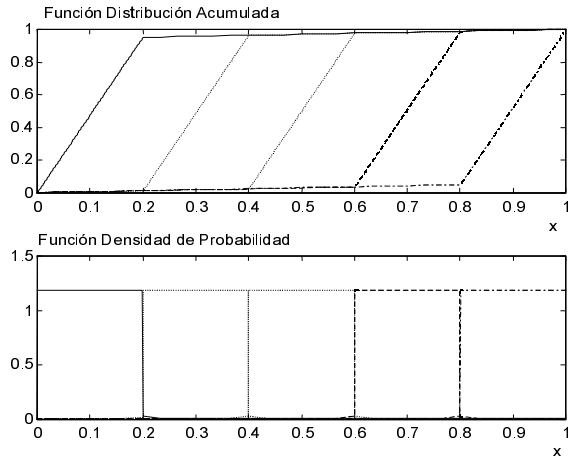


Figura 12: Función de distribución y densidad de probabilidad de los datos usados para la identificación de la función $y = x^2$

4 Influencia por la presencia de ruido

El análisis que se realiza en la sección anterior supone que el sistema no está influenciado por perturbaciones externas al sistema.

Puede haber diferentes factores para que aparezca ruido en un sistema. El ruido en un sistema se presenta por una carencia de información. El más común es el producido por la naturaleza de los sensores. Un ejemplo de este son las tensiones eléctricas espúreas producidas por la agitación térmica. Para el caso del ruido térmico, si se pudiera determinar o medir la señal producida por la agitación térmica, esta se podría transformar en una entrada del sistema y en este caso el sistema dejaría de ser ruidoso.

Una mala determinación del conjunto de entradas del sistema puede verse a la salida como si el sistema estuviera afectado por ruido. En la figura 13 y 14 se observan datos muestreados de la función

$$y = \sin(x_1) + 0.1\cos(x_2)$$

donde $x_1 \in [0, 2\pi]$ y $x_2 \in [0, 4\pi]$. En la figura (13) se observa la gráfica de $y = f(x_1)$. Nótese que para cada valor de la entrada x_1 existen múltiples valores para la salida y . En la figura (14) se puede apreciar la verdadera forma de la función.

Por lo tanto cuando se desconocen la entradas del sistema, este aparenta ser ruidoso y se hace complicado establecer la estructura del modelo.

Este tipo de inconveniente no se puede salvar fácilmente cuando se desconocen las entradas del sistema ya que la identificación de las entradas de un sistema depende de la estructura del modelo.

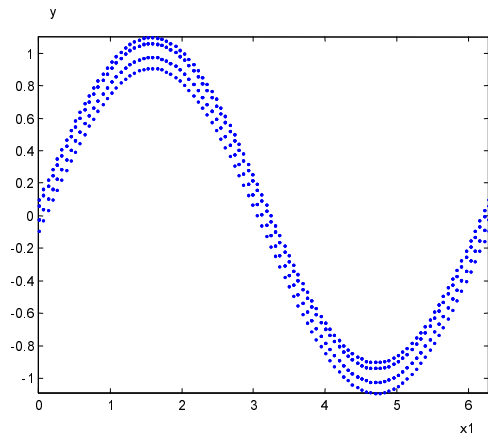


Figura 13: Función $y = \sin(x_1) + 0.1\cos(x_2)$. Gráfica $y = f(x_1)$.

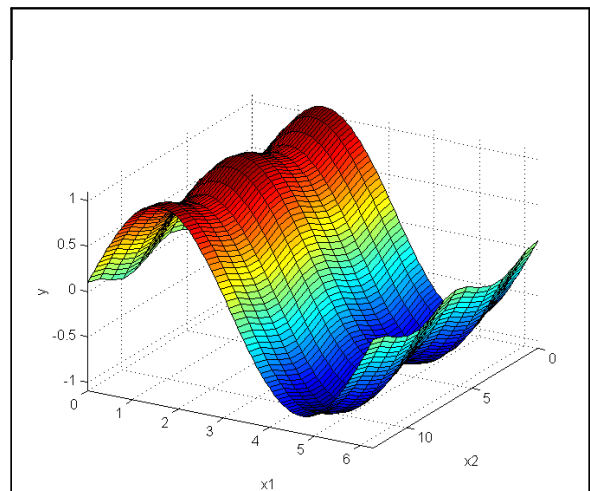


Figura 14: Función $y = \sin(x_1) + 0.1\cos(x_2)$. Gráfica $y = f(x_1, x_2)$.

En esta sección se entiende que hay presencia de *ruido* en un sistema cuando existe ruido de medición el cual no está correlacionado con ninguna variable del sistema y puede ser modelado como una variable aleatoria. Para analizar el problema de la existencia de ruido de medición en el sistema se presenta un ejemplo.

Considérese un sistema $y = f(x)$ donde $f(x) = \sin(x)$ para el que se han realizado mediciones tanto en la entrada como en la salida. Estas mediciones están afectadas por ruido aditivo tanto a la entrada r_x como a la salida r_y . Las mediciones realmente obtenidas son $x_r = x + r_x$ y $y_r = x + r_y$. Un esquema general se puede apreciar en la figura 15.

Suponiendo que a) solo existe ruido de medición en la salida del sistema (en la figura 15 se supone que $r_x = 0$), b) se toman muestras en valores discretos de la entrada, c) se toman los suficientes datos, d) el ruido en la salida r_y es no correlacionado con la entrada y tiene una distribución normal con media cero, entonces las muestras obtenidas tomarían la forma mostrada en la figura 16 (a). En este caso para poder obtener una buena identificación se debería primero promediar todos los valores de salida para cada valor discreto de entrada x_i (para $i = 1, \dots, m$) como

$$\hat{y}_{xi} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{xi}^j \quad (16)$$

obteniendo un valor promedio de la salida para cada muestra (donde n_i es el número de muestras de la salida para cada valor de la entrada). El índice para medir el error se puede expresar como

$$V = \frac{1}{m} \sum_{i=1}^m (y_{xi} - \hat{y}_{xi})^2 \quad (17)$$

reemplazando (16) en (17).

Si se cumple a) a d) entonces el valor estimado para cada salida tenderá al valor real del mismo y_{xi}^* , por lo tanto

$$V = \frac{1}{m} \sum_{i=1}^m (y_{xi} - y_{xi}^*)^2$$

Minimizar (17) con la consideración (16) es equivalente a minimizar el error cuadrático medio de la salida (10). Este es el índice más comúnmente usado en la identificación de sistemas.

De este ejemplo se puede deducir que cuando se minimiza el error cuadrático medio de la salida (10) se deberían tener en cuenta las consideraciones a) a d), especialmente la que indica que solo existe ruido en la salida.

En la figura 16 (b) se muestran los resultados cuando existe ruido tanto en la entrada como en la salida, para datos muestreados en el espacio de entrada (sin ruido) bien separados. En este caso se podría realizar una doble promediación tanto para los datos de la salida como para los datos de la entrada, y luego aplicar un criterio de error cuadrático medio para ajustar la curva. Para este caso

$$y_i = \frac{1}{n} \sum_{j=1}^n y_i^j \quad (18)$$

$$x_i = \frac{1}{n} \sum_{j=1}^n x_i^j \quad (19)$$

y el índice quedaría

$$V = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}(x_i))^2 \quad (20)$$

Reemplazando (18) y (19) en (20)

$$V = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{n} \sum_{j=1}^n y_i^j - \hat{y} \left(\frac{1}{n} \sum_{j=1}^n x_i^j \right) \right)^2 \quad (21)$$

Si ahora se realiza otra consideración e) en la que se añade que el modelo sea lineal (no el caso de la función $f(x) = \sin(x)$) es decir que $\hat{y}(x) = a_0 + a_1 x$ entonces (21) resulta

$$V = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{n} \sum_{j=1}^n y_i^j - \frac{1}{n} \sum_{j=1}^n (a_0 + a_1 x_i^j) \right)^2$$

como $\hat{y}(x_i^j) = a_0 + a_1 x_i^j$

$$V = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{n} \sum_{j=1}^n (y_i^j - \hat{y}(x_i^j)) \right)^2 \quad (22)$$

Por lo tanto aún cuando el sistema es lineal, el índice del error cuadrático medio en la salida no es equivalente al índice de desempeño (20) con (18) y (19).

La situación anterior se complica cuando en la entrada los valores datos muestreados (sin ruido) no están bien separados como se muestra en la figura 16 (c).

Una situación más real ocurre cuando los datos en el espacio de entrada no están suficientemente separados pero la amplitud del ruido (su varianza o desviación estándar) es igual tanto a la entrada como a la salida. En este caso como no se conoce cuál es el punto verdadero sobre la curva $y = f(x)$,

se puede considerar que cada muestra proviene del punto más próximo sobre la curva. El índice de desempeño para una correcta identificación debería ser de la forma

$$V = \sum_{i=1}^N d[(x_i, y_i), (x, f(x))] \quad (23)$$

donde $d[(x_i, y_i), (x, f(x))]$ es la distancia desde el dato de muestra (x_i, y_i) hasta el punto más próximo sobre la curva $y = f(x)$. Para que el punto (x_{if}, y_{if}) (donde $y_{if} = f(x_{if})$) sea el punto perteneciente a la función $y = f(x)$ más próximo al punto (x_i, y_i) , es condición necesaria que la dirección de la derivada de la función en el punto (x_{if}, y_{if}) sea perpendicular con el vector que une al punto (x_{if}, y_{if}) con el punto (x_i, y_i) , por lo tanto

$$\begin{bmatrix} x_i - x_{if} & y_i - y_{if} \end{bmatrix} \begin{bmatrix} 1 \\ dy(x)/dx|_{x=x_{if}} \end{bmatrix} = 0 \quad (24)$$

En estas circunstancias el problema de optimización para encarar la identificación es encontrar el conjunto de parámetros del modelo que minimicen el índice de desempeño

$$V = \sum_{i=1}^N d[(x_i, y_i), (x_{if}, y_{if})] \quad (25)$$

agregando a las variables de optimización los puntos (x_{if}, y_{if}) para $(i = 1 \text{ a } N)$ y las restricciones

$$(x_i - x_{if}) + (y_i - y_{if}) \frac{dy(x)}{dx} \Big|_{x=x_{if}} = 0 \quad (26)$$

$$y_{if} = f(x_{if}) \quad (27)$$

Este problema de optimización es bastante complicado y solo tiene una solución práctica para casos sencillos.

Una situación normal es que las distintas variables de entrada y salida estén afectadas por ruido de distinta magnitud. En este caso la distancia entre el dato de muestra y la función se ve afectada por este hecho. En la figura 18 se representan con elipses (para diferenciarlos con los círculos de la figura 17) marcadas con círculos alrededor de los puntos sobre la función).

Considerando que la amplitud del ruido (desviación estándar) de la entrada es R_x y la de la salida R_y el índice resultante es

$$V = \sum_{i=1}^N d[(x_i, y_i), (x_{if}, y_{if})]_R$$

donde

$$d[(x_i, y_i), (x_{if}, y_{if})]_R = \sqrt{\frac{1}{R_x^2} (x_i - x_{if})^2 + \frac{1}{R_y^2} (y_i - y_{if})^2}$$

y la condición que se debe cumplir como restricción¹ es

$$\begin{bmatrix} x_i - x_{if} & y_i - y_{if} \end{bmatrix} \begin{bmatrix} \frac{1}{R_x} & 0 \\ 0 & \frac{1}{R_y} \end{bmatrix} \begin{bmatrix} 1 \\ \frac{dy(x)}{dx} \Big|_{x=x_{if}} \end{bmatrix} = 0 \quad (28)$$

En la figura 19 se muestra que en los casos extremos se debe medir la distancia del punto de muestra a la función solo en x cuando $R_y = 0$ y solo en y cuando $R_x = 0$. Este último es el caso normalmente usado para la identificación de sistemas.

En la figuras 20 a 22 se observan los resultados de la identificación de un conjunto de datos de la función $y = A \sin(Wx + \phi)$. En este ejemplo para cada uno de los 20 datos uniformemente distribuidos en x sobre el intervalo $[0, \pi/2]$ se tomaron 10 muestras con ruido aleatorio con distribución normal tanto en la entrada como en la salida de la misma amplitud. De ello surgen las 200 muestras con las que se realiza el experimento.

En la figura 20 se observa el resultado de la identificación considerando el índice tradicional, es decir que $R_x = 0$. En este y lo que se minimiza es la suma de las distancias en la salida y . El índice resultante es

$$V = \sum_{j=1}^n \|y_i - \hat{y}_i\|$$

En la figura 21 se minimiza con el índice propuesto en (23), (27), (28) considerando $R_x = R_y$. En este caso, lo que se minimiza es la suma de las distancias mínimas entre los puntos de muestra y la curva.

En la figura 22 se minimiza con el índice propuesto en (23), (27), (28) considerando $R_y = 0$ y $R_x \neq 0$. En este caso, lo que se minimiza es la suma de distancias entre los puntos de muestra y la curva en forma paralela al eje x .

De estas tres figuras se concluye que la curva identificada que mejor se ajusta a la real es la que corresponde al índice propuesto en (23), (27), (28) en donde se considera $R_x = R_y$. Se debe notar que para aplicar este índice se debe disponer de una estima de las potencias de ruido de cada una de las variables involucradas en el proceso.

¹Para no tener problemas numéricos, se pueden dividir los vectores en su norma.

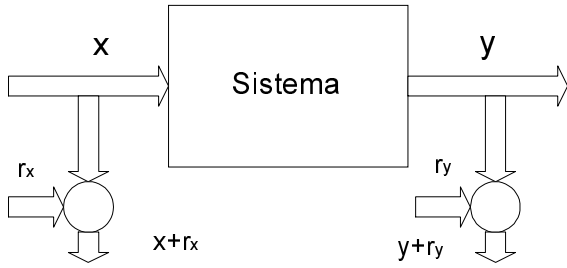


Figura 15: Sistema con presencia de ruido en la entrada y en las salidas

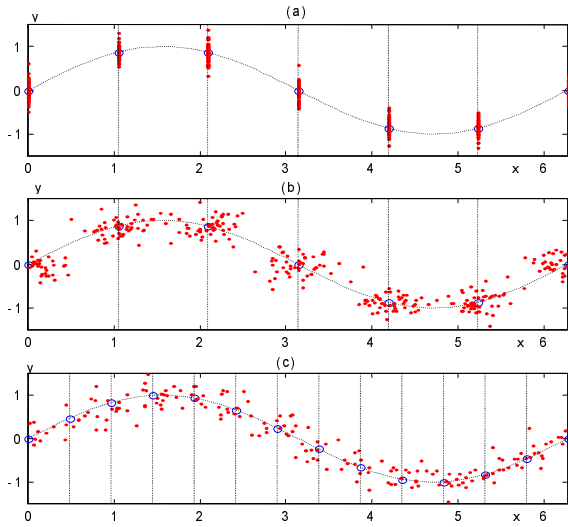


Figura 16: Influencia del ruido. (a) Presencia de ruido en la salida. (b,c) Presencia de ruido en la entrada y en la salida

Si bien los resultados de identificar con un índice (23), (27), (28) presenta mejores resultados, el problema computacional crece considerablemente al necesitar mayores recursos de cálculo y mejores algoritmos de optimización. Esto hace que con los recursos normalmente disponibles solo se puedan resolver problemas sencillos.

5 Definición de zona de validez del modelo

En los sistemas reales, las señales de los mismos están acotadas. Las cotas se presentan tanto en las entradas, en las salidas y en los estados. Las entradas de los sistema son acciones de control que provienen de un actuador cuyos valores están limitados por el sistema físico del mismo. Todo los sistemas, aún los que tienen característica integral, poseen una salida acotada después de cierto tiempo. Por último, cuando los estados tienen un

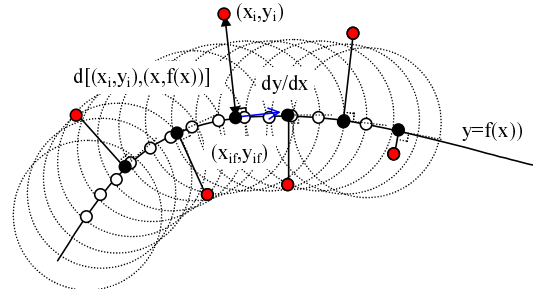


Figura 17: Distancia cuando $R_x = R_y$.

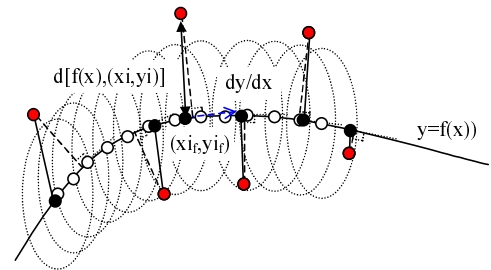


Figura 18: Distancia cuando $R_x \neq R_y$.

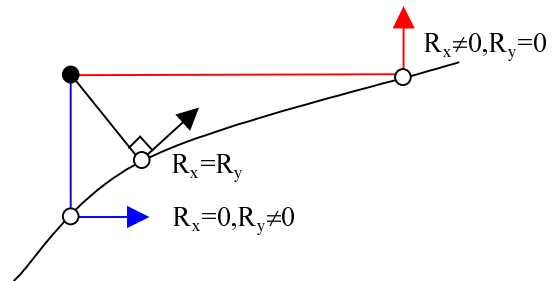


Figura 19: Casos límites en la aplicación del índice propuesto.

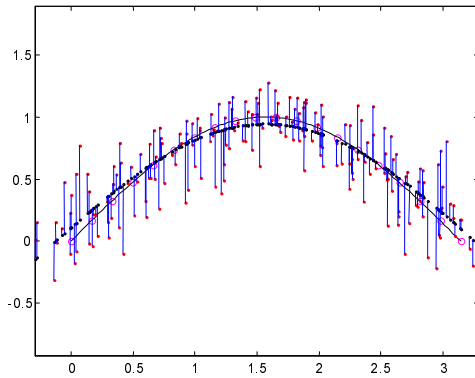


Figura 20: Resultados de la identificación considerando $R_x = 0$. Valores estimados: $A = 0.94846$, $W = 0.92859$, $\phi = 0.11944$.

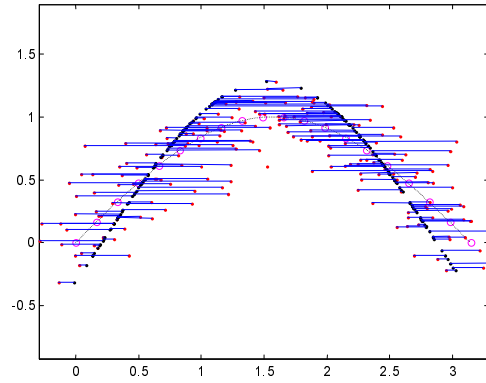


Figura 22: Resultados de la identificación considerando $R_x = 0$. Valores estimados: $A = 1.2874$, $W = 1.1757$, $\phi = -0.23507$.

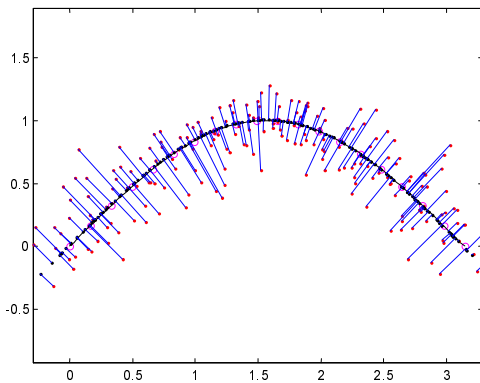


Figura 21: Resultados de la identificación considerando $R_x = R_y$. Valores estimados: $A = 1.0102$, $W = 0.99772$, $\phi = 0.019404$.

significado físico, estos también están acotados.

Estas cotas en general son conocidas, por lo menos en un rango amplio y deben estar contempladas en el modelo matemático que representa el sistema. El rango de variación de las señales de entrada determinan el espacio de entrada del sistema.

En general es muy difícil que las señales de entrada de un sistema cubran todo el espacio de entrada. Por ejemplo, cuando se utilizan como entradas un regresor de un sistema dinámico con valores pasados de la entrada y con un período de muestreo pequeño, los datos se encuentran correlacionados entre sí sobre la recta $x(k) = x(k-1)$. En la figura 23 se muestra esta condición. Cuando se realiza la identificación, el modelo es válido estrictamente en la zona del espacio de entrada donde las señales

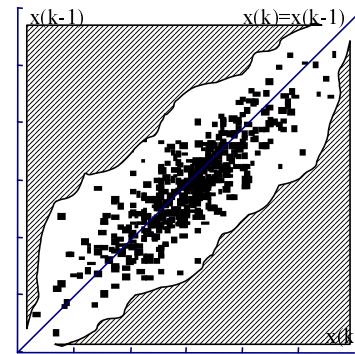


Figura 23: Datos de entrada correlacionados.

de entrada han excitado al sistema. En la figura 23 la zona del espacio de entrada rallada no ha sido excitada por las señales de entrada. Se denomina *Zona de Validez del Modelo* a la región del espacio de entrada en donde se han encontrado datos durante el proceso de identificación. Los modelos obtenidos por un proceso de identificación, deben especificar bajo qué condiciones se realizó la identificación estableciendo su zona de validez.

Ejemplo:

En la figura 24 se muestra la identificación de la función

$$y = x_1^3 + x_2^2$$

para dos conjuntos de datos de entrada diferentes por un modelo lineal a tramos [8]. En la figura 24 (a) se observa la función original identificada. En la figura 24 (b) se muestra el modelo lineal a tramos resultado de la identificación con un conjuntos de datos distribuidos en forma uniforme en el espacio $[-1,1] \times [-1,1]$. Se puede observar que la

aproximación lineal a tramos presenta pequeños errores con la función original. En la figura 24 (c) se observa el resultado de la identificación con un conjunto de datos que solo cubren una parte del espacio de entrada. En este caso se puede definir la zona de validez del modelo por el conjunto de regiones en las que por lo menos hay un dato. En la figura 24 (d) se muestra con rayas la zona en el espacio de entrada en donde los datos no son válidos. En este caso, identificación con un modelo lineal a tramos, la zona de validez se puede indicar con un dato auxiliar por cada región que indique si han habido suficientes datos en la misma durante el proceso de identificación.

Conclusiones

En este trabajo se analizan algunos inconvenientes que surgen en la identificación con un conjunto limitado de datos. Primero se analiza cómo debe ser el muestreo de los datos para obtener una buena identificación. Debido a la “maldición de la dimensionalidad” se puede comprobar que generalmente el conjunto de datos está mal condicionado para realizar la identificación.

Posteriormente se analiza la influencia del índice de desempeño y de la distribución de los datos en el proceso de identificación cuando se trabaja con un número reducido de datos. La diferente distribución de los datos de entrada y salida y la utilización de diferentes índices de desempeño puede provocar una aproximación diferente. Se puede observar que los índices de desempeño que comúnmente se utilizan desprecian la información contenida en los datos aislados. Muchas veces esta es la única información con la que se cuenta en algunas regiones del espacio de entrada del sistema. Un factor de peso diferenciando los datos puede mejorar la identificación del sistema.

A continuación se analiza la influencia en el índice de desempeño de la presencia de ruido de medición en el sistema. Se verifica que los índices normalmente usados no contemplan la presencia de ruido en las entradas del sistema. Se propone una metodología para tener en cuenta esta influencia.

Agradecimientos

Los autores agradecen a los Ministerios de Educación y Cultura y Deporte y de Ciencia y Tecnología (España) y al Consejo Nacional Investigaciones Científicas y Tecnológicas (Argentina) por financiar este trabajo bajo los proyectos SB2000-0285 and DPI 2002-04375-C03-01

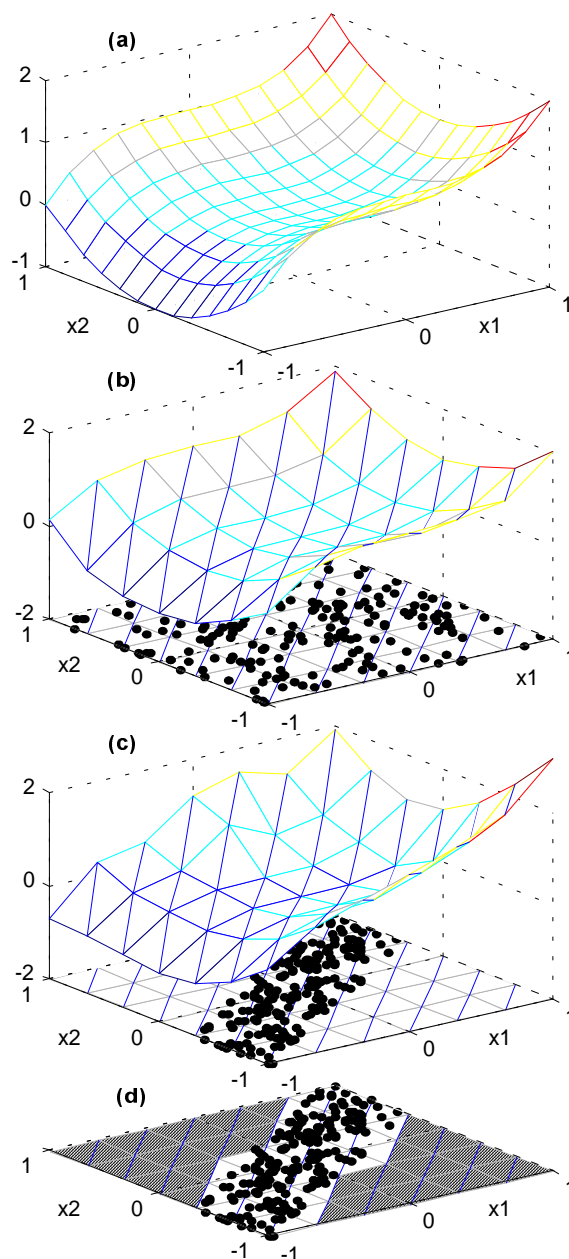


Figura 24: Función identificada $y = x_1^3 + x_2^2$.

Referencias

- [1] R. Babuska. *Fuzzy Modeling for Control*. Kluwer Academic Publisher, Massachusetts, USA, 1998.
- [2] K. Bossley. *Neurofuzzy Modelling Approaches in System Identification*. PhD thesis, University of Southampton, Highfield, Southampton, UK, 1997.
- [3] P. Eykhoff. *System Identification, Parameter and State Estimation*. John Wiley and Sons, Great Britain, UK, 1974. Reprinted 1979, ISBN 0-471-24980-7.
- [4] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, 1995.
- [5] S. Haykin. *Neural Networks, A Comprehensive Foundation*. Prentice-Hall Inc., New Jersey, USA, second edition, 1999. ISBN 0-13-273350-1.
- [6] H. Hellendoorn and D. Driankov. *Fuzzy Model Identification*. Springer-Verlag Cop., Berlin, Germany, 1997. ISBN 3-540-62721-9.
- [7] T. A. Johansen. Robust identification of Takagi-Sugeno-Kang fuzzy models using regularization. In *1997 IEEE International Conference on Robotics and Automation*, pages 180–186, New Orleans, USA, 1997.
- [8] P. Julian. *A High Level Canonical Piecewise Linear Representation, Theory and Applications*, by Pedro Julin,. PhD thesis, Departamento de Ingeniera Elctrica, Universidad Nacional del Sur, Bahia Blanca, Argentina, 1999.
- [9] L. Ljung. *System Identification, Theory for the user*. Printice Hall Inc., New Jersey, USA, 1987. ISBN 0-13-881640-9.
- [10] L. Ljung and T. Glad. *Modeling of Dynamic Systems*. Printice Hall Inc., New Jersey, USA, 1994. ISBN 0-13-597097-0.
- [11] T. Sderstrm and P. Stoica. *Systems Identification*. Prentice Hall, London, U.K., 1989. ISBN 0-13-127606-9.
- [12] J. Sjberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P-Y.Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.