

500879

DIVISION DE ESTUDIOS DE POSGRADO

Facultad de Ingeniería

IDENTIFICACION DE SISTEMAS VARIABLES EN EL TIEMPO

RAFAEL MAURICIO CANETTI WASSER

TESIS

Presentada a la División de Estudios de
Posgrado de la

FACULTAD DE INGENIERIA

de la

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

como requisito para obtener

el grado de

MAESTRO EN INGENIERIA

(CONTROL)

CIUDAD UNIVERSITARIA

OCTUBRE DE 1986



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

IDENTIFICACION DE SISTEMAS VARIABLES EN EL TIEMPO

Créditos asignados a la Tesis: 12 (Doce)

APROBADO POR EL JURADO:

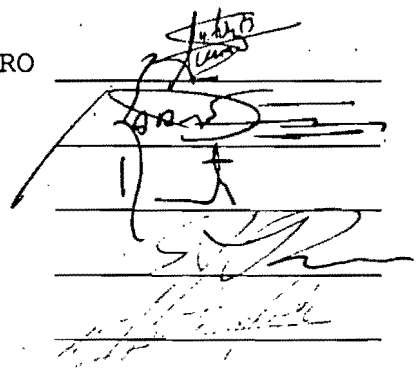
PRESIDENTE : DR. ANTONIO ALONSO CONCHEIRO

VOCAL : DR. MARTIN ESPAÑA

SECRETARIO : DR. ROMEO ORTEGA

SUPLENTE : DR. STANISLAW RACZYNSKI

SUPLENTE : DRA. CRISTINA VERDE RODARTE



Handwritten signatures and initials on a grid of lines. The signatures are written in black ink and are somewhat stylized. The grid consists of several horizontal lines, with the signatures written across them. The signatures appear to correspond to the names listed in the text to the left.

A MODO DE PROLOGO

El objetivo central de este trabajo era estudiar la posibilidad de incorporar esquemas de identificación a un predictor lineal multipasos para predecir el comportamiento de cuencas hidrológicas.

Este tema es de interés en la Sección de Automatización del Instituto de Ingeniería, donde tuvo lugar este trabajo.

Como se trata de un sistema no lineal, variable en el tiempo, etc, nos propusimos estudiar distintas alternativas existentes en la literatura para identificar sistemas variantes en el tiempo. Por eso, esta tesis dedica un capítulo a la descripción (sin estudiarlas con demasiado detalle) de las técnicas que nos parecieron más importantes para abordar este problema.

Por razones que se evidencian en el trabajo, elegimos una técnica para utilizarla en las aplicaciones que nos interesaban. Se trata del algoritmo de mínimos cuadrados con factor de olvido variable, al que dedicamos un capítulo aparte.

Dicho algoritmo fue estudiado con bastante detalle, y como ocurre frecuentemente, cuanto más avanzábamos en su estudio, más preguntas sobre su funcionamiento nos hacíamos, muchas de las cuales no han sido totalmente respondidas en la literatura.

Apoyándonos en trabajos previos de Rogelio Lozano Leal y Johnstone y Anderson (entre otros), pudimos obtener resultados de convergencia más fuertes que los ya existentes en la literatura, y aclaramos un poco más los orígenes de los problemas de "blow-up" y "bursting" tantas veces reportados como una limitación en el funcionamiento del algoritmo. Estos resultados se describen con todo detalle, así como simulaciones que verifican las previsiones teóricas. A pesar de que las simulaciones que se describen son el resumen de una enorme cantidad de ejemplos simulados en computadora, el estudio de este algoritmo ocupa un capítulo bastante abultado de la tesis.

Por los resultados obtenidos, este capítulo es la parte medular del trabajo desarrollado.

Finalmente se cumplió el objetivo deseado, aplicando esta técnica de identificación para la predicción de caudales en cuencas hidrológicas. A eso están dedicados los últimos dos capítulos.

Quiero agradecer el aporte de varias personas a este trabajo. Al Dr. Rogelio Lozano y al Dr. Romeo Ortega por sus comentarios que me ayudaron a aclarar dudas importantes y por la ayuda en tener "a mano" bibliografía nada fácil de obtener (y menos rápidamente). Al Dr. Martín D. España por la dirección de este trabajo de tesis, tarea en la que mostró una enorme paciencia, orientándome a buscar respuestas más precisas a los problemas planteados. A mis padres, cuyo apoyo permanente fue invaluable para la realización no solo de este trabajo sino de muchas otras cosas incluyendo toda mi carrera.

Finalmente quiero agradecer a Yolanda Rivera Sánchez el excelente trabajo de mecanografiado de la parte más difícil de la tesis.

I N D I C E

	Pág
1. INTRODUCCION	1
1.1 Identificación de sistemas dinámicos	1
1.2 Objetivo	6
2. ALGUNAS FORMAS DE ABORDAR EL PROBLEMA	8
2.1 Notación	10
2.2 Modelo de variación de parámetros	14
2.2.1 Filtro de Kalman	14
2.2.2 Doble recursión de aproximaciones estocásticas (STANKOVIĆ)	18
2.2.3 Suavizamiento óptimo	23
2.3 Observación de un horizonte finito (ventana rectangular)	25
2.3.1 Ventana rectangular, modelo paramétrico	26
2.3.2 Ventana rectangular, modelo no paramétrico	28
2.4 Estimador de dos niveles	38
3. MINIMOS CUADRADOS CON OLVIDO VARIABLE	42
3.1 El algoritmo	42
3.1.1 Introducción	42
3.1.2 Mínimos cuadrados pesados y con factor de olvido	44
3.2 Factor de olvido constante	52

	Pág
3.2.1 Velocidad de seguimiento. Ventana efectiva	53
3.2.2 Simulación	57
3.3 Dinámica del error de estimación	64
3.3.1 Superposición	66
3.3.2 Matriz de transición	68
3.3.3 Comportamiento de la matriz de covarianzas. Excitación persistente	75
a) Resultados generales	75
b) Comportamiento asintótico de $P_t(t \rightarrow \infty)$	78
c) Comportamiento de P_t si $\lambda \rightarrow 1$	79
d) Incidencia de λ_t en la adaptabilidad	80
e) Simulaciones	81
3.3.4 Evolución de la componente debida al error inicial: resultados generales	85
3.3.5 Evolución de la componente del error debido a variación en el parámetro	88
a) Variaciones esporádicas del parámetro	89
b) Variaciones arbitrarias del parámetro	103
3.3.6 Influencia de las perturbaciones	105
3.4 Selección del factor de olvido λ_t	108
3.4.1 Algoritmo de información constante	110
3.4.2 Algoritmo de traza constante	120
3.4.3 Detección basada en el error de predicción	122
3.4.4 Mecanismos robustos de detección	124
3.5 Algunas aplicaciones reportadas	133
3.6 Conclusiones	134

	Pág
4. APLICACION A PREDICCIÓN AUTOAJUSTABLE	137
4.1 Introducción	137
4.2 Predictor óptimo multipasos, parámetros conocidos	139
4.3 Predictor multipasos autosintonizable	143
4.4 Simulaciones	147
4.4.1 Simulaciones con modelos ARMA	149
4.4.2 Pruebas con modelos ARMAX	155
4.5 Adaptabilidad del predictor recursivo multipasos	162
4.5.1 Sistema sin perturbaciones	162
4.5.2 Sistema con perturbaciones	168
4.6 Conclusiones	170
5. APLICACION A LA PREDICCIÓN DE CAUDALES EN CUENCAS HIDROLOGICAS	173
5.1 Modelo simplificado de una cuenca	174
5.2 Pruebas con datos reales. Cuenca de Cantón	183
5.3 Cuenca de Cantón, modelo lineal puro	191
5.4 Pruebas con datos simulados	193
5.5 Conclusiones	203
APENDICE A - Matrices elementales de Householder	205
APENDICE B - Algunas desigualdades para autovalores	208

	Pág
APENDICE C - Demostraciones	211
REFERENCIAS	224

1. INTRODUCCION

1.1 Identificación de sistemas dinámicos

Una de las áreas de la Ingeniería de Control que ha recibido mucha atención en los últimos años ha sido la de sistemas autoadaptivos, orientada hacia esquemas de control o de predicción autoadaptiva de sistemas dinámicos.

Numerosas técnicas han sido desarrolladas (ver p. ej. R. Isermann [4]) en trabajos relativamente recientes, y han encontrado un rango bastante amplio de aplicaciones.

Todas estas técnicas se basan en un elemento común: la identificación en línea de sistemas dinámicos.

Desde un punto de vista formal se puede definir la identificación a partir de una familia \mathcal{M} de modelos y de la observación de un conjunto de señales de un sistema como: dadas las señales observadas, determinar el modelo $M \in \mathcal{M}$ que mejor ajuste las observaciones en el sentido de minimizar un criterio C .

Una clase particular de identificación es la que más impulso ha tenido. Se trata de la observación de las señales de entrada y salida de un sistema dinámico, lineal, de parámetros concentrados con perturbaciones estocásticas, en tiempo discreto. Estos sistemas siguen una ley de recurrencia del tipo:

$$y_t = a_1 y_{t-1} + \dots + a_n y_{t-n} + b_0 u_t + \dots + b_m u_{t-m} + e_t \quad (1.1)$$

donde u_t y y_t son entradas y salidas respectivamente, observadas en el tiempo t ; e_t es una v.a., y a_i , b_j son coeficientes constantes.

El problema es: estimar el valor de los parámetros $a_1 \dots a_n$, $b_0 \dots b_m$ a partir de la observación de las sucesiones u_t y y_t de entradas y salidas del sistema que se quiere identificar (ver R. Canetti [17] para una definición más precisa de esta cuestión).

Este problema se ha denominado de identificación paramétrica de sistemas dinámicos en tiempo discreto. Muchos algoritmos de solución han sido desarrollados por otros tantos autores.

Se pueden encontrar estos algoritmos por ejemplo en Goodwin/Payne [19], Goodwin/Sin [20], Isermann [4].

Las razones de la cantidad de contribuciones son muchas, pero hay dos de mayor importancia: la cantidad de sistemas físicos que pueden ser modelables con (1.1) y la extremada sencillez de dicho modelo. En particular, cualquier sistema describible por un sistema de ecuaciones diferenciales ordinarias, de coeficientes constantes, tiene, en tiempo discreto, una representación de entrada-salida como (1.1).

Muchos de los algoritmos desarrollados para resolver este problema poseen un atractivo adicional: son esquemas de cálculo recursivo que admiten una actualización a medida que nuevas observaciones son accesibles.

Esto permite el cálculo en línea, y la utilización de sus resultados también en línea.

Ejemplos de estos algoritmos son: mínimos cuadrados, máxima verosimilitud recursiva, variables instrumentales, mínimos cuadrados generalizados recursivos.

Por otro lado tienen una desventaja: presuponen que el sistema se rige según una ley como (1.1), aunque los sistemas reales son por lo general de parámetros distribuidos (orden no finito), variantes en el tiempo, y no lineales.

Los algoritmos de identificación arrojarán resultados tan buenos como bueno sea el modelo (1.1) para representar el sistema.

Es de sumo interés extender los algoritmos mencionados al caso en que los parámetros a_i , b_i sean variables en el tiempo. En efecto, esto permite abordar una clase mucho más amplia de sistemas.

Como ser:

a) Sistemas variables en el tiempo

Una gran cantidad de sistemas lineales son variables en el tiempo. Como ejemplo:

- Sistemas con envejecimiento. Sus características dependen de factores que evolucionan a lo largo del tiempo.
- Sistemas periódicos. Por ejemplo: sistemas que dependen de ciclos anuales, sistemas que dependen de características climáticas.

Hay sistemas que tienen ciclos diarios, p. ej.: sistemas biológicos. Un caso típico es el de plantas de tratamiento de aguas servidas.

b) Sistemas no lineales

Una gran clase de sistemas no lineales son linealizables

alrededor de su punto de operación actual.

Por ejemplo: . procesos que involucran fenómenos de transferencia de calor (transferencia radiativa, convección en condiciones no estacionarias, etc).

. Procesos con términos viscoelásticos, con términos de disipación no lineales.

Supongamos un sistema no lineal regido por

$$\begin{aligned} \dot{x}(t) &= f[x(t), u(t)] \\ y(t) &= g[x(t), u(t)] \end{aligned} \quad (1.2)$$

donde

$$\left\{ \begin{array}{l} u - \text{entrada.} \\ y - \text{salida.} \\ f, g - \text{funciones no lineales.} \end{array} \right.$$

Es sabido, que 1.2 se puede aproximar mediante:

$$\begin{aligned} \dot{\bar{x}}(t) &\cong A(t) \bar{x}(t) + B(t) \bar{u}(t) \\ \bar{y}(t) &\cong C(t) \bar{x}(t) + D(t) \bar{u}(t) \end{aligned} \quad (1.3)$$

donde

$$\left\{ \begin{array}{l} A, B, C, D \text{ son matrices dependientes de } x(t), \\ \quad \quad \quad u(t), f(\dots), g(\dots) \\ x(t) = \bar{x}(t) + x_0(t) \\ y(t) = \bar{y}(t) + y_0(t) \\ u(t) = \bar{u}(t) + u_0(t) \end{array} \right.$$

Siempre que se cumpla que $x_0(t), y_0(t), u_0(t)$ son solución de (1.2); $\bar{x}(t), \bar{u}(t), \bar{y}(t)$ son pequeñas variaciones alrededor de las funciones anteriores, y f, g satisfacen condiciones de regularidad (sobre sus derivadas parciales).

Obsérvese que se ha obtenido una aproximación lineal (1.3) de un sistema no lineal.

Obviamente (1.3) admite una representación como (1.1) con coeficientes a_i, b_i variantes en el tiempo.

Si el punto de operación se mantiene, se puede "identificar" el modelo lineal con los algoritmos mencionados. Es de suma utilidad poder extender esos algoritmos cuando el punto de operación sigue una ley de variación en el tiempo.

1.2 Objetivo

Como hemos visto la clase de sistemas físicos que se podrían identificar se ampliaría enormemente si se pudieran extender los algoritmos de identificación admitiendo variación temporal de los parámetros.

El objetivo del trabajo que se resume en este informe fue por un lado investigar en la literatura que soluciones se daban a este problema, y por otro lado aplicarlas a problemas de interés de la Coordinación: predicción autoadaptativa, y posible

aplicación a la predicción de caudales en cuencas hidrológicas.

En el capítulo 2 se discuten varias técnicas descritas en la literatura que hemos considerado de interés, así como la descripción de algunas aplicaciones sobre sistemas físicos reportados.

El capítulo 3 está dedicado a una de las técnicas de identificación : el método de mínimos cuadrados con factor de olvido variable.

En el capítulo 4 se desarrolla un predictor multipasos autoadaptivo y se discuten simulaciones realizadas para verificar sus propiedades y para compararlo con otro predictor multipasos autoadaptivo usado por varios autores.

En el capítulo 5 se discute una aplicación de la predicción autoadaptiva: la predicción de caudales en ríos a partir de la observación de las precipitaciones en sus cuencas. Se observa el comportamiento del predictor con datos reales obtenidos en la cuenca de Cantón.

2. ALGUNAS FORMAS DE ABORDAR EL PROBLEMA

Se han encontrado en la literatura distintos enfoques que intentan resolver el problema de la identificación de sistemas dinámicos variantes en el tiempo.

Las soluciones propuestas son "ad hoc", orientadas a la solución de un problema concreto y no existe ningún método que resuelva el problema en el caso general.

La solución exacta (o al menos muy aproximada) requeriría del conocimiento preciso de las leyes de variación del sistema o al menos limitar las formas de variación.

Los métodos descritos en la literatura presuponen: o bien que se conocen las leyes de variación del sistema, o que este varía lentamente y se ajusta un modelo a las últimas observacioio

nes, o bien que aún cuando haya cambios abruptos, hay largos períodos sin esta clase de cambios. Los métodos se pueden clasificar de distintas maneras, por ejemplo:

. De acuerdo a la clase de modelos del sistema:

- { paramétricos (ecuación en diferencias).
- { no paramétricos (respuesta a impulso).

. De acuerdo a cómo se supone que varía el sistema en el tiempo.

- { con modelo de variación de parámetros.
- { sin modelo de variación de parámetros.

. De acuerdo a cómo se toman en cuenta las observaciones.

- { observación de un horizonte finito (ventana rectangular, identificación por tramos).
- { observación de un horizonte no acotado con pesos relativos.

De todas las combinaciones posibles de estas clasificaciones veremos cuatro métodos (con algunos casos particulares), por ser los más importantes:

- i) Modelo de variación de parámetros.
- ii) Observación de un horizonte finito.
- iii) Observación de un horizonte no acotado, con pesos relativos.
- iv) Partición del espacio de parámetros.

En la próxima sección se unifica la notación y en las siguientes se describen los métodos.

2.1 Notación

Se considera un sistema S (ver fig. 2.1), descrito desde el punto de vista de entrada-salida. Dicho sistema es dinámico, lineal, de orden finito n , y satisface la recursión (2.1):

$$y_t + a_{1t} y_{t-1} + \dots + a_{nt} y_{t-n} = b_{0t} u_t + b_{1t} u_{t-1} + \dots + b_{mt} u_{t-m} + e_t \quad (2.1)$$

donde:

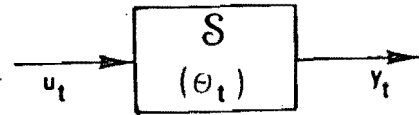
$$(2.2) \quad \left\{ \begin{array}{l} \{y_t\}_0^\infty \text{ es la sucesión de salidas (observables) del sistema.} \\ \{u_t\}_0^\infty \text{ es la sucesión de entradas (observables) del sistema.} \\ \{a_{it}\}_0^\infty \quad i=1..n, \{b_{jt}\}_0^\infty \quad j=0..m \text{ son las sucesiones de parámetros variables en el tiempo.} \\ \{e_t\}_0^\infty \text{ es una sucesión de v.a. centradas, independientes y de varianza acotada.} \end{array} \right.$$

llamando:

$$(2.3) \quad \left\{ \begin{array}{l} x_t^T = [y_{t-1}, y_{t-2}, \dots, y_{t-n}, u_t, u_{t-1}, \dots, u_{t-m}], x_t \in \mathbb{R}^p \text{ con } p=m+n+1 \\ \theta_t^T = [-a_{1t}, \dots, -a_{nt}, b_{0t}, b_{1t}, \dots, b_{mt}], \theta_t \in \mathbb{R}^p \end{array} \right.$$

Entonces $y_t = x_t^T \theta_t + e_t \quad (2.4)$

Esta es la clase de sistemas que consideraremos. Como se vé, estos quedan caracterizados por su vector de parámetros θ_t , variable en el tiempo.



(Fig 2.1)

De lo que se trata es de "identificar" el sistema, estimando el valor del parámetro θ_t , con un estimador $\hat{\theta}_t$, también variable en el tiempo.

La forma de calcular el estimador es la que diferenciará un método de otro. Lo común a todos los métodos es que las entradas y salidas $u_i, y_i, 1 \leq i \leq t$ son los datos a partir de los cuales se infiere $\hat{\theta}_t$.

Si no se restringiera más el problema, se notaría la primera dificultad: así planteado, el problema de identificación es irresoluble puesto que existe un número infinito de soluciones.

En efecto: aún supuestas conocidas las entradas, salidas, e incluso las perturbaciones, existen infinitas maneras de elegir la sucesión de parámetros θ_t de manera de satisfacer exactamente la ecuación (2.4).

¿Cómo distinguir la verdadera sucesión θ_t ?

Más aún: aunque θ_t fuera obtenida, digamos "casualmente", su

conocimiento no serviría a efectos prácticos (por ejemplo para propósitos de control o predicción) si θ_{t+1} fuera independiente de θ_t .

El problema radica en permitir una variación arbitraria de θ_t .

Como problema teórico, es un problema sin solución única. Pero como problema práctico carece de sentido: no trabajaremos con sistemas cuyas características sean completamente independientes en dos instantes consecutivos de muestreo.

Existen restricciones adicionales para la evolución de θ_t :

1. El parámetro θ_t debe estar contenido en un subconjunto de parámetros admisibles por restricciones físicas.
2. Los parámetros θ_t y θ_{t+1} no son independientes sino que la evolución de θ_t depende de cómo varía el sistema real a lo largo del tiempo. Esta variación está también limitada por razones físicas.

Nótese que los primeros métodos de identificación presuponían la restricción 2) más fuerte: $\theta_t = \theta_{t+1} \forall t$.

El primer algoritmo que veremos supone una ley precisa de la evolución del parámetro. Luego veremos dos métodos que pueden considerarse como una extensión de las ideas fundamentales

del primero.

2.2 Modelo de variación de parámetros

2.2.1 Filtro de Kalman

Supongamos el sistema lineal regido por (2.4):

$$y_t = x_t^T \theta_t + e_t$$

y un modelo estocástico de variación del parámetro θ_t :

$$\theta_{t+1} = F_t \theta_t + v_t \quad (2.5)$$

donde: $\left\{ \begin{array}{l} F_t \in R^{P \times P} \quad \forall t \\ v_t \in R^P \text{ es una va. con } E\{v_t\} = 0 \\ \text{y además } \text{cov}\{v_t, e_{t+\tau}\} = 0 \quad \forall \tau \quad \text{cov}\{v_i, v_k\} = R_i \delta_{ik} \end{array} \right.$

El filtro de Kalman nos dá una expresión recursiva para estimar el parámetro θ_{t+1} con las observaciones (y_t, x_t) hasta el instante $t+1$. En ese caso, la expresión del filtro en su forma de estimador (y no de predicador!) sería:

$$\begin{cases}
 \hat{\theta}_{t+1} = F_t \hat{\theta}_t + K_{t+1} [y_{t+1} - x_{t+1}^T F_t \hat{\theta}_t] \\
 K_{t+1} = P_t x_{t+1} / (\sigma^2 + x_{t+1}^T P_t x_{t+1}) \\
 P_{t+1} = F_{t+1} P_t F_{t+1}^T + R_{t+1} - F_{t+1} P_t x_{t+1} x_{t+1}^T P_t F_{t+1}^T / \\
 (\sigma^2 + x_{t+1}^T P_t x_{t+1})
 \end{cases} \quad (2.6)$$

Suponiendo $\begin{cases} F_0 = I \\ R_0 = \underline{0} \end{cases}$

con las condiciones iniciales $\begin{cases} \hat{\theta}_0 = E\{\theta_0\} \\ P_0 = \text{cov}\{\hat{\theta}_0\} \end{cases}$

Notación - Designaremos $\hat{\theta}_{t/t}$ al estimador de θ_t con las observaciones hasta el tiempo t , y $\hat{\theta}_{t+1/t}$ al estimador de θ_{t+1} con las observaciones hasta el instante t .

Por comodidad de notación escribimos $\hat{\theta}_{t+1}$ en lugar de $\hat{\theta}_{t+1/t+1}$

La derivación de las ecuaciones (2.6) puede encontrarse en Sage y Melsa [21], Anderson y Moore [22].

La matriz P_{t+1} está relacionada con la matriz de covarianzas de la estimación a través de:

$$P_t = F_t \Sigma_t F_t^T + R_t \quad (2.7)$$

$$\text{donde } \Sigma_t = \text{Var} \{ \hat{\theta}_t \} \quad (2.8)$$

Optimalidad

- 1) De los posibles estimadores de θ_{t+1} dadas las observaciones hasta el tiempo $t+1$, el filtro de Kalman provee la mejor estimación ($\hat{\theta}_{t+1}$) lineal en las observaciones $x_i^T, y_i, i=1 \dots t+1$, en el sentido de variancia mínima.
- 2) Si además e_t, v_t, θ_0 son variables aleatorias gaussianas independientes, entonces el filtro de Kalman no es sólo el mejor estimador lineal sino el mejor estimador (en sentido de variancia mínima) de θ_{t+1} .

Observación - Se aclaró la utilización de la forma estimador del filtro de Kalman porque parece que hay una utilización un poco confusa de este algoritmo en parte de la literatura revisada.

A efectos de la identificación conviene tener una expresión recursiva como la (2.6) que permita la actualización de $\hat{\theta}_{t+1}$ una vez que llegan las nuevas observaciones y_{t+1}, x_{t+1} .

A veces se usa una expresión en forma de predictor $\hat{\theta}_{t+1/t}$, útil en otros casos, que difiere de (2.6) solamente en un término en la primer ecuación, pero que corresponde a un caso esencialmente distinto.

Esquema de cálculo

Como se desprende de las ecuaciones (2.6), estas permiten un cálculo recursivo que puede dar lugar a aplicaciones en línea (en tiempo real).

Esto está facilitado por el hecho de ser las matrices P_t y R_t simétricas. Además por lo general se "ayuda" al cálculo con matrices de transición F_t particulares.

En la figura (2.2) se representa el esquema de cálculo en tiempo real del estimador $\hat{\theta}_{t+1}$, en relación al funcionamiento del sistema S que genera los datos.

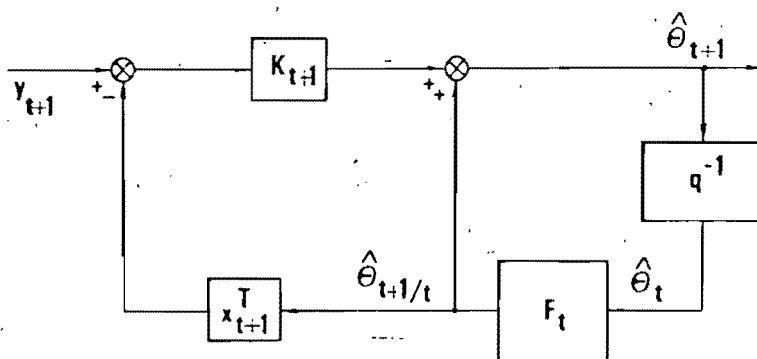


Fig. 2.2 Esquema de Cálculo

Aplicaciones

De las ecuaciones (2.4) y (2.5) se desprende la clase de sistemas variantes en el tiempo que pueden ser identificados por este método.

Usado preferentemente para sistemas que varían lentamente en el tiempo, se supone por lo general un modelo de variación de parámetros de la forma:

$$\theta_{t+1} = \theta_t + v_t \quad (2.9)$$

Es decir que $F_t = I \forall t$, y la matriz de covarianzas de v_t (por lo general de forma diagonal) expresa la velocidad con la que se espera que cambie el vector de parámetros. Es el esquema clásico de "marcha aleatoria simple".

De entre las aplicaciones reportadas podemos citar a Wittenmark [11], que utiliza este estimador como un estimador "fino" dentro de otro esquema general que se describirá más adelante.

Limitaciones - Todos los autores coinciden en algunas limitaciones que han encontrado en la aplicación de este algoritmo, ver p. ej.: [6], [9] y [11].

La primera limitación es que funciona bastante mal cuando el sistema no varía lentamente en el tiempo.

La segunda limitación es que es necesario conocer la dinámica del parámetro, esto es: F_t, R_t .

En simulaciones y aplicaciones se ha encontrado que si el sistema varía lentamente en el tiempo, el algoritmo funciona bastante bien.

También se ha encontrado que errores cometidos en la estimación de σ^2 y R_t influyen bastante en el error de estimación, pudiendo llegar a ser muy grande especialmente si σ^2 y R_t varían en el tiempo, o cuando se desprecia la influencia de los valores iniciales del algoritmo.

Veremos enseguida dos extensiones que tratan de aprovechar las propiedades del filtro de Kalman y reducir sus inconvenientes.

2.2.2 Doble recursión de aproximaciones estocásticas (STANKOVIĆ)

Seguiremos el razonamiento de Stanković, ver ref. [9]. Parte de las ecuaciones del filtro de Kalman (2.6), en forma de predictor, esto es que la primer ecuación es:

$$\hat{\theta}_{t+1/t} = F_t \hat{\theta}_{t/t-1} + F_t K_{t+1} [y_t - x_t^T \hat{\theta}_{t/t-1}] \quad (2.10)$$

Notando que por el modelo de variación del parámetro (2.5) se cumple que:

$$\hat{\theta}_{t+1} = \hat{\theta}_{t+1/\theta} = F_t \hat{\theta}_{t/t} \quad (2.11)$$

$$\text{y que } \hat{\theta}_t \triangleq \hat{\theta}_{t/t} \quad (2.12)$$

llegamos a la misma ecuación (2.6)

$$\hat{\theta}_{t+1} = F_t \hat{\theta}_t + K_{t+1} [y_{t+1} - x_{t+1}^T F_t \hat{\theta}_t] \quad (2.13)$$

o lo que es lo mismo:

$$\hat{\theta}_{t+1} = F_t \hat{\theta}_t + \Gamma_{t+1} x_{t+1} [y_{t+1} - x_{t+1}^T F_t \hat{\theta}_t] \quad (2.14)$$

$$\text{con } \Gamma_{t+1} = P_t / (\sigma_t^2 + x_{t+1}^T P_t x_{t+1}). \quad (2.15)$$

El algoritmo descrito por (2.14) es un algoritmo general de aproximaciones estocásticas, del que el filtro de Kalman es un caso particular de selección de la matriz de ganancias Γ_{t+1} , con las ecuaciones (2.15) y (2.6).

Como se dijo, el filtro de Kalman es sensible al conocimiento de la caracterización estocástica del modelo. Lo que plantea Stanković es producir entonces una matriz de ganancias Γ_{t+1} adaptable respecto a esa caracterización.

Al algoritmo (2.14)-(2.15) se llega considerando una función de costo $J(\hat{\theta}_t, t) = E \{ (y_t - x_t^T \hat{\theta}_t)^2 | \hat{\theta}_t \}$ (2.16)

minimizándola a partir de realizaciones de

$$\nabla_{\theta} J(\hat{\theta}_i, i)$$

y utilizando (para extrapolar), el modelo (2.5). Supongamos desconocida r_{t+1} y consideremos la función

$$J(r_{t+1}, t+1, t) = J(\hat{\theta}_{t+1}(r_{t+1}), t+1) \quad (2.17)$$

Es decir reemplazando $\hat{\theta}_{t+1}$ por la expresión (2.14)

El objetivo ahora es minimizar $J(r_{t+1}, t+1, t)$ respecto la matriz desconocida r_{t+1} , para cada $t+1$.

Siguiendo ese razonamiento, Stankovic concluye la expresión general para $\hat{r}_{t+1} = \text{Arg} \{ \min_{r_{t+1}} J(r_{t+1}, t+1, t) \}$, utilizando el al

goritmo de aproximaciones estocásticas:

$$\hat{r}_{t+1} = \hat{r}_t + W_t F_t^T x_{t+1} x_t^T \epsilon_{Ft} [\epsilon_{pt+1} - x_{t+1}^T F_t \hat{r}_t x_t \epsilon_{Ft}] \quad (2.18)$$

donde $\epsilon_{Ft} = y_t - x_t^T \hat{\theta}_t$ (error de filtrado) (2.19)

$$\epsilon_{pt+1} = y_{t+1} - x_{t+1}^T F_t \hat{\theta}_t \quad (\text{error de predicción}) \quad (2.20)$$

Y la W_t es una sucesión "adecuada" de matrices de peso para la recursión de r_{t+1} .

Finalmente el algoritmo queda descrito por las ecuaciones (2.14) y (2.18). Como se vé, son dos procedimientos acoplados de aproximaciones estocásticas, uno para la evolución del parámetro y otro para la evolución de la matriz de ganancias del procedimiento anterior. Esto permite la "adaptación" de la matriz de ganancias a los cambios en la caracterización estocástica de la evolución de θ_t .

Solución subóptima

Si en lugar de minimizar $J(\Gamma_{t+1}, t+1, t)$ sobre $\Gamma_{t+1} \in R^{p \times p}$ se minimiza sobre las Γ_{t+1} de forma diagonal, se obtiene.

$$\hat{\gamma}_{t+1} = \hat{\gamma}_t + W_t F_t^T x_{t+1} \text{diag } x_t^T \epsilon_{Ft} [\epsilon_{Pt+1} - F_t^T x_{t+1} \text{diag } x_t^T \hat{\gamma}_t \epsilon_{Ft}] \quad (2.21)$$

donde una elección "conveniente" de la sucesión W_t es:

$$W_t = W_t^1 / (\beta_t^T W_t^1 \beta_t + \lambda) \quad (2.22)$$

$$\text{con } \left\{ \begin{array}{l} \beta_t = F_t^T x_{t+1} \text{diag } x_t^T \epsilon_{Ft} \end{array} \right. \quad (2.23)$$

$$\left\{ \begin{array}{l} W_{t+1}^1 = \frac{1}{\lambda} [W_t^1 + \Delta W - W_t^1 \beta_t (\beta_t^T W_t^1 \beta_t + \lambda)^{-1} \beta_t^T W_t^1] \end{array} \right. \quad (2.24)$$

$$\text{las constantes: } \left\{ \begin{array}{l} \lambda : \quad 0 < \lambda \leq 1 \\ \Delta W \quad (\Delta W \geq 0) \end{array} \right.$$

2.2.3 Suavizamiento óptimo

Este procedimiento sirve para la identificación fuera de línea, o al menos cuando la estimación del parámetro θ_t se necesita con retraso, por ejemplo en $t + d$.

Supongamos conocer las entradas y salidas del sistema desde $t=1$ hasta $t=N$, es obvio que las propiedades del estimador $\hat{\theta}_{t/N}$ son mejores que las del estimador $\hat{\theta}_{t/t}$ porque la información disponible en este caso está contenida en el primero.

Reescribamos las ecuaciones del filtro de Kalman por comodidad de notación

$$\left\{ \begin{array}{l} \hat{\theta}_{t+1/t+1} = F_t \hat{\theta}_{t/t} + \frac{S_{t+1/t} x_{t+1}}{1 + x_{t+1}^T S_{t+1/t} x_{t+1}} [y_{t+1} - x_{t+1}^T F_t \hat{\theta}_{t/t}] \\ S_{t+1/t+1} = S_{t+1/t} - \frac{x_{t+1} S_{t+1/t} S_{t+1/t} x_{t+1}^T}{1 + x_{t+1}^T S_{t+1/t} x_{t+1}} \\ S_{t+1/t} = F_t S_{t/t} F_t^T + r_t w_t r_t^T \end{array} \right. \quad (2.25)$$

donde hemos llamado

$$\left\{ \begin{array}{l} S_{t+1/t+1} = (P_{t+1})/\sigma_t^2 \\ \bar{w}_t = (w_t)/\sigma_t^2 \end{array} \right. \quad (2.26)$$

Las formas más simples de computar $\hat{\theta}_{t/N}$ son (ver ref. 6):

$$(I) \begin{cases} \hat{\theta}_{t/N} = F_t^{-1} (\hat{\theta}_{t+1/N} - r_t^T Q_t r_t^T \mu_t) \\ \mu_t = F_{t+1}^T \mu_{t+1} - x_{t+1} [y_{t+1} - x_{t+1}^T \hat{\theta}_{t+1/N}] \end{cases} \quad (2.27)$$

$$(II) \begin{cases} \hat{\theta}_{t/N} = \hat{\theta}_{t/t} - S_{t/t} F_t^T \mu_t \\ \mu_t = \left(I - \frac{x_{t+1} x_{t+1}^T S_{t+1/t}}{1 + x_{t+1}^T S_{t+1/t} x_{t+1}} \right) (F_{t+1}^T \mu_{t+1} - x_{t+1} [y_{t+1} - x_{t+1}^T (F_t \hat{\theta}_{t/t})]) \end{cases} \quad (2.28)$$

Obviamente el algoritmo I es más conveniente en lo que se refiere al uso de memoria (no hay que almacenar las estimaciones $\hat{\theta}_{t/t}$, $S_{t/t}$).

El procedimiento es sencillo: primero se computa $\hat{\theta}_{t/t}$ de $t = 1$ a N y luego se computa $\hat{\theta}_{t/N}$ desde $t=N$ hasta $t=1$ con valor inicial $\mu_N = 0$.

El problema con el algoritmo I consiste en que es una recursión inestable (puede demostrarse) y si se utiliza en una longitud de pasos grande, el error acumulado es bastante grande también.

El procedimiento más conveniente sería recurrir a I por tramos, y actualizar periódicamente con el algoritmo II. Para ello hay que almacenar previamente $\hat{\theta}_{i/i}$, $S_{i/i}$, eligiendo algunos valores

de i , $1 \leq i \leq N$.

Aplicaciones

En la ref. [6] se encuentran resultados de la aplicación exitosa del procedimiento sobre datos simulados de modelos ARMAX con variación sinusoidal de algún parámetro.

Se reporta también la aplicación del algoritmo sobre datos reales, de la identificación de un modelo precipitación-escorrentía del río Huon, en Tasmania.

2.3 Observación de un horizonte finito (ventana rectangular)

Con este método lo que se propone es utilizar la observación de entradas y salidas del sistema en un intervalo de N instantes de muestreo, para obtener una estimación del modelo del sistema a identificar.

La suposición básica es que el sistema no varía mucho entre t y $t+N$.

En el caso de cambios abruptos se considera que la identificación tiene un retardo de N instantes de muestreo en "seguir" el cambio.

Hay muchas maneras de aplicar esta idea de la ventana rectan-

gular, tantas como algoritmos de identificación en los que sea posible poner en forma explícita que se utilizarán datos desde $t-N$ hasta t para obtener $\hat{\theta}_t$.

Se distinguen sin embargo dos variantes:

- i) para obtener $\hat{\theta}_t$ se descartan los datos en $t-N-1$ y se incorporan los datos en t . Es una forma de desplazar continuamente la "ventana rectangular".
- ii) Entre KN y $(K+1)N$ se mantiene $\hat{\theta}_{KN}$ mientras se estima $\hat{\theta}_t$ con $KN < t < (K+1)N$ cuando $t = (K+1)N$ se adopta $\hat{\theta}_{(K+1)N}$ como nuevo estimador, y se descarta al anterior.

Veremos dos ejemplos de utilización de la ventana rectangular, uno con modelo paramétrico y otro con modelo no paramétrico.

2.3.1 Ventana rectangular, modelo paramétrico

Veremos el algoritmo de mínimos cuadrados con ventana rectangular.

Este algoritmo se obtiene de minimizar la función de costo:

$$J(\bar{\theta}_{i+N, i+1}) = \sum_{k=i+1}^{i+N} [y_k - x_k^T \bar{\theta}_{i+N, i+1}]^2 \quad (2.29)$$

Donde $\bar{\theta}_{i+N, i+1}$ es un estimador genérico de θ_{i+N} con datos desde $t=i+1$ hasta $t=i+N$.

Se obtienen expresiones recursivas para el cálculo de

$$\hat{\theta}_{i+N, i+1} = \text{Arg}\{\min_{\bar{\theta}_{i+N, i+1} \in R^p} J(\bar{\theta}_{i+N, i+1})\} \quad (2.30)$$

Estas recursiones comprenden dos partes; estando en $t=i+N-1$, primero: se incorporan los datos en $t=i+N$ (y_{i+N} , x_{i+N}), segundo: se descartan los datos en $t=i$ (y_i , x_i).

1) incorporación de datos en $t=i+N$:

$$\left\{ \begin{aligned} \hat{\theta}_{i+N, i} &= \hat{\theta}_{i+N-1, i} + K_{i+N, i} [y_{i+N} - x_{i+N}^T \hat{\theta}_{i+N-1, i}] \\ K_{i+N, i} &= P_{i+N-1, i} \left[\frac{x_{i+N}}{1 + x_{i+N}^T P_{i+N-1, i} x_{i+N}} \right] \\ P_{i+N, i} &= \left[I - P_{i+N-1, i} \frac{x_{i+N} x_{i+N}^T}{(1 + x_{i+N}^T P_{i+N-1, i} x_{i+N})} \right] P_{i+N-1, i} \end{aligned} \right. \quad (2.31)$$

2) descarte de los datos en $t=i$:

$$\left\{ \begin{aligned} \hat{\theta}_{i+N, i+1} &= \hat{\theta}_{i+N, i} - K_{i+N, i+1} [y_i - x_i^T \hat{\theta}_{i+N, i}] \\ K_{i+N, i+1} &= P_{i+N, i} \left[\frac{x_i}{1 - x_i^T P_{i+N, i} x_i} \right] \\ P_{i+N, i+1} &= \left[I + P_{i+N, i} \frac{x_i x_i^T}{1 - x_i^T P_{i+N, i} x_i} \right] P_{i+N, i} \end{aligned} \right. \quad (2.32)$$

De (2.31) y (2.32) se desprende la sencillez del cálculo recursivo del estimador $\hat{\theta}_{i+N, i+1}$. Si se va a utilizar un horizonte N muy largo, hay que tomar en cuenta que se necesita almacenar en la memoria los datos $y_k, x_k \quad k=i, \dots, i+N$.

En este caso se recomienda una variante del algoritmo que es el de "memoria oscilante" (ver ref. [19]), que consiste en incorporar los nuevos datos de uno en uno, y descartarlos por bloques de largo N , cada N instantes de muestreo.

El nombre de "memoria oscilante" proviene del hecho que dicho algoritmo considera un horizonte de datos cuya longitud va oscilando entre N y $2N$ instantes de muestreo.

La versión de "ventana rectangular" de mínimos cuadrados ha sido ampliamente usada pero no parece presentar ventajas frente al algoritmo de mínimos cuadrados con olvido exponencial que veremos más adelante. Incluso este último posee ventajas adicionales: poder modular el horizonte de datos que se toman en cuenta, y no necesita almacenar las observaciones pasadas.

2.3.2 Ventana rectangular. Modelo no paramétrico

Como en el algoritmo anterior, la idea básica es utilizar las observaciones entre $t+1$ y $t+N$ para identificar el sistema. En este caso el sistema está modelado por la respuesta al pulso, es decir por la sucesión $h(k)$, $k=0, 1, 2, \dots$

$$y_t = \sum_{k=0}^{\infty} h_k u_{t-k} \quad (2.33)$$

Podría pensarse que no tiene sentido identificar el sistema por su respuesta al pulso, sin embargo hay que tener en cuenta dos cosas: i) que hay situaciones donde la respuesta al pulso es una caracterización apropiada para propósitos de control, y ii) que con dispositivos microelectrónicos cada vez más rápidos y con mayor capacidad de almacenamiento resulta un problema cada vez menor el almacenar una cantidad grande de valores de la sucesión de peso y operar con ella.

Veremos esto con dos ejemplos de aplicaciones reportadas:

CONTROL DE UN PROCESO DE FUSION POR ZONA

Esta aplicación está reportada por C.Wiss (referencia [10]). Se trata de un proceso de fusión de una muestra metálica por inducción.

La muestra metálica está rodeada por una bobina de inducción que avanza a una velocidad v constante. Esto permite considerar la evolución del sistema en función de una variable espacial x (posición del centro de la bobina), en lugar de la variable temporal t .

El objetivo final es mantener la zona de fundido alrededor de x con una longitud constante $= 2 \delta$.

La variable que se puede observar $T_E(x)$ es la temperatura en el extremo de la muestra, mientras la bobina está en la posición x . La variable que se puede controlar $U(x)$, es una señal que regula la potencia en la bobina de inducción.

Experimentalmente se sabe que el objetivo de mantener en 2δ el espesor fundido, se cumple si la curva $T_E(x)$ tiene un valor $T_D(x)$ representado por C en la figura 2.5.

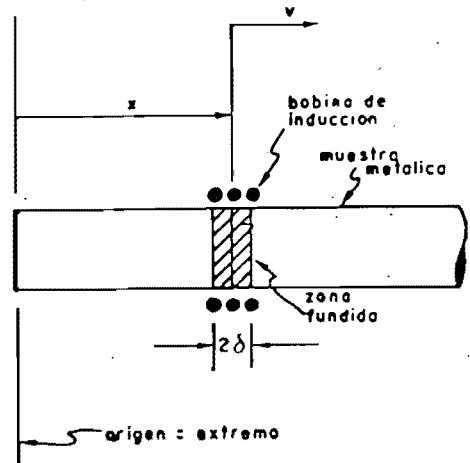


Fig. 2.4

En la misma figura se designa como \underline{b} el valor de $T_E(x)$ si se mantuviera $U(x) = \text{cte.}$

Obviamente se trata de un sistema variable en el tiempo (ahora parametrizado en x en lugar de t).

Se divide x en intervalos $\theta, 2\theta, 3\theta, \dots, k\theta, \dots$ (øcte.) y se mantiene $U(x)$ constante en esos intervalos:

$$U(x) = U(k) \quad \text{si } k\theta \leq x < (k+1)\theta$$

Se considera un modelo lineal:

$$T_E(k\theta + \xi) = T_p(k\theta + \xi) + q(k\theta + \xi)\Delta U_k \quad 0 \leq \xi < \theta \quad (2.34)$$

donde: $\begin{cases} T_p(k\theta + \xi) = \text{temp. de salida si } \Delta U_k = 0 \text{ (potencia constante)} \\ q(k\theta + \xi) = \text{relaci3n incremento de temperatura de salida} \\ \text{a incremento de la variable de control } U_k. \end{cases}$

La curva T_p se calcula experimentalmente (es la de temperaturas a potencia constante) como sigue:

En el intervalo $[(k-1)\theta, k\theta)$ se mide $T_E(x)$, entonces $T_p[k\theta + \xi]$ es $T_E(x)$ prolongada al intervalo siguiente con $\Delta U_k = 0$. Para estimar $T_E(x)$ se ajusta una curva:

$$T_p(x) = \alpha_1 x_i + \alpha_2 (1 - e^{-\alpha_3 x_i}) \quad (2.35)$$

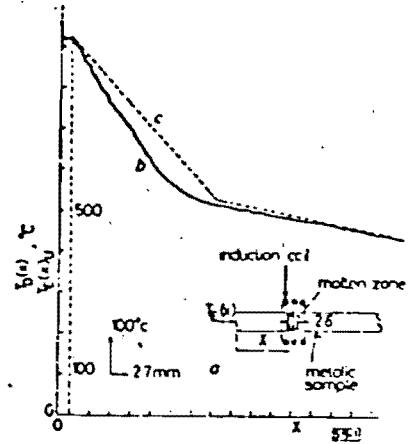
donde $k\theta \leq x_i < (k+1)\theta$

$\alpha_1, \alpha_2, \alpha_3$ se calculan en $(k-1)\theta \leq x < k\theta$.

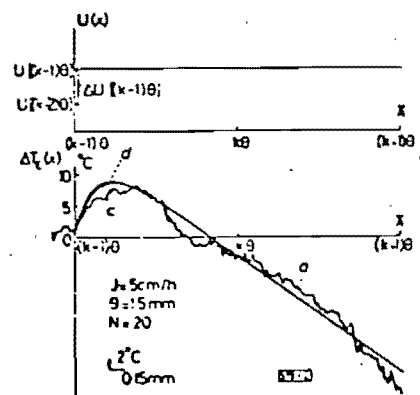
finalmente se estima:

$$\hat{q}(k\theta + \xi) = \frac{T_E(k\theta + \xi) - T_p(k\theta + \xi)}{\Delta U_k} \quad (2.36)$$

con la estimaci3n $\hat{q}(k\theta + \xi)$ y la ley (2.34) se calcula en el intervalo siguiente



(Fig 2.5)



(Fig 2.6)

$(k+1)\theta < x < (k+2)\theta$ una ley de control que optimiza una función de costo cuadrática que penaliza el alejamiento de T_E del valor T_D deseado y la señal de control ΔU_k .

En la figura 2.7 se muestra un resultado experimental, donde ϵ_1 representa el error $(T_D - T_E)$ cuando $U(x) = \text{cte}$, y ϵ_2 representa el mismo error para el sistema controlado.

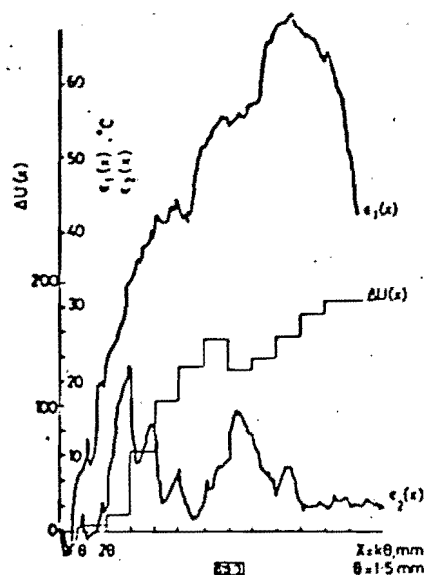


Fig. 3 Performance of the method
 $\epsilon_1(x)$ = error at constant $U(x)$
 $\epsilon_2(x)$ = error for the controlled system
 $\Delta U(x)$ = corresponding to $\epsilon_1(x)$, the computed and applied control input variations $\Delta U(x)$

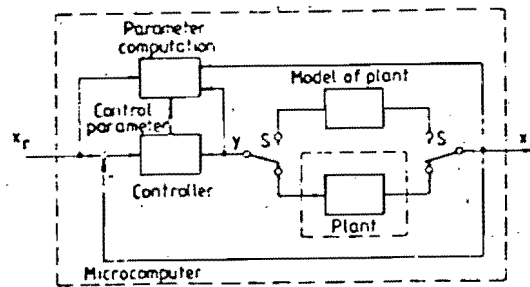
(Fig 2.7)

CONTROLADOR P.I.D. AUTOADAPTIVO

Esta aplicación reportada por G.Fromme [3] muestra la potencialidad del uso de microprocesadores con técnicas clásicas de control.

Se trata de un controlador tipo P.I.D. implementado en una microcomputadora, con la propiedad de autoajustarse a las características de la planta a controlar. Esta aplicación está orientada a desarrollar un controlador de carácter general, instalable en una amplia clase de plantas.

La figura 2.8 muestra el diagrama esquemático del controlador.



Self-adjusting controller with search procedure.

(Fig 2.8)

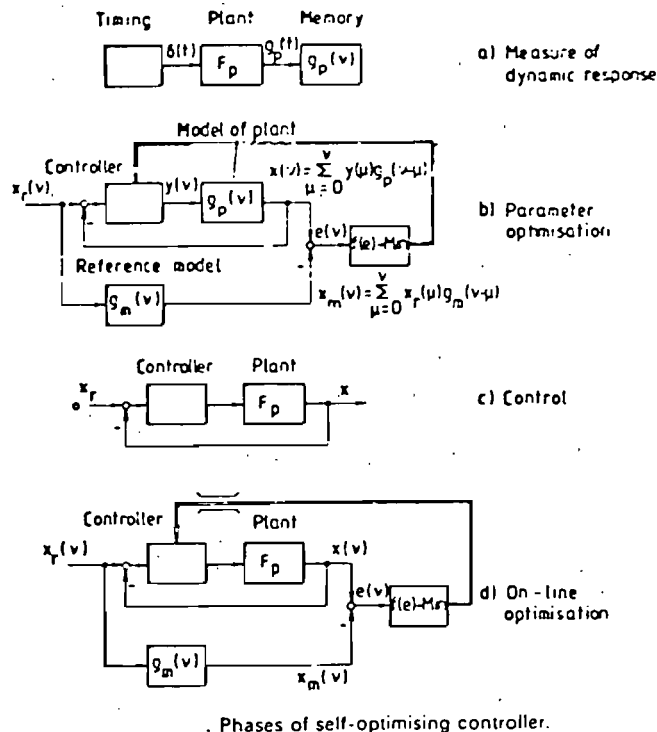
El funcionamiento tiene 3 etapas (fig. 2.9). En primera instancia se excita el sistema a controlar y se registra la sucesión de peso del sistema: $g_p(t)$ $t=1,2,3\dots$. Esta se almacena como modelo de la planta a controlar.

La segunda fase corresponde al ajuste de los parámetros del controlador, fuera de línea. Para ello el controlador desconecta la planta y con el modelo almacenado realiza una búsqueda multidimensional minimizando el error:

$$OF = \frac{1}{N} \sum_{k=0}^{N-1} |e(k)|$$

Donde $e(k)$ es la diferencia entre las respuestas a escala del lazo de control simulado y una respuesta objetiva (modelo de referencia).

Una vez concluida la búsqueda de los parámetros del controlador, la tercera fase es el control en tiempo real de la planta.



(Fig 2.9)

Hay una cuarta fase que es posible activar a solicitud del operador: la realización de la búsqueda de los parámetros del controlador operando en línea. En tal caso opera la planta y no el modelo almacenado.

Limitaciones

Como es obvio de la descripción del controlador hay algunas suposiciones implícitas que deben ser satisfechas:

- La planta debe ser lentamente variable en el tiempo, ó en forma intermitente.

- La planta debe ser estable en lazo abierto.
- No hay no-linealidades importantes.
- Las perturbaciones durante las mediciones son pequeñas.
- No existe gran sensibilidad respecto de los parámetros del controlador.

El controlador fue implementado en una pequeña microcomputadora con un microprocesador INTEL 8086 con 8k de programa y 4k de RAM.

Los convertidores tienen una resolución de 12 bits y se agregó un multiplicador rápido (TRWTDC 1010).

Las dimensiones del controlador son aprox: 8 x 10 x 22 cm.

Fue utilizado en el control de una prensa hidráulica para ensayo de materiales, para trabajo tanto a la tracción como a la compresión.

El modelo de la planta depende de la prueba en ensayo.

Un diagrama de bloques y un diagrama esquemático se muestran en la figura 2.10.

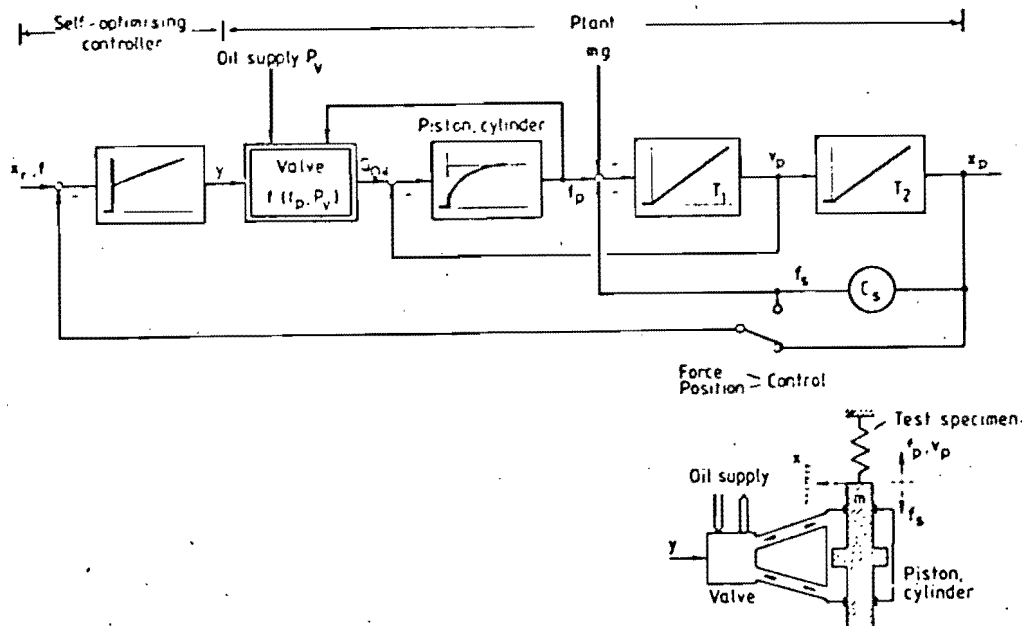
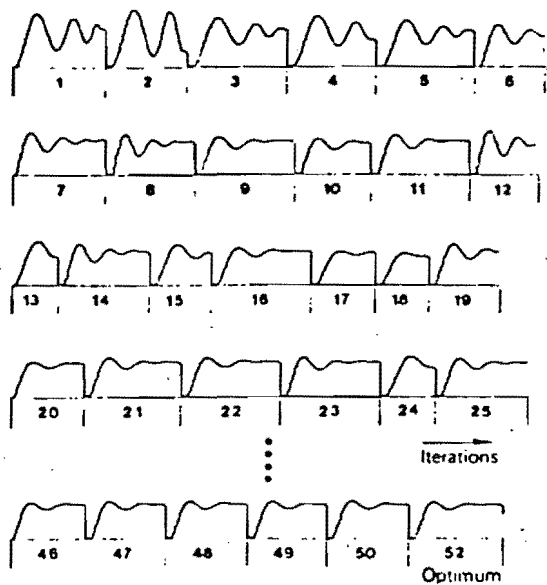


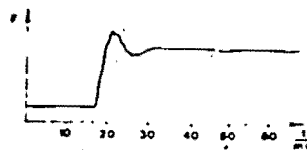
Fig. 2.10

El controlador aplica un pequeño pulso para la determinación de la respuesta del sistema. Tarda unos 6 segundos en el cálculo de los parámetros, y con el dato de la carga máxima a aplicar al espécimen calcula el segundo y último pulso.

En la figura (2.11) se muestra una respuesta a escalón óptima obtenida en un ensayo. El No. de iteraciones necesario para obtenerla fue de 52.



a) Step responses of force control loop modelled in the computer during parameter search.



b) Step response of force control loop with optimised parameters in connection with the plant

Application example hydraulic materials testing machine.

Fig. 2.11

2.4 Estimador de dos niveles

En la sección 2.2 se comentó que el Filtro de Kalman podría funcionar adecuadamente en el caso de sistemas lentamente variables en el tiempo.

Para atacar el caso en que estas variaciones lentas se combinan con variaciones más rápidas o saltos abruptos, B. Wittenmark [11] propone un esquema de estimación de parámetros con dos niveles.

Se supone el sistema dinámico como en (2.1).

El primer nivel corresponde a un "estimador fino", constituido por el filtro de Kalman (2.6), si se supone un modelo aproximado de variación de parámetros como en (2.5).

Este nivel de estimación está actualizando recursivamente el estimador $\hat{\theta}_t$, y se encarga de "seguir" las variaciones lentas de θ_t .

El segundo nivel de estimación prueba sistemáticamente N modelos θ^i , invariantes, predefinidos.

La idea es que si se produce una variación abrupta del parámetro real, se detecte que el sistema tiene ahora un parámetro más próximo a alguno de los N modelos, y se corrija la estimación fina para lograr un seguimiento más rápido.

El cálculo del estimador "fino" ya ha sido discutido en 2.2.1, veremos el estimador "grueso".

Denominamos θ^i $i=1\dots N$, a los parámetros predefinidos; Y_t son las observaciones de entradas y salidas hasta el tiempo t .

Entonces:

$$p(\theta^i | Y_t) = p(\theta^i | Y_{t-1}, y_t) = \frac{p(y_t | \theta^i, Y_{t-1}) p(\theta^i | Y_{t-1})}{p(y_t | Y_{t-1})} \quad (2.37)$$

$$\therefore p(\theta^i | Y_t) = \frac{p(e_t | \theta^i, Y_{t-1}) p(\theta^i | Y_{t-1})}{p(y_t | Y_{t-1})} \quad (2.38)$$

si además se supone $e(t)$ gaussiana.

$$p(e_t | \theta^i, Y_{t-1}) = \frac{1}{2\sigma_e^2} e^{-(y_t - x_t^T \theta^i)^2 / 2\sigma_e^2} \quad (2.39)$$

$$y \quad p(y_t | Y_{t-1}) = \prod_{i=1}^N p(e_t | \theta^i, Y_{t-1}) p(\theta^i | Y_{t-1}) \quad (2.40)$$

Nótese que la ecuación (2.38) nos dá un esquema recursivo de actualización de $p(\theta^i | Y_t)$ con la llegada de la nueva observación.

Se proponen dos formas de determinar el modelo más probable:

- 1) encontrar i tal que sea $\max_i p(\theta^i | Y_t)$
- 2) maximizar el logaritmo de la función de verosimilitud:

$$L_i(t) = \sum_{j=1}^t [y_j - x_j^T \theta^i]^2 \quad (2.41)$$

Llamando θ_t^* al mejor modelo obtenido de esta manera, se propone corregir el estimador fino con uno de dos métodos:

- 1) si $\|\hat{\theta}_{t-1} - \theta_t^*\|$ es muy grande hacer crecer la matriz de ganancias K_t ó la de covarianzas P_t .
- 2) si $\|\hat{\theta}_{t-1} - \theta_t^*\|$ es muy grande forzar $\hat{\theta}_t$ a aproximarse a θ_t^* por ejemplo con un valor interpolado.

Observaciones:

- La inicialización de (2.38) puede hacerse suponiendo que los modelos son igualmente probables "apriori", con probabilidad $1/N$.

- Como los sistemas son variables en el tiempo hay que modificar las ecuaciones (2.38) y (2.41) para que solo influyan las observaciones más recientes, por ejemplo computando (2.38) en una ventana rectangular (1), entre los instantes $t+1-l$ y t . En la ecuación (2.41) se puede introducir un factor de olvido $\lambda < 1$ y queda la minimización de:

$$L_i(t) = \sum_{j=1}^t \lambda^{t-j} [y_j - x_j^T \theta^i]^2 \quad (2.42)$$

- Se tienen unos cuantos parámetros de diseño:
 - . El factor de olvido λ , o la ventana rectangular l .
 - . El número de modelos a checar: N .
 - . Los parámetros de los modelos fijos θ^i , $i=1 \dots N$.
 - . La forma en que la estimación "gruesa" modifica la estimación "fina".

Debe notarse que en los hechos se está haciendo una partición del espacio de parámetros en regiones próximas a los modelos fijos.

La idea de este estimador de dos niveles está comprendida dentro del esquema general de sistemas adaptivos con un identificador, ver por ejemplo Rajbman/Chadeev [7].

3. MINIMOS CUADRADOS CON OLVIDO VARIABLE

3.1 El algoritmo

3.1.1 Introducción

El algoritmo que estudiaremos con mayor profundidad es el de mínimos cuadrados con factor de olvido variable. La característica principal de este algoritmo es que el estimador se obtiene minimizando una función cuadrática en los errores de predicción a un paso. Dichos errores de predicción se pesan en forma creciente con el tiempo, dándole más relevancia entonces a las observaciones más cercanas al instante actual.

Esta idea puede clasificarse (según lo visto en el capítulo 2) como de "observación de un horizonte no acotado, con pesos relativos", puesto que se toman en cuenta todas las observa-

ciones hasta el presente.

Como en el caso de la "ventana rectangular", se jerarquizan más unas observaciones que otras.

En la ventana rectangular el peso de las observaciones pasadas que caían fuera de la ventana era nulo, y el de las observaciones interiores a la ventana era uniforme. En este caso todos los pesos son mayores que cero y crecientes con el tiempo; su forma de crecimiento, como veremos, está gobernada por una sucesión de números reales λ_i , $1 \leq i \leq t$, llamada: factor de olvido.

El estudio del algoritmo será en el siguiente orden: primero se lo describe; en la sección 3.2 se discute en caso particular: factor de olvido constante, que permite extraer ciertas ideas básicas y definir conceptos como el de ventana efectiva y velocidad de seguimiento del cambio en los parámetros. Además se realizan algunas simulaciones que servirán como patrón de referencia para comparar distintos métodos de elección del factor de olvido (no constante).

En la sección 3.3 se discute detalladamente la dinámica del error de estimación, se estudia la convergencia del estimador y se plantean algunos problemas vinculados a la elección del factor de olvido y al efecto de mal condicionamiento numérico

del problema, si las señales del sistema no son adecuadas.

En la sección 3.4 se discuten y comparan distintos métodos para elegir el factor de olvido y para detectar cuando hubo cambios en el sistema que origina los datos.

3.1.2 Mínimos cuadrados pesados y con factor de olvido

Consideraremos el sistema (2.1), representado por la ecuación (2.4):

$$y_t = x_t^T \theta_t + e_t \quad (3.1)$$

donde hemos mantenido la notación del capítulo 2, definida en las ecuaciones (2.1), (2.2), (2.3) y (2.4).

Consideraremos la función de costo $J_t(\hat{\theta})$ definida por:

$$J_t(\hat{\theta}) = \sum_{i=1}^t \alpha_{i_t} (y_i - x_i^T \hat{\theta})^2 \quad (3.2)$$

donde los "pesos" α_{i_t} son positivos, no decrecientes y dependen del instante actual t .

$$\hat{\theta}_t = (X_t^T X_t)^{-1} X_t^T y_t \quad (3.7)$$

Las ecuaciones (3.6) se denominan ecuaciones normales por su interpretación geométrica, que no discutiremos ahora. Resolverlas, y encontrar $\hat{\theta}_t$ de acuerdo a (3.7) es un problema numéricamente complejo, pero en el siguiente TEOREMA 3.1 veremos un esquema recursivo (de Plackett), que permite resolver las ecuaciones normales en forma bastante sencilla.

El estimador de mínimos cuadrados pesados da una posibilidad interesante: si se eligen los pesos crecientes (como lo hemos hecho), permite abordar el problema de identificación de sistemas variables en el tiempo, ya que se da más importancia a las últimas observaciones. Se obtiene por lo tanto un estimador $\hat{\theta}_t$ que ajusta más las observaciones cercanas al instante actual, que las pasadas.

Esta idea, de pesar más las observaciones más confiables por cierto no es nueva, está explícitamente mencionada en el primer trabajo publicado acerca del método de mínimos cuadrados, por K.F. GAUSS en 1809 (ver ref. [37]). Incluso en este trabajo se menciona el uso de distintos pesos relativos en trabajos realizados en 1795 (sin publicar) para determinar los parámetros de las órbitas de cuerpos celestes tomando en cuenta que distintas observaciones tenían distinta precisión en su medida.

Observación:

Dada la sucesión $0 < \alpha_{1_t} \leq \alpha_{2_t} \leq \dots \leq \alpha_{t_t} = 1$, existe otra sucesión $\{\lambda_{i_t}\}$, $i = 1, 2, \dots, t-1$ con $0 < \lambda_{i_t} \leq 1$ tal que

$$\alpha_{i_t} = \prod_{j=i}^{t-1} \lambda_{j_t}, \quad i=1, 2, \dots, t-1 \quad (3.8)$$

Para demostrarlo, basta con construir la sucesión:

$$\lambda_{j_t} = \frac{\alpha_{j_t}}{\alpha_{j+1_t}}, \quad j=1, 2, \dots, t-1$$

luego, por (3.3) $\lambda_{j_t} \leq 1$.

Solamente consideraremos sucesiones $\{\alpha_{i_t}\}$ tales que $\lambda_{j_t} = \lambda_j$ ($j=1, 2, \dots$), solo es función del índice j y no del instante actual t .

A este caso particular de selección de pesos lo denominaremos algoritmo de mínimos cuadrados con factor de olvido variable.

A la sucesión λ_j , $i=1, 2, \dots$, la denominaremos sucesión de factores de olvido.

TEOREMA 3.1 (Algoritmo de Plackett, ver [38]).

El estimador de mínimos cuadrados con factor de olvido variable que minimiza la función de costo (3.2) con $\alpha_{i_t} = \lambda_i \dots \lambda_t$, satisface la recursión:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{P_t x_{t+1}}{1 + x_{t+1}^T P_t x_{t+1}} (y_{t+1} - x_{t+1}^T \hat{\theta}_t) \quad (3.9)$$

$$P_{t+1} = \frac{1}{\lambda_{t+1}} \left[P_t - \frac{P_t x_{t+1} x_{t+1}^T P_t}{1 + x_{t+1}^T P_t x_{t+1}} \right] \quad (3.10)$$

$$\forall t \geq p \quad (p = \dim \theta)$$

Inicialización

La recursión (3.7) (3.8) es válida para $t \geq p$ (p = número de parámetros).

Existen dos maneras usuales de inicializar el algoritmo:

- i) Inicialización exacta. Se resuelven las ecuaciones normales que dan lugar a la recursión, para $t = p$. Se obtienen así $\hat{\theta}_p$ y P_p ; con estos valores se prosigue la recursión para $t > p$.

- ii) Inicialización arbitraria. Se inicializa el algoritmo

con $\hat{\theta}_0$ finito (usualmente $\hat{\theta}_0 = 0$) y $P_0 \gg I$, (usualmente $P_0 = kI$ con $k \gg 1$). En tal caso, llamando $\hat{\theta}_t^*$ y P_t^* los valores de $\hat{\theta}_t$ y P_t obtenidos con esta inicialización, resulta que $\hat{\theta}_t^* \rightarrow \hat{\theta}_t$ y $P_t^* \rightarrow P_t$, rápidamente.

La demostración del teorema 3.1 se omite por razones de espacio, puede ser hallada en las referencias [16], [17] y [20].

TEOREMA 3.2

La función de costo (3.2), que dá lugar al algoritmo de mínimos cuadrados con factor de olvido, evaluada en el estimador (3.7), satisface la recursión:

$$J_{t+1}(\hat{\theta}_{t+1}) = \lambda_t J_t(\hat{\theta}_t) + \frac{\epsilon_{t+1}^2}{1 + x_{t+1}^T P_t x_{t+1}} \quad (3.9)$$

Donde ϵ_{t+1} es el error de predicción "a priori", definido como:

$$\epsilon_{t+1} = y_{t+1} - y(t+1/\hat{\theta}_t) = y_{t+1} - x_{t+1}^T \hat{\theta}_t \quad (3.10)$$

La demostración de este teorema también se omite por razones de espacio, puede encontrarse en las referencias [13] y [17].

Observaciones

El algoritmo de mínimos cuadrados con factor de olvido, y el esquema de cálculo recursivo de Plackett tienen características atractivas:

- i) Se actualiza el estimador a partir del valor anterior y las observaciones x_{t+1} , y_{t+1} que llegan en el nuevo instante de muestreo.
- ii) El cálculo tiene una gran sencillez numérica, especialmente si se le compara con el problema original de solución de las ecuaciones normales. Para resolver las ecuaciones normales es necesario invertir una matriz de dimensión $p \times p$ que resulta del producto de dos matrices $p \times t$ y $t \times p$ respectivamente.

Puede comprobarse (ver (3.7) y (3.8)) que en el esquema recursivo se evita la inversión de matrices y que el cálculo de $\hat{\theta}_t$ y P_t es sencillo.

Además las necesidades de memoria son limitadas (no hay arreglos de dimensión creciente), y la matriz P_t es simétrica.

Estos hechos hacen que el algoritmo sea apropiado para

aplicaciones en tiempo real (sistemas en línea).

iii) La matriz P_t es:

$$P_t = \left[\sum_{i=1}^t (\lambda_i \dots \lambda_t) x_i x_i^T + P_0^{-1} (\lambda_1 \dots \lambda_t) \right]^{-1} \quad (3.11)$$

Esta matriz se denomina usualmente "matriz de covarianzas". El nombre se debe a que si $\lambda_t = 1 \quad \forall_t$, bajo ciertas condiciones de regularidad (excitación persistente), puede demostrarse que $\sigma^2 P_t$ representa asintóticamente ($t \rightarrow \infty$) la matriz de covarianzas del estimador $\hat{\theta}_t$.

En el caso que nos ocupa: $\lambda_t < 1 \quad \forall_t$, esta propiedad ya no es cierta, a pesar de lo cual muchos autores mantienen la misma denominación para P_t .

iv) Como resultado de que $\lambda_t < 1$, el algoritmo puede hacer un "seguimiento" de la variación de los parámetros del sistema, al ir descartando paulatinamente los datos pasados, por la vía de "pesarlos" menos que los datos presentes.

Cómo elegir el valor adecuado de λ_t en cada instante, es un problema aún abierto.

Distintas maneras de elegir λ_t se estudiarán en las siguientes

tes secciones así como las ventajas o desventajas de cada una de ellas.

Antes de un estudio más formal, veremos la elección de λ_t más sencilla: $\lambda_t = \lambda$ (constante) $< 1 \quad \forall_t$.

Con esto buscamos entender el efecto de seleccionar distintos valores del factor de olvido constante λ .

3.2 Factor de olvido constante

La forma más extendida de selección de λ_t es la de asignarle un valor constante $\lambda_t = \lambda < 1$. También se le llama "olvido exponencial" o "ventana exponencial", por razones que veremos inmediatamente.

La función de costo minimizada por el algoritmo es (3.2) con α_{i_t} dado por $\alpha_{i_t} = \frac{1}{\pi} \sum_{j=1}^{t-1} \lambda^j = \lambda^{t-i}$, luego:

$$J_t(\hat{\theta}_t) = \sum_{i=1}^t \lambda^{t-i} (y_i - x_i^T \hat{\theta}_t)^2 \quad (3.12)$$

El peso de las observaciones decrece en forma exponencial para $\lambda < 1$.

Cuanto más próximo a cero sea λ , más rápidamente se dejan de tomar en cuenta los datos pasados. Cuanto más próxima a 1

sea λ , más uniforme es el peso de los datos.

El valor de λ es un parámetro de diseño del algoritmo. La figura 3.1 muestra el decaimiento del peso de las observaciones para distintos valores de λ : $0.95 \leq \lambda \leq 0.995$.

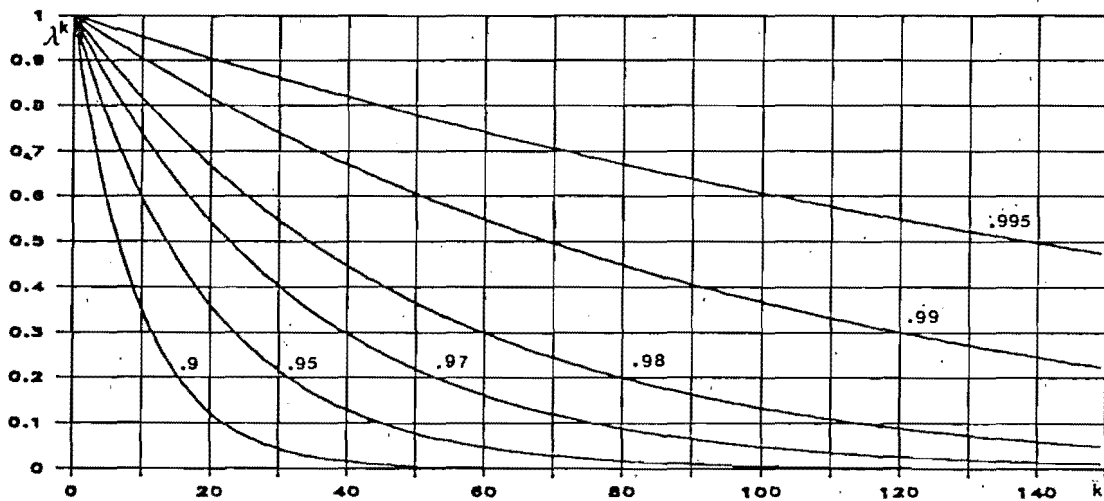


Fig. 3.1

3.2.1 Velocidad de seguimiento. Ventana efectiva

El valor de λ modula la velocidad con que el algoritmo sigue las variaciones en los parámetros del sistema.

Si $0 < \lambda < 1$, cuanto más próximo a 0 está λ , mayor es la rapidez con que se siguen los cambios del parámetro del sistema.

La dependencia entre la adaptabilidad y el valor de λ puede cuantificarse.

Tiempo de subida (o de levantamiento)

Supongamos que $\theta_t = \theta_1$ (constante) si $t \leq T$, y que se produce un salto abrupto a $\theta_t = \theta_2 \quad \forall t > T$.

Para que el algoritmo haga que $\hat{\theta}_t \rightarrow \theta_2$ ($t > T$), es necesario que las observaciones generadas por el nuevo parámetro pesen más que las generadas por el anterior.

Si $t \rightarrow \infty$, para que las últimas k observaciones pesen una fracción ρ (en por unidad) del total de los pesos, dado que:

$$\sum_{i=t-k+1}^t \lambda^{t-i} = \frac{1 - \lambda^k}{1 - \lambda} \quad (\text{peso de las últimas } k) \quad (3.13)$$

deben ser:

$$\frac{1 - \lambda^k}{1 - \lambda} = \rho \cdot \sum_{-\infty}^t \lambda^{t-i} = \rho \frac{1}{1 - \lambda} \quad (3.14)$$

$$\therefore \lambda^k = 1 - \rho \quad (3.15)$$

Una magnitud que puede dar idea de la velocidad con que se sigue un cambio en el parámetro, es el valor de k (no. de pasos) necesario para que $\rho = 0.9$. Esto es una especie de extensión del concepto de "tiempo de subida" (rise-time), normalmente

usado en ingeniería.

La figura 3.2 muestra los valores $\lambda = \lambda(k)$ para $\rho = 0.9$. Al parámetro $k = k(\lambda)$ así definido, lo denominaremos "tiempo de subida", o de "levantamiento". (N_1). Resulta entonces, de (3.15), que

$$N_1 = \ln(0.1) / \ln(\lambda) \quad (3.16)$$

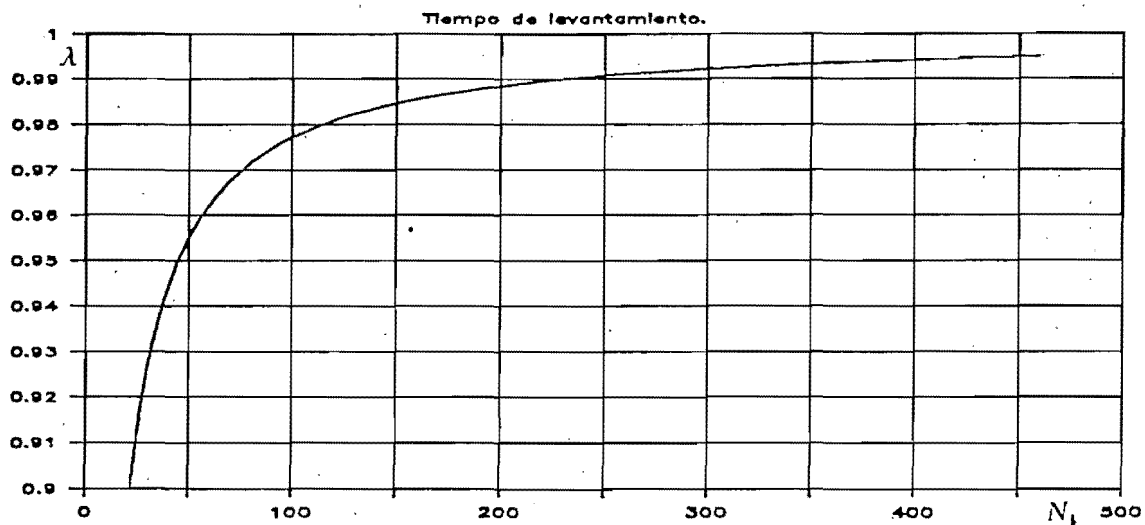


Fig. 3.2

Ventana efectiva. Otra magnitud importante es la que se deriva del concepto de "constante de tiempo". A este valor lo llamaremos "ventana efectiva" (N_{ef}), y lo definimos como:

$$N_{ef} = - \frac{1}{\left. \frac{d\lambda^\tau}{d\tau} \right|_{\tau=0}} \triangleq - \frac{1}{\ln(\tau)}$$

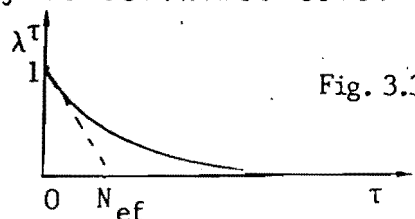


Fig. 3.3

$$\therefore N_{ef} = \frac{1}{\ln(\tau)} \quad (3.17)$$

Si $\lambda = 1$ (que será usualmente nuestro caso), entonces (3.17) se puede aproximar

$$N_{ef} \approx \frac{\lambda}{1-\lambda} \quad (3.18)$$

$$N_{ef} \cong \frac{1}{1-\lambda} \quad (3.19)$$

De la ecuación (3.15) puede deducirse que N_{ef} es el valor de k que corresponde a $\rho = 1 - e^{-1} = 0.6321$, es decir que los últimos N_{ef} pesos, suman un 63.21% del total.

El nombre de ventana efectiva se desprende de otra forma de deducir este valor. Consideremos la función de costo (3.12). La suma de todos los coeficientes (pesos) de dicha función es (asintóticamente):

$$\lim_{t \rightarrow \infty} \frac{1 - \lambda^t}{1 - \lambda} = \frac{1}{1 - \lambda} \quad (3.20)$$

Compárese (3.20) con (3.19). Entonces, el número de coeficientes iguales a 1 equivalente a infinitos coeficientes menores que 1 de la función de costo es la "ventana efectiva" N_{ef} .

Por otro lado, de (3.16) y (3.17) se observa que

$$N_1 \cong 2.3 N_{ef} \quad (3.21)$$

Observación

Como se ha visto, el asignar valores bajos a λ se traduce en una tendencia a seguir más rápidamente la variación del parámetro del sistema que genera los datos. La contrapartida es que al reducir la "ventana efectiva", la estimación del parámetro se hace sobre la base de un menor número de observaciones, lo que se traduce en una mayor influencia del ruido e_t en las observaciones. El resultado es un estimador "menos suave", con mayor incidencia de las perturbaciones. A menores valores de λ , mayor es la covarianza del estimador $\hat{\theta}_t$. Más adelante volveremos más formalmente sobre este problema.

En resumen la elección de λ se basa en el compromiso entre la precisión con la que se desea estimar los parámetros, y la velocidad de adaptación del algoritmo.

3.2.2 Simulación

Para ejemplificar las ideas expuestas, se muestran resultados de simulaciones. El sistema simulado fué:

$$\begin{cases} y_t - 0.9 y_{t-1} = u_{t-1} + \sigma e_t & (1 \leq t \leq 299) \\ y_t - 0.4 y_{t-1} = u_{t-1} + \sigma e_t & (300 \leq t \leq 500) \end{cases} \quad (3.22)$$

Se usaron distintos valores de σ para obtener distintos valores de la relación señal a ruido en la observación de la salida y_t .

σ	S/N
0.2	25
0.4	6.25
0.707	2

La señal u_t es una sucesión binaria pseudo aleatoria de longitud máxima con período de 127 muestras, y e_t es una realización (simulada) de ruido blanco gaussiano, con media cero y varianza unitaria.

Todas las simulaciones de los algoritmos de identificación se hicieron sobre las mismas señales del sistema (3.22) para poder comparar los resultados obtenidos.

Los resultados con factor de olvido constante servirán como patrón de referencia.

Las figuras (3.4), (3.5) y (3.6) muestran los estimados de

los parámetros del sistema (3.22), obtenidos con mínimos cuadrados con factor de olvido constante.

La figura (3.4) corresponde a una relación señal a ruido $(S/N) = 25$, la fig (3.5) a $S/N = 6.25$, y la figura (3.6) a $S/N = 2$.

En cada caso se aplicaron 3 factores de olvidos $\lambda_1 = 0.99$, $\lambda_2 = 0.98$ y $\lambda_3 = 0.96$.

En las 9 simulaciones se puede verificar las tendencias previstas:

- . a mayor nivel del ruido, mayor es la imprecisión (varianza) de la estimación.
- . a menor valor del factor de olvido, mayor es la velocidad de adaptación de algoritmo, a costa de una mayor incidencia de las perturbaciones; y por lo tanto de una mayor imprecisión en la estimación.
- . finalmente, se puede observar que N_1 (tiempo de levantamiento), es una buena medida de la velocidad de adaptación. En efecto: de acuerdo a 3.16, los valores de N_1 que corresponden a $\lambda = 0.99$, 0.98 y 0.96 son 229, 114 y 56 respectivamente. Compárese estos valores con la for-

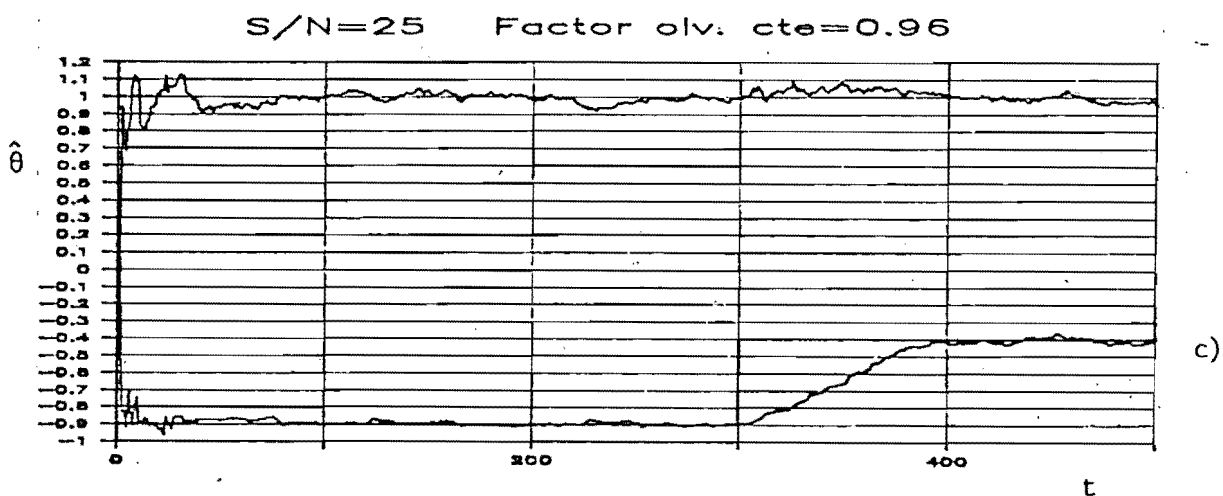
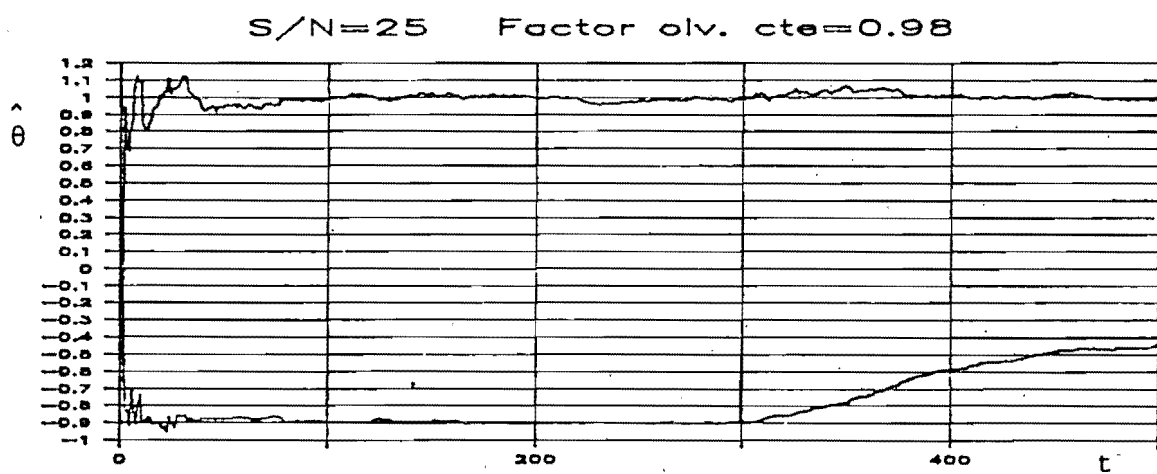
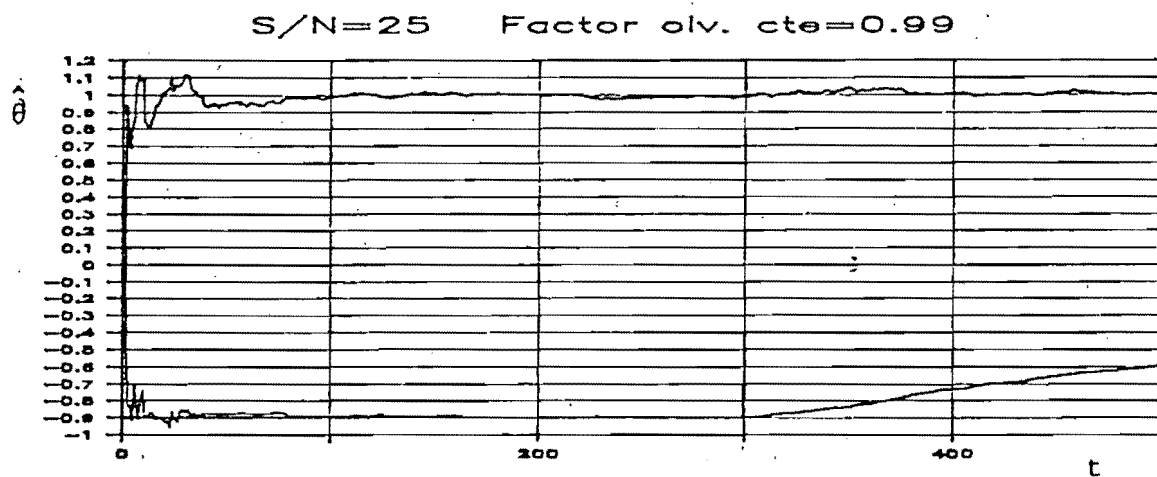
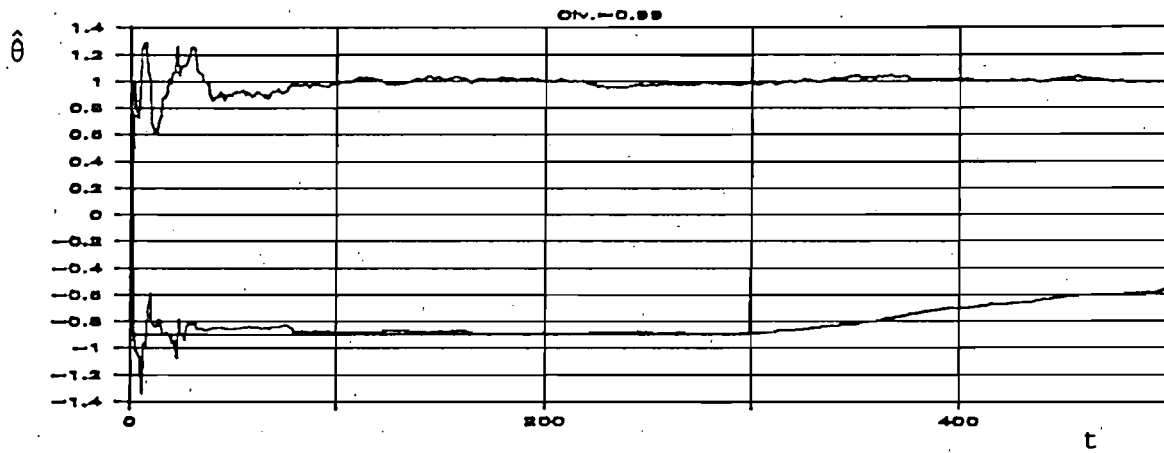
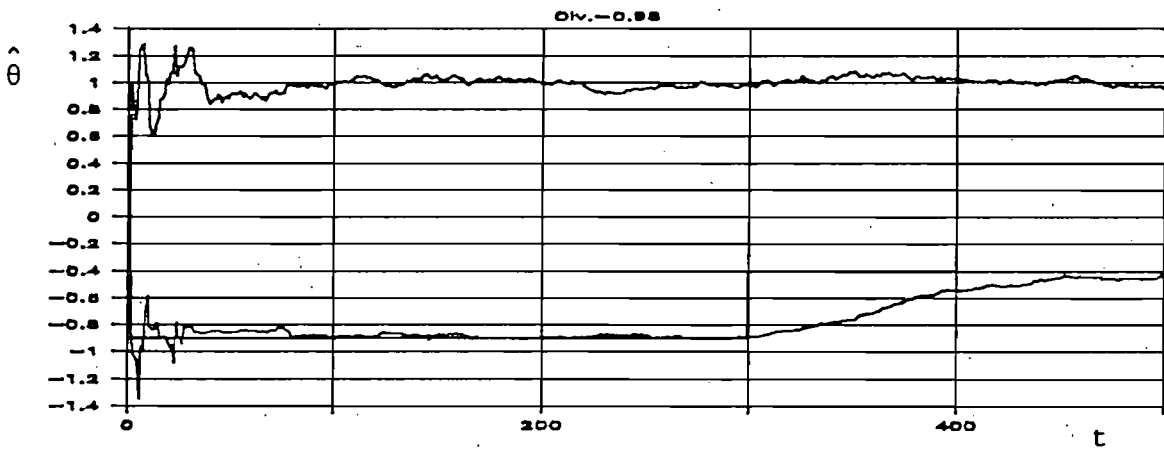


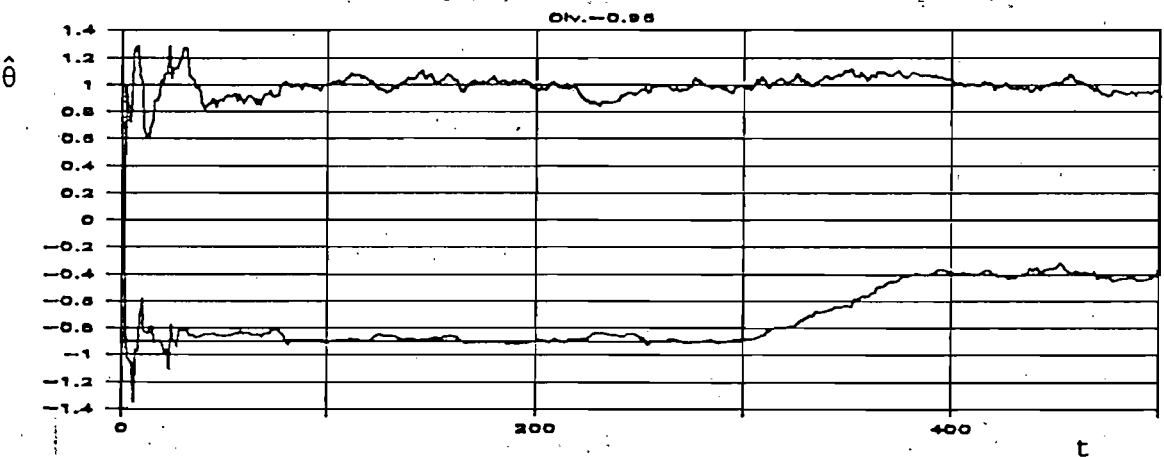
Figura 3.4



a)



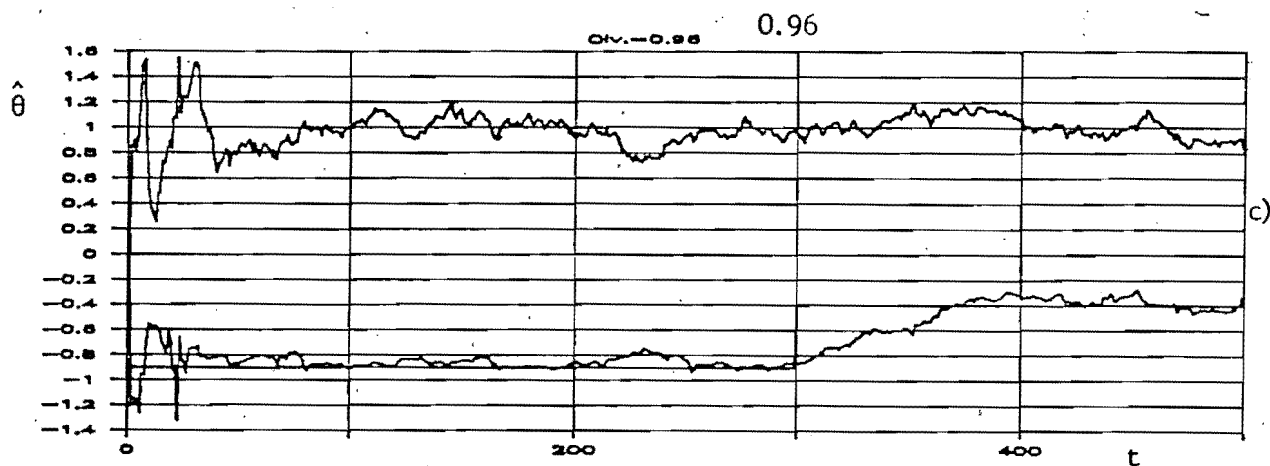
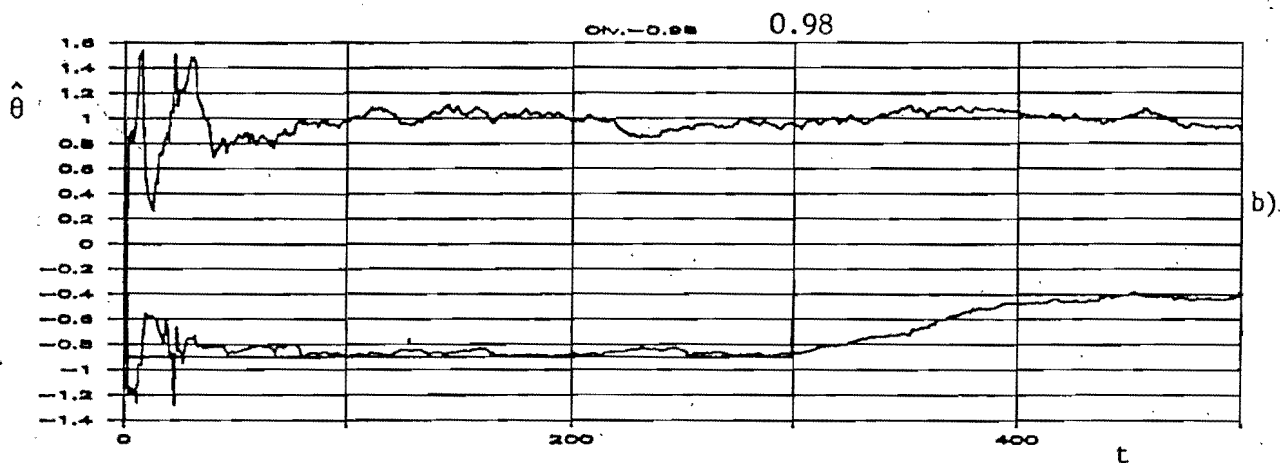
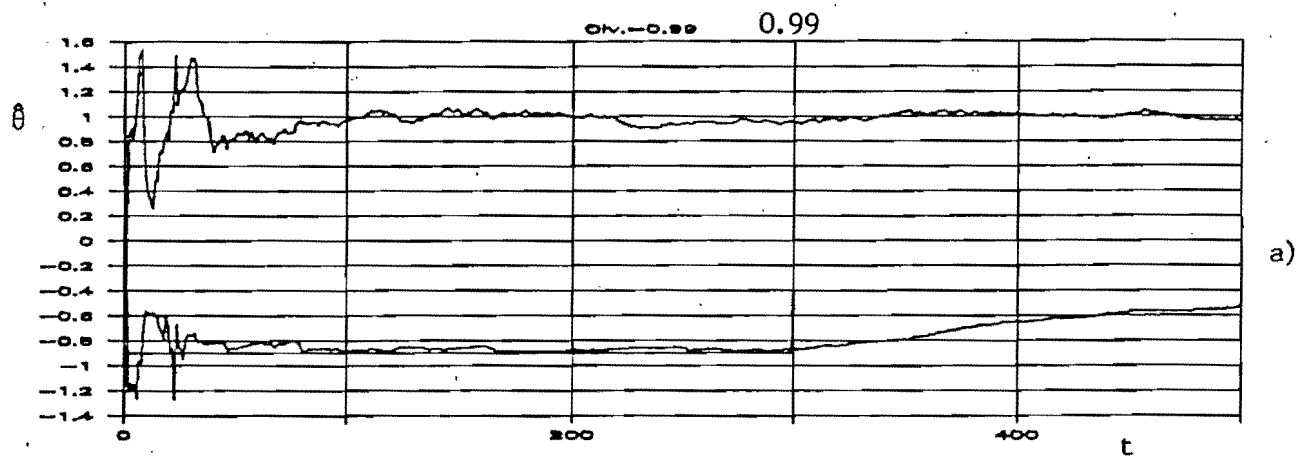
b)



c)

Relación señal a ruido $S/N=6.25$

Figura 3.5



Relación señal a ruido $S/N=2$.
Figura 3.6

ma en que el parámetro que cambia en el tiempo, es seguido por el algoritmo en todos los casos.

Puede observarse que la velocidad de seguimiento, no depende de la relación señal a ruido, sino que depende del valor del factor de olvido. El problema es que no se puede elegir esta velocidad independientemente del problema concreto de identificación. En efecto: si eligieramos por ejemplo $\lambda = 0.96$, los resultados obtenidos con $S/N = 25$ (fig 3.4.c.) pueden considerarse sumamente satisfactorios para un gran rango de aplicaciones, pero si la relación señal a ruido fuera 2 (fig 3.6.c), las estimaciones aparecen fuertemente perturbadas, quizá más de lo que podamos admitir en una aplicación específica.

Aunque conociéramos aproximadamente el valor de la varianza de las perturbaciones, el hecho de estar tratando con sistemas variables en el tiempo nos pondría ante la dificultad de que no conocemos el valor de la relación señal a ruido.

Por esta clase de problemas hay que recurrir a algoritmos del estilo de mínimos cuadrados con factor de olvido variable.

El problema es cómo seleccionar la sucesión λ_t , de manera de aprovechar todas las observaciones cuando no hay cambios en los parámetros, y detectar dichos cambios, disminuyendo el factor de olvido, para aumentar la velocidad de seguimiento.

Este problema es sumamente complejo y el encontrar una solución óptima sigue siendo un problema abierto. Para el caso determinístico ($\sigma = 0$), existen por supuesto, bastantes más resultados que para el caso estocástico.

En la siguiente sección se estudia la evolución del estimador $\hat{\theta}_t$, en relación a las observaciones y al factor de olvido, y posteriormente se discuten distintos esquemas de selección de λ_t .

3.3 Dinámica del error de estimación

Dado un sistema dinámico variable en el tiempo, regido por:

$$y_t = x_t^T \theta_t + e_t \quad (3.23)$$

con $\left\{ \begin{array}{l} x_t, \theta_t \in R^p \\ e_t : \text{v.a.} \end{array} \right.$

con $\left\{ \begin{array}{l} E\{e_t\} = 0 \\ E\{e_i e_j\} = \sigma^2 \delta_{ij} \end{array} \right.$

Definimos:

$$\left\{ \begin{array}{l} \Delta \theta_t = \theta_t - \theta_{t-1} \quad (\text{variación del parámetro en el instante } t) \\ \tilde{\theta}_t = \theta_t - \hat{\theta}_t \quad (\text{error de estimación en el instante } t) \\ \varepsilon_t = y_t - \hat{y}_{t/t-1} \stackrel{\Delta}{=} y_t - x_t^T \hat{\theta}_{t-1} \quad (\text{error "a priori" de predicción en } t) \end{array} \right. \quad (3.24)$$

$$\left\{ \begin{array}{l} \tilde{\theta}_t = \theta_t - \hat{\theta}_t \quad (\text{error de estimación en el instante } t) \\ \varepsilon_t = y_t - \hat{y}_{t/t-1} \stackrel{\Delta}{=} y_t - x_t^T \hat{\theta}_{t-1} \quad (\text{error "a priori" de predicción en } t) \end{array} \right. \quad (3.25)$$

$$\left\{ \begin{array}{l} \varepsilon_t = y_t - \hat{y}_{t/t-1} \stackrel{\Delta}{=} y_t - x_t^T \hat{\theta}_{t-1} \quad (\text{error "a priori" de predicción en } t) \end{array} \right. \quad (3.26)$$

$$\bar{\varepsilon}_t = y_t - \hat{y}_{t/t} \stackrel{\Delta}{=} y_t - x_t^T \hat{\theta}_t \quad (\text{error "a posteriori" en } t) \quad (3.27)$$

de (3.24) y (3.25) se desprende que:

$$\hat{\theta}_t = \theta_{t-1} + \Delta \hat{\theta}_t - \tilde{\theta}_t \quad (3.28)$$

considerando el estimador de mínimos cuadrados con factor de olvido:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \lambda_{t+1} P_{t+1} x_{t+1} (y_{t+1} - x_{t+1}^T \hat{\theta}_t) \quad (3.29)$$

$$P_{t+1} = \frac{1}{\lambda_{t+1}} \left[P_t - \frac{P_t x_{t+1} x_{t+1}^T P_t}{1 + x_{t+1}^T P_t x_{t+1}} \right] \quad (3.30)$$

de (3.28) y (3.29):

$$\theta_t + \Delta \theta_{t+1} - \tilde{\theta}_{t+1} = \hat{\theta}_t + \lambda_{t+1} P_{t+1} x_{t+1} [y_{t+1} - x_{t+1}^T \hat{\theta}_t] \quad (3.31)$$

Entonces:

$$\begin{aligned} \tilde{\theta}_{t+1} &= \tilde{\theta}_t + \Delta \theta_{t+1} - \lambda_{t+1} P_{t+1} x_{t+1} [x_{t+1}^T \theta_t + \\ &\quad + x_{t+1}^T \Delta \theta_{t+1} + e_{t+1} - x_{t+1}^T \hat{\theta}_t] = \\ \tilde{\theta}_{t+1} &= [I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T] (\tilde{\theta}_t + \Delta \theta_{t+1}) - \\ &\quad - \lambda_{t+1} P_{t+1} x_{t+1} e_{t+1} \end{aligned} \quad (3.32)$$

La ecuación (3.32) representa la dinámica del error de estimación $\tilde{\theta}_t$. Se trata de un sistema dinámico, lineal, de entradas: $\Delta\theta_{t+1}$ y e_{t+1} y variante en el tiempo.

Tanto la matriz de transición, como la matriz asociada a la entrada e_{t+1} dependen del "vector de observaciones" x_{t+1} y de la "matriz de covarianza" P_{t+1} .

3.3.1. Superposición

La ecuación (3.32) permite descomponer la solución por superposición, en tres partes con una interpretación inmediata; definiendo:

$$\tilde{\theta}_t = \tilde{\theta}_t^1 + \tilde{\theta}_t^2 + \tilde{\theta}_t^3 \quad (3.33)$$

con

$$\left\{ \begin{array}{l} \tilde{\theta}_0^1 = \tilde{\theta}_0 \\ \tilde{\theta}_0^2 = 0 \\ \tilde{\theta}_0^3 = 0 \end{array} \right.$$

$$\tilde{\theta}_{t+1}^1 = [I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T] \tilde{\theta}_t^1 \quad a)$$

$$\tilde{\theta}_{t+1}^2 = [I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T] (\tilde{\theta}_t^2 + \Delta\theta_{t+1}) \quad b) \quad (3.34)$$

$$\tilde{\theta}_{t+1}^3 = [I - \lambda_{t+1} p_{t+1} x_{t+1} x_{t+1}^T] \tilde{\theta}_t^3 - \lambda_{t+1} p_{t+1} x_{t+1} e_{t+1} \quad c)$$

$\tilde{\theta}_t^1$ es la parte debida al error en la condición inicial (respuesta libre), $\tilde{\theta}_t^2$ el error de estimación debido a los cambios en el parámetro, y $\tilde{\theta}_t^3$ es el error de estimación debido a la incidencia de las perturbaciones en el proceso de identificación.

Esta descomposición permite estudiar por separado los efectos debidos a las tres causas que originan el error de estimación.

Aunque estudiaremos las tres contribuciones, usualmente sólo se toman en cuenta $\tilde{\theta}_t^1$ y $\tilde{\theta}_t^3$. Es decir que se suponen los cambios en el parámetro lo suficientemente espaciados en el tiempo como para permitir la convergencia entre cambio y cambio.

Se trata este problema como de condiciones iniciales no nulas ($\tilde{\theta}_0 = \Delta\theta$). Análogamente, el error de predicción a un paso, al ser lineal en el error de estimación y el proceso del ruido, admite una descomposición similar.

Definiendo:

$$\epsilon_{t+1} = \epsilon_{t+1}^1 + \epsilon_{t+1}^2 + \epsilon_{t+1}^3 \quad (3.35)$$

entonces, por (3.26):

$$\epsilon_{t+1} = y_{t+1} - x_{t+1}^T \theta_t = x_{t+1}^T (\theta_t + \Delta\theta_{t+1}) + e_{t+1} - x_{t+1}^T \theta_t$$

$$\epsilon_{t+1} = x_{t+1}^T (\tilde{\theta}_t + \Delta\theta_{t+1}) + e_{t+1}$$

y convenimos en definir:

$$\left\{ \begin{array}{l} \epsilon_{t+1}^1 = x_{t+1}^T \tilde{\theta}_t^1 \\ \epsilon_{t+1}^2 = x_{t+1}^T (\tilde{\theta}_t^2 + \Delta\theta_{t+1}) \\ \epsilon_{t+1}^3 = x_{t+1}^T \tilde{\theta}_t^3 + e_{t+1} \end{array} \right. \quad (3.36)$$

3.3.2 Matríz de transición

La respuesta libre $\tilde{\theta}_t^1$ (ver ecs. 3.34), está totalmente determinada por $\tilde{\theta}_0$ y la matríz de transición. Esta matríz está presente como matríz de transición en las tres partes en que hemos descompuesto el error de estimación, razón por la que la estudiaremos antes de seguir adelante.

De la forma de la matríz A_t :

$$A_{t+1} = [I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T] \quad (3.37)$$

es inmediato que se trata de una "matriz elemental" en el sentido de Householder [36] (ver apéndice A), del tipo:

$$A_{t+1} = I - \rho_{t+1} u_{t+1} v_{t+1}^T \quad \text{donde} \quad \begin{cases} \rho_{t+1} = \lambda_{t+1} & \in R^1 \\ u_{t+1} = P_{t+1} x_{t+1} & \in R^p \\ v_{t+1} = x_{t+1} & \in R^p \end{cases} \quad (3.38)$$

por lo tanto (ver apéndice), A_{t+1} es una matriz cuadrada ($p \times p$), que tiene: $p-1$ autovalores unitarios asociados al hiperplano de R^p ortogonal al vector x_{t+1} de observaciones, y un autovalor λ_p asociado al autovector $u_{t+1} = P_{t+1} x_{t+1}$, cuyo valor es:

$$\lambda_p = 1 - \rho_{t+1} v_{t+1}^T u_{t+1} = 1 - \lambda_{t+1} x_{t+1}^T P_{t+1} x_{t+1} = (\text{por (3.30)}) =$$

$$\lambda_p = \frac{1}{1 + x_{t+1}^T P_t x_{t+1}} \quad (3.39)$$

Por (3.11), es obvio que eligiendo P_0 positiva definida (así debe hacerse), entonces P_t también es positiva definida. De este hecho y de (3.39) resulta que $\lambda_p < 1$.

En resúmen, en cada recursión de (3.34.a), al pasar de $\tilde{\theta}_t^1$ a $\tilde{\theta}_{t+1}^1$, la matriz de transición A_{t+1} mantiene invariante la componente de $\tilde{\theta}_t^1$ contenida en el hiperplano ortogonal a x_{t+1} , y se contrae en la componente colineal con $P_{t+1} x_{t+1}$.

El hecho de que la componente de $\tilde{\theta}_t^1$ contenida en \mathcal{F} se mantiene invariante, se debe a que es precisamente la parte ortogonal al vector de observaciones x_{t+1} , que el algoritmo utiliza para la actualización a través del error de predicción:

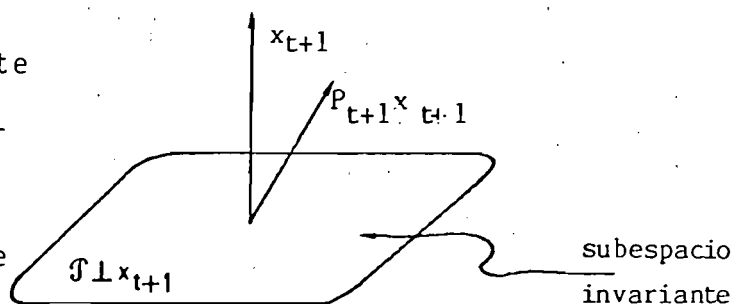


Fig. 3.7

$$\varepsilon_{t+1} = y_{t+1} - x_{t+1}^T \hat{\theta}_t$$

El hecho de que la contracción sea en la dirección de $P_{t+1} x_{t+1}$ se debe a que esa es la dirección de actualización del estimador, ver ec. (3.29).

Cabe observar que:

- i) $P_{t+1} x_{t+1}$ es colineal con $P_t x_{t+1}$. En efecto, de (3.30) resulta que

$$\lambda_{t+1} P_{t+1} x_{t+1} = \frac{1}{1 + x_{t+1}^T P_t x_{t+1}} P_t x_{t+1} \quad (3.40)$$

- ii) El problema geométrico no degenera, es decir x_{t+1} y $P_{t+1} x_{t+1}$ no son ortogonales, mientras se asegure que P_t sea positiva definida.

En efecto: x_{t+1} y $P_{t+1} x_{t+1}$ no son ortogonales si no lo son x_{t+1} y $P_t x_{t+1}$, pero $x_{t+1}^T P_t x_{t+1} > 0$ si P_t es positiva definida.

Este sería el caso de "mal condicionamiento" del problema de mínimos cuadrados, cosa que no admitiremos; P_t es siempre positiva definida.

En secciones siguientes veremos como se asegura esto.

Para la respuesta libre, y para la contribución de los saltos en el parámetro sería ideal que A_{t+1} contrayera el espacio lo más posible. El problema es que como A_{t+1} no es simétrica, a pesar de que todos sus autovalores son menor o iguales a la unidad, no se puede asegurar la contracción del error en cada paso.

Para estudiar esto hay que analizar el problema de los valores singulares de A_{t+1} . Desafortunadamente no es tan sencillo como el problema de los valores característicos.

Veamos que sucede "en norma" euclidiana (2), al transformar con A_{t+1} el vector genérico z ; llamando w al transformado de z :

$$\|w\|^2 = w^T w = z^T z - \frac{z^T x_{t+1} x_{t+1}^T P_t z + z^T P_t x_{t+1} x_{t+1}^T z}{1 + x_{t+1}^T P_t x_{t+1}} +$$

$$\|w\|^z = \|z\|^2 + \frac{z^T x_{t+1} x_{t+1}^T P_t P_t x_{t+1} x_{t+1}^T z}{(1 + x_{t+1}^T P_t x_{t+1})^2} - \frac{2(z^T x_{t+1})(x_{t+1}^T P_t z)}{1 + x_{t+1}^T P_t x_{t+1}} + \frac{(z^T x_{t+1})^2 x_{t+1}^T P_t P_t x_{t+1}}{(1 + x_{t+1}^T P_t x_{t+1})^2} \quad (3.41)$$

para que $\|w\| \leq \|z\|$, se debe cumplir que:

$$\frac{2(z^T x_{t+1})(x_{t+1}^T P_t z)}{1 + x_{t+1}^T P_t x_{t+1}} \geq \frac{(z^T x_{t+1})^2 x_{t+1}^T P_t P_t x_{t+1}}{(1 + x_{t+1}^T P_t x_{t+1})^2} > 0 \quad (3.42)$$

$$\therefore \begin{cases} \text{si } z^T x_{t+1} \geq 0, & x_{t+1}^T P_t z \geq \frac{(z^T x_{t+1})(x_{t+1}^T P_t P_t x_{t+1})}{1 + x_{t+1}^T P_t x_{t+1}} \\ \text{si } z^T x_{t+1} < 0, & x_{t+1}^T P_t z \leq \frac{(z^T x_{t+1})(x_{t+1}^T P_t P_t x_{t+1})}{1 + x_{t+1}^T P_t x_{t+1}} \end{cases} \quad (3.43)$$

La zona que viola las condiciones (3.43), es una "doble cuña", como la representada en R^3 en la figura (3.8), situada entre los hiperplanos $\mathcal{F} \perp x_{t+1}$, y Q definido como:

$$Q = \left\{ z: \frac{x_{t+1}^T P_t z}{z^T x_{t+1}} = \frac{(x_{t+1}^T P_t P_t x_{t+1})}{2(1 + x_{t+1}^T P_t x_{t+1})} \right\} \quad (3.44)$$

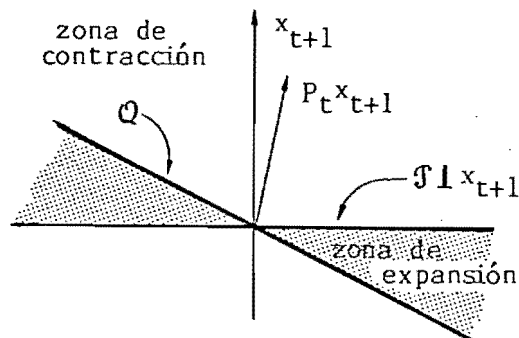


Fig. 3.8

Desde aquí se puede ver una parte de la dificultad para elegir una sucesión óptima λ_t . Aún cuando el sistema fuera determinístico, e invariante en el tiempo, elegir λ_t a un paso con base en el error de estimación ϵ_{t+1} lleva al siguiente problema:

Lo único que se conoce es de $\tilde{\theta}_t^1$ es:

$$\epsilon_{t+1} = \epsilon_{t+1}^1 = x_{t+1}^T \tilde{\theta}_t^1 = |x_{t+1}| |\tilde{\theta}_t^1| \cos\theta \quad (3.45)$$

Esto es que $\tilde{\theta}_t^1$ es un vector en algún lado de la zona representada en la fig (3.9). Pero no se sabe siquiera si pertenece a la zona de contracción o expansión a un paso. La elección

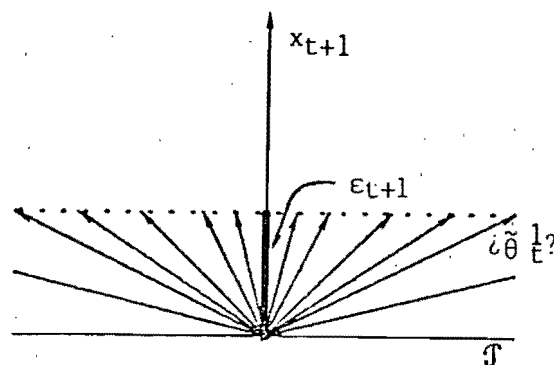


Fig. 3.9

de λ_t es crucial: El autovalor asociado a $P_t x_{t+1}$ es λ_p :

$$\lambda_p = \frac{1}{1 + x_{t+1}^T P_t x_{t+1}} = \frac{\lambda_t}{\lambda_t + x_{t+1}^T P_t^* x_{t+1}} \quad (3.46)$$

donde

$$P_t^* = \lambda_t P_t = \left[x_t x_t^T + \lambda_{t-1} x_{t-1} x_{t-1}^T + \dots + \lambda_{t-1} \dots \lambda_1 x_1 x_1^T + \lambda_1 \dots \lambda_t P_0^{-1} \right]^{-1} \quad (3.47)$$

y P_t^* no depende de λ_t . Para λ_t creciente ($0 < \lambda_t \leq 1$); resulta ρ_p monótona creciente

$$0 < \rho_p \leq \frac{1}{1 + x_{t+1}^T P_t^* x_{t+1}} \quad (3.48)$$

En resumen, aún para el caso de la respuesta libre, no se puede asegurar la reducción (en norma) del error de estimación, en cada paso.

Además, por las razones expuestas no se puede elegir a un paso, con base solamente en el error de estimación e_{t+1} , una sucesión óptima (en el sentido de reducir a cada paso $\|\tilde{\theta}\|$) de factores de olvido λ_t .

Si se tratara de la parte del error de estimación debida a las perturbaciones y a cambios en el parámetro, el problema se complica aún más, pues e_{t+1} en lugar de tener la expresión (3.45), es:

$$e_{t+1} = x_{t+1}^T (\tilde{\theta}_t + \Delta\theta_{t+1}) + e_{t+1} \quad (3.49)$$

A pesar de no poder asegurar la contracción de $|\tilde{\theta}_0^1|$ a un paso veremos que bajo ciertas hipótesis sobre la clase de vectores x_{t+1} , junto con $\lambda_t < 1$ se asegura la convergencia a largo plazo por la vía de garantizar que la contracción se realice en una sucesión de direcciones que barran todo el espacio en el sentido

de no dejar zonas "privilegiadas" sin contraer. Veremos esto en seguida.

3.3.3 Comportamiento de la matriz de covarianzas. Excitación persistente

Definición 1. Diremos que la sucesión de vectores de observación aleatorios: x_{t+1} es "persistentemente excitante" de orden p , si existen reales m y M , $0 < m \leq M$, tales que para cualquier $j(j \geq 0)$ entero, se cumpla:

$$mI \leq \sum_{i=j+1}^{j+p} x_i x_i^T \leq MI \quad \text{c.s. (casi seguramente) (3.50)}$$

Definición 2. Diremos que p_t es una "matriz de excitación persistente" si se cumple que existen ℓ_m y ℓ_M , reales, $0 < \ell_m \leq \ell_M$ tales que:

$$0 < \ell_m I \leq P_t \leq \ell_M I \quad \text{c.s. } \forall t \quad (3.51)$$

Es decir que acotan P_t uniformemente en t .

a). Resultados generales

Proposición 1. Si x_t es una sucesión persistentemente excitante, y λ_t está acotado por λ_m y λ_M :

$$0 < \lambda_m \leq \lambda_t < \lambda_M < 1 \quad \forall_t \quad (3.52)$$

entonces P_t es una matriz de excitación persistente, cumpliéndose:

$$i) \quad 0 < \lambda_m I \leq P_t \leq \lambda_M I \quad \text{c.s.} \quad (3.53)$$

$$\text{donde} \quad \begin{cases} \lambda_m = (\lambda_M^p \lambda_{\text{máx}} P_0^{-1} + \frac{M \lambda_M}{1 - \lambda_M^p})^{-1} \end{cases} \quad (3.54)$$

$$\begin{cases} \lambda_M = \frac{1}{m \lambda_m^p} \end{cases} \quad (3.55)$$

ii) Además, $\forall_t > 0$

$$\left\{ \begin{aligned} & (\lambda_M^t \lambda_{\text{máx}} P_0^{-1} + M \lambda_M \frac{1 - \lambda_M^{(r+1)p}}{1 - \lambda_M^p})^{-1} I \leq P_t \quad \text{c.s.} \quad (3.56a) \\ & \frac{1}{m \lambda_m^p} \frac{1 - \lambda_m^p}{1 - \lambda_m^{rp}} I \geq (\lambda_m^t \lambda_{\text{mín}} P_0^{-1} + m \lambda_m^p \frac{1 - \lambda_m^{rp}}{1 - \lambda_m^p})^{-1} I \geq P_t \quad \text{c.s.} \quad (3.56b) \end{aligned} \right.$$

donde $\begin{cases} r = \xi\{t/p\} \quad , \quad \xi = \text{parte entera.} \\ \lambda_{\text{mín}} P_0^{-1} \text{ y } \lambda_{\text{máx}} P_0^{-1} \text{ son los autovalores mínimo} \\ \text{y máximo de } P_0^{-1} \end{cases}$

La demostración se detalló en el apéndice C.

Comentarios:

Las cotas para los autovalores de P_t dadas por (3.56)

son estrechas, es decir que no son cotas "excesivamente sobradas" como veremos en ejemplos simulados poco más adelante.

Las cotas dadas por (3.54) y (3.55) son bastante más holgadas, pero son uniformes en el tiempo.

La condición de que x_t sea una sucesión persistentemente excitante evita el mal condicionamiento numérico del problema, y asegura convergencia como veremos enseguida.

Como el algoritmo actualiza la identificación a través del error de predicción a un paso ϵ_{t+1} (esto vale para las tres componentes en que se expandió ϵ_{t+1}), ver ec (3.29), y como $\epsilon_{t+1} = x_{t+1}^T (\tilde{\theta}_t + \Delta\theta_{t+1}) + e_{t+1}$, se puede concluir que:

- i) El error de predicción consta de 2 partes, una que corresponde a la "observación" del error de estimación $\tilde{\theta}_t$ según la dirección $x_{t+1} : x_{t+1}^T \tilde{\theta}_t$, y otra que corresponde a las perturbaciones y modificación del parámetro: $x_{t+1}^T \Delta\theta_{t+1} + e_{t+1}$
- ii) La condición de excitación persistente asegura la "observación" de $\tilde{\theta}_t$ en todas las direcciones en el sentido de que como $\sum_{i=j+1}^{j+p} x_i x_i^T$ es no singular, cada p pasos no que

dan direcciones de R^p
 en las que $\tilde{\theta}_t$ no pueda
 ser "observado" a trav
 vés de $x_{t+1}^T \tilde{\theta}_t$.

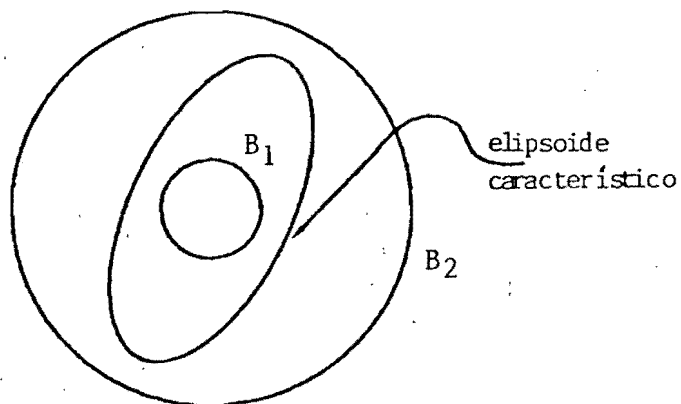


Fig. 3.10

Más aún, como los elipsoides característicos de $\sum_{i=t-p+1}^t x_i x_i^T$ y de P_t (en virtud de la proposición 1) quedan limitados entre sen das bolas de R^p , se asegura a través de los radios de estas bolas mayor uniformidad en la observación mencionada, en el sentido de que se limita el peso relativo que tiene cada di rección frente a las otras.

b) Comportamiento asintótico de P_t ($t \rightarrow \infty, \lambda_M < 1$)

La desigualdad (3.56) rige para todo instante $t > 0$. Es útil en el estudio del transitorio inicial. Pero una vez desaparecido éste, la contribución del término en $P_0^{-1} \rightarrow 0$.

Si el proceso de identificación está llevándose a cabo desde mucho tiempo atrás, puede suponerse que se aproxima como si se hubiera iniciado en $t_0 = -\infty$, o bien en $t_0 = 0$ y $t \rightarrow \infty$.



DEFI

En tal caso, si $\lambda_M < 1$, entonces asintóticamente (3.56) se transforma en:

$$\left(\frac{1 - \lambda_M^p}{M \lambda_M}\right) I \leq P_t \leq \left(\frac{1 - \lambda_m^p}{m \lambda_m^p}\right) I \quad (t \rightarrow \infty) \quad (3.57)$$

c) Comportamiento de P_t , si $\lambda \rightarrow 1$, $t \rightarrow \infty$

Una prueba importante para verificar si las cotas de (3.57) son buenas, es estudiar la consistencia de las conclusiones que de ella se derivan, con los resultados ampliamente conocidos de mínimos cuadrados ordinarios ($\lambda_t = 1 - \psi_t$).

Para ello consideremos una sucesión de experimentos con las mismas señales, x_t , y_t , pero cambiando las sucesiones de factores de olvido, de manera que $\lambda_m \rightarrow 1^-$, $\lambda_M \rightarrow 1^-$.

Ya no se puede seguir el mismo razonamiento de b) en el sentido de considerar $1 - \lambda_m^{rp} \rightarrow 1$ para $r \rightarrow \infty$.

Sin embargo puede demostrarse (por L'Hôpital ó desarrollando en series) que las cotas de desigualdad (3.56) son convergentes a cero ambas, por lo que $P_t \rightarrow 0$ para $\lambda_m \rightarrow 1^-$, $\lambda_M \rightarrow 1^-$, y $t \rightarrow \infty$.

Esto es totalmente consistente con el resultado de mínimos

cuadrados ordinarios (ver p. ej. [42] [43]), en que la matriz de covarianzas converge a la matriz nula si $t \rightarrow \infty$.

Esta convergencia de la matriz de covarianzas es crucial para la convergencia del algoritmo de mínimos cuadrados ordinarios en el caso de sistemas invariantes.

La consistencia de las cotas de la desigualdad (3.56) es importante. Se han encontrado otras cotas (ver p.ej. Lozano [39]), que tienen el problema que para $\lambda_t \rightarrow 1$ no garantizan la convergencia de la matriz de covarianzas y no dan lugar a resultados consistentes con los resultados clásicos. Esto es sumamente inconveniente puesto que los valores utilizados usualmente para el factor de olvido son próximos a la unidad.

d) Incidencia de λ_t en la adaptabilidad

Para asegurar la adaptabilidad del algoritmo de identificación, es necesario evitar que $P_t \rightarrow 0$ si $t \rightarrow \infty$, como ocurre con mínimos cuadrados ordinarios. En efecto: de (3.29) se desprende que si $P_t \rightarrow 0$, el algoritmo pierde su adaptabilidad puesto que se actualiza el parámetro según la dirección $P_{t+1} x_{t+1}$ (colineal con $P_t x_{t+1}$).

El mantener un factor de olvido acotado por $\lambda_M < 1$, asegura de acuerdo a (3.57) que P_t tenga esta propiedad, es decir que ase

gura que P_t tenga una cota inferior para su autovalor mínimo, asegurando que P_t no converge a la matriz nula.

A este efecto usualmente se le denomina "mantener despierto" al estimador. Naturalmente, cuanto más bajo sea el factor de olvido mayor será la cota inferior de P_t , de acuerdo a la desigualdad (3.57).

e) Simulaciones

En c) vimos una verificación de la "bondad" de las desigualdades 3.56 y 3.57. Para mostrar más claramente qué tan buenas son estas desigualdades, hemos recurrido también a simulaciones.

Como ejemplo veremos la simulación que corresponde al sistema de primer orden:

$$y_t = 0.9 y_{t-1} + u_t \quad (3.58)$$

Se excitó el sistema con la entrada periódica 1, -0.8, 1, -0.8, ... y se identificó el sistema con $P_0 = 1000 \times I$ y factor de olvido $\lambda = 0.96$, cte.

Con la entrada elegida se aseguró

$$1.27662 \leq \sum_{i=t+1}^{t+2} x_i x_i^T \leq 2.812132 \quad (3.59)$$

luego

$$\begin{cases} m = 1.27662 \\ M = 2.812132 \end{cases}$$

La figura 3.11.a muestra la evolución de los autovalores de P_t para $1 \leq t \leq 300$, contra los extremos λ_{\min} , λ_{\max} de la desigualdad 3.56.

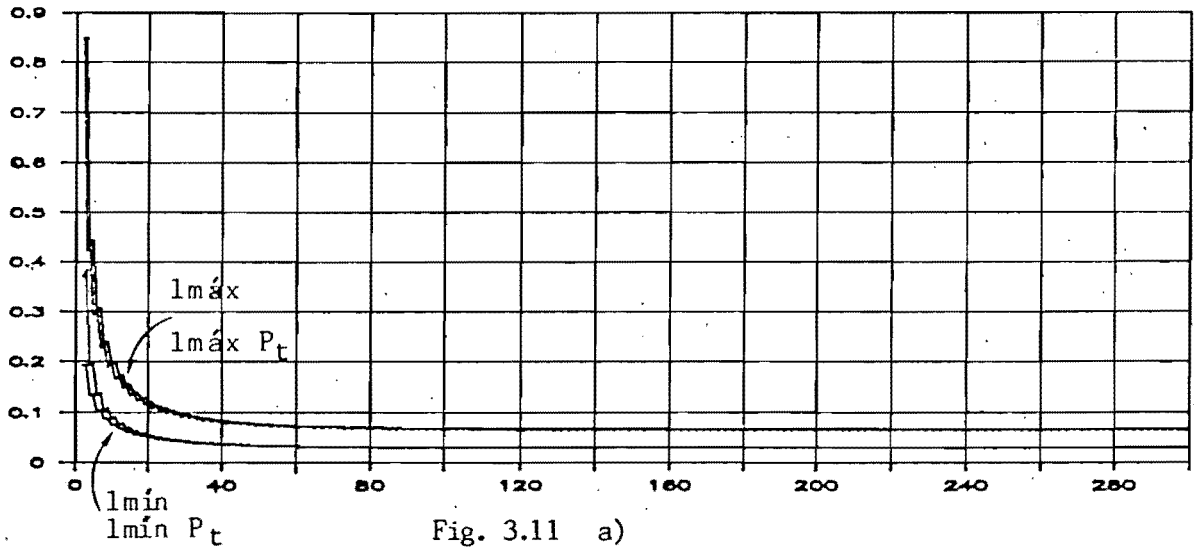
La figura 3.11.b muestra el detalle para $200 \leq t \leq 300$.

En la figura no puede distinguirse separación entre los valores de los autovalores y los extremos alcanzables.

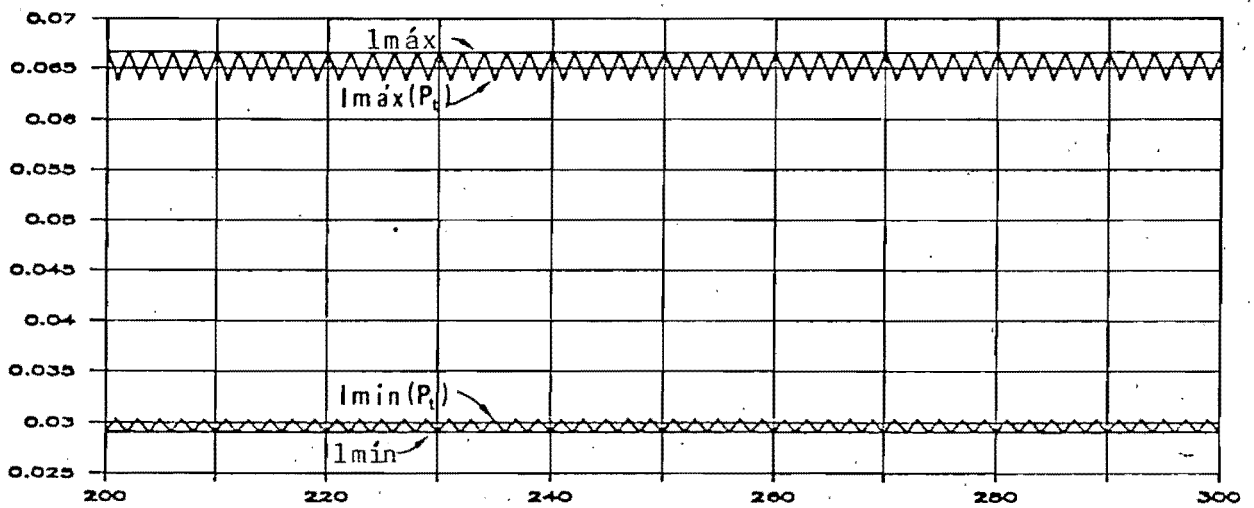
La tabla 3.II muestra los valores de λ_{\min} y λ_{\max} (extremos según 3.56) y los autovalores de $\lambda_{\min} P_t$ y $\lambda_{\max} P_t$ para $1 \leq t \leq 24$ y $295 \leq t \leq 300$.

Se aprecia que las cotas teóricas de los autovalores, dadas por la desigualdad (3.56) se alcanzan, dentro de la precisión del programa LOTUS empleado para el cálculo.

En efecto, en $1 \leq t \leq 24$ la cota superior se alcanza cada dos instantes de muestreo, dentro de un error que varía de 0 a 4×10^{-6} .



Autovalores de $P(t)$ y extremos



lmin(P)	lmax(P)	olvido	lmin	lMax	tiempo t
0.815070	1041.666	0.96	0.370281	1041.666	1
0.370296	0.849281	0.96	0.192731	0.849288	2
0.385148	0.425095	0.96	0.192731	0.849288	3
0.192736	0.442145	0.96	0.133664	0.442149	4
0.200721	0.294461	0.96	0.133664	0.442149	5
0.133666	0.306661	0.96	0.104226	0.306664	6
0.139218	0.229594	0.96	0.104226	0.306664	7
0.104227	0.239130	0.96	0.086640	0.239132	8
0.108560	0.190852	0.96	0.086640	0.239132	9
0.086641	0.198787	0.96	0.074980	0.198789	10
0.090245	0.165168	0.96	0.074980	0.198789	11
0.074981	0.172038	0.96	0.066707	0.172039	12
0.078101	0.146944	0.96	0.066707	0.172039	13
0.066708	0.153057	0.96	0.060550	0.153058	14
0.069484	0.133380	0.96	0.060550	0.153058	15
0.060551	0.138931	0.96	0.055803	0.138932	16
0.063071	0.122924	0.96	0.055803	0.138932	17
0.055804	0.128040	0.96	0.052043	0.128041	18
0.058126	0.114641	0.96	0.052043	0.128041	19
0.052044	0.119413	0.96	0.049000	0.119414	20
0.054210	0.107938	0.96	0.049000	0.119414	21
0.049001	0.112431	0.96	0.046495	0.112432	22
0.051040	0.102419	0.96	0.046495	0.112432	23
0.046495	0.106683	0.96	0.044402	0.106684	24

0.030250	0.063972	0.96	0.029041	0.066636	295
0.029041	0.066636	0.96	0.029040	0.066636	296
0.030250	0.063971	0.96	0.029040	0.066636	297
0.029041	0.066636	0.96	0.029040	0.066636	298
0.030250	0.063971	0.96	0.029040	0.066636	299
0.029041	0.066636	0.96	0.029040	0.066636	300

TABLA 3.II.

La cota inferior se alcanza con un instante de retraso, y en los últimos instantes ($295 \leq t \leq 300$), se alcanza cada dos instantes de muestreo dentro de un error de 4×10^{-6} .

Más aún, de acuerdo a la desigualdad (3.57) se esperan cotas asintóticas $\lambda_{\min} = 0.290404$ y $\lambda_{\max} = 0.0666364$, alcanzadas con una precisión de $1/10^6$ para los últimos puntos tabulados.

En resumen, las cotas encontradas para los autovalores de P_t son cotas finas, como lo han mostrado las observaciones de la sección 3.3.3c y el ejemplo simulado, que es bastante elocuente.

3.3.4 Evolución de la componente debida al error inicial: resultados generales

Basados en los resultados anteriores, podemos ahora estudiar la evolución de los tres componentes del error de estimación $\tilde{\theta}_t$.

Comenzaremos con el "error inicial", es decir la contribución de las condiciones iniciales al error de estimación: $\tilde{\theta}_t^1$.

La siguiente proposición permite analizar $\tilde{\theta}_t^1$.

Proposición 2

El término $\tilde{\theta}_t^1$, la parte del error de estimación debido a la condición inicial $\tilde{\theta}_0^1$, satisface:

$$i) \quad \tilde{\theta}_t^1 = \lambda_1 \lambda_2 \dots \lambda_t P_t P_0^{-1} \tilde{\theta}_0^1 \quad (3.60)$$

$$ii) \quad \|\tilde{\theta}_t^1\| \leq \left(\prod_{i=1}^t \lambda_i \right) \frac{\lambda_{\max} P_t}{\lambda_{\min} P_0} \|\tilde{\theta}_0^1\| \quad (3.61)$$

$$iii) \quad \|\tilde{\theta}_t^1\| \geq \left(\prod_{i=1}^t \lambda_i \right) \frac{\lambda_{\min} P_t}{\lambda_{\max} P_0} \|\tilde{\theta}_0^1\| \quad (3.62)$$

La demostración puede encontrarse en el apéndice C.

Debe notarse que no se ha hecho ninguna hipótesis sobre la clase de vectores de observación x_t , ni sobre los valores de la sucesión de factores de olvido. Sin embargo para establecer la convergencia de $\tilde{\theta}_t^1$ a cero, es necesario hacer ciertas hipótesis adicionales:

Proposición 3 (Convergencia de $\tilde{\theta}_t^1$ ver Lozano [39])

Si además la sucesión x_t es persistentemente excitante y los factores de olvido están acotados por:

$$0 < \lambda_m \leq \lambda_t \leq \lambda_M < 1 \quad \forall t \quad (3.63)$$

entonces se cumple:

$$i) \quad \text{La sucesión } \|\tilde{\theta}_t^1\|_{P_t^{-1}}^2 \triangleq \tilde{\theta}_t^{1T} P_t^{-1} \tilde{\theta}_t^1$$

es positiva, decreciente, y además:

$$\lim_{t \rightarrow \infty} \tilde{\theta}_t^{1T} P_t^{-1} \tilde{\theta}_t^1 = 0 \quad \text{c.s.} \quad (3.64)$$

$$ii) \quad \lim_{t \rightarrow \infty} \|\tilde{\theta}_t^1\| = 0 \quad \text{c.s.} \quad (3.65)$$

La demostración se encuentra en el apéndice C.

Los resultados anteriores establecen condiciones de convergencia a cero del error de estimación debido a condiciones ini-ciales.

El resultado ii) de la proposición 3 establece convergencia salvo en un conjunto de realizaciones de medida nula.

Cabe hacer notar que los resultados ii) y iii) de la propo-sición 2, junto con la hipótesis de excitación persistente ase-guran una convergencia de carácter exponencial.

Para el caso que nos ocupa ahora esto no es muy importante, puesto que alcanza con saber que la contribución de las condi

ciones iniciales tiende a cero, y luego de un período transitorio, dejan de influir en el error de estimación. Por el contrario, veremos en la siguiente sección que el carácter exponencial de dicha convergencia sí importa si hay cambios esporádicos en el parámetro θ_t del sistema.

3.3.5 Evolución de la componente del error debido a variación en el parámetro

Estudiaremos ahora la componente del error de estimación $\tilde{\theta}_t^2$ debido a cambios en el parámetro θ_t del sistema que genera los datos.

Como ya se discutió en el capítulo 2, no existe ninguna forma de abordar el problema de identificación de un sistema arbitrariamente variable en el tiempo. El caso que más nos interesa estudiar es cuando existen variaciones abruptas en el parámetro, lo suficientemente distanciadas entre sí como para permitir cierta adaptación de la identificación entre salto y salto.

Suponemos además que el proceso de identificación se inicia en $-\infty$, por lo que vale el comportamiento asintótico de P_t :
(3.57).

Estudiaremos dos casos:

a) Variaciones esporádicas del parámetro

b) Variaciones arbitrarias del parámetro

a) Variaciones esporádicas del parámetro

Consideramos la segunda ecuación del sistema (3.34), que corresponde a la evolución de $\tilde{\theta}_t^2$. Donde $\Delta\theta_t = 0$, excepto en t_{o+1} , en que $\Delta\theta_{t_{o+1}} = \Delta\theta$, un cambio aislado del parámetro.

El problema se reduce al caso considerado en la sección anterior, de evolución de la condición inicial $\Delta\theta$, a partir de t_o .

Entonces, de acuerdo a (3.61) y (3.62):

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_{o+\tau}}^2\| \leq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{\lambda_{\text{máx}}^P t_{o+\tau}}{\lambda_{\text{mín}}^P t_o} \|\Delta\theta\| \quad \text{c.s.} \\ \|\tilde{\theta}_{t_{o+\tau}}^2\| \geq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{\lambda_{\text{mín}}^P t_{o+\tau}}{\lambda_{\text{máx}}^P t_o} \|\Delta\theta\| \quad \text{c.s.} \end{array} \right. \quad (3.66)$$

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_{o+\tau}}^2\| \leq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{\lambda_{\text{mín}}^P t_{o+\tau}}{\lambda_{\text{máx}}^P t_o} \|\Delta\theta\| \quad \text{c.s.} \\ \|\tilde{\theta}_{t_{o+\tau}}^2\| \geq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{\lambda_{\text{máx}}^P t_{o+\tau}}{\lambda_{\text{mín}}^P t_o} \|\Delta\theta\| \quad \text{c.s.} \end{array} \right. \quad (3.67)$$

luego, por (3.57):

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_{o+\tau}}^2\| \leq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{M}{m} \frac{1 - \lambda_m^P}{\lambda_m^P} \frac{\lambda_M}{1 - \lambda_M^P} \|\Delta\theta\| \quad \text{c.s.} \\ \|\tilde{\theta}_{t_{o+\tau}}^2\| \geq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{m}{M} \frac{\lambda_m^P}{1 - \lambda_m^P} \frac{1 - \lambda_M^P}{\lambda_M} \|\Delta\theta\| \quad \text{c.s.} \end{array} \right. \quad (3.68)$$

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_{o+\tau}}^2\| \leq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{m}{M} \frac{\lambda_m^P}{1 - \lambda_m^P} \frac{1 - \lambda_M^P}{\lambda_M} \|\Delta\theta\| \quad \text{c.s.} \\ \|\tilde{\theta}_{t_{o+\tau}}^2\| \geq \left(\prod_{i=t_{o+1}}^{t_{o+\tau}} \lambda_i \right) \frac{M}{m} \frac{1 - \lambda_m^P}{\lambda_m^P} \frac{\lambda_M}{1 - \lambda_M^P} \|\Delta\theta\| \quad \text{c.s.} \end{array} \right. \quad (3.69)$$

Las ecuaciones (3.68) y (3.69) permiten sacar varias conclusiones.

- i) El error de estimación $\tilde{\theta}_{t_0+\tau}^2$ debido a un cambio aislado del parámetro en el tiempo t_0 , converge exponencialmente a cero puesto, que como $\lambda_i \leq \lambda_M \psi_i$, y $\lambda_M < 1$ entonces:

$$\|\tilde{\theta}_{t_0+\tau}^2\| \leq (\lambda_M^\tau) \left[\frac{M}{m} \frac{1 - \lambda_m^p}{\lambda_m^p} \frac{\lambda_M}{1 - \lambda_M^p} \|\Delta\theta\| \right] \quad \text{c.s.} \quad (3.70)$$

Como el término entre corchetes es constante, entonces $\tilde{\theta}_{t_0+\tau}^2$ converge a cero exponencialmente conforme $\tau \rightarrow \infty$.

- ii) El error $\tilde{\theta}_{t_0+\tau}^2$ evoluciona dentro de una franja que se estrecha y tiende exponencialmente a cero conforme $\tau \rightarrow \infty$.

Esta franja está definida por constantes que dependen de m , M , λ_m , λ_M .

- iii) Para ejemplificarlo en forma sencilla veamos qué sucede si $\lambda_m = \lambda_M$ es decir factor de olvido constante λ . En tal caso:

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_0+\tau}^2\| \leq \lambda^\tau \left(\frac{M}{m} \lambda^{1-p} \|\Delta\theta\| \right) \quad \text{c.s.} \quad (3.71) \\ \|\tilde{\theta}_{t_0+\tau}^2\| \geq \lambda^\tau \left(\frac{m}{M} \lambda^{p-1} \|\Delta\theta\| \right) \quad \text{c.s.} \quad (3.72) \end{array} \right.$$

El decaimiento exponencial del error de estimación le dá sentido a las definiciones de "ventana efectiva" y "tiempo de subida" dadas en la sección 3.2.1, así como explica el comportamiento de las simulaciones de la sección 3.2.2.

En la figura (3.12) se muestra la evolución de $\tilde{\theta}_{t+\tau}^2$ (normalizada) dentro de la franja definida por las ecuaciones (3.71) y (3.72).

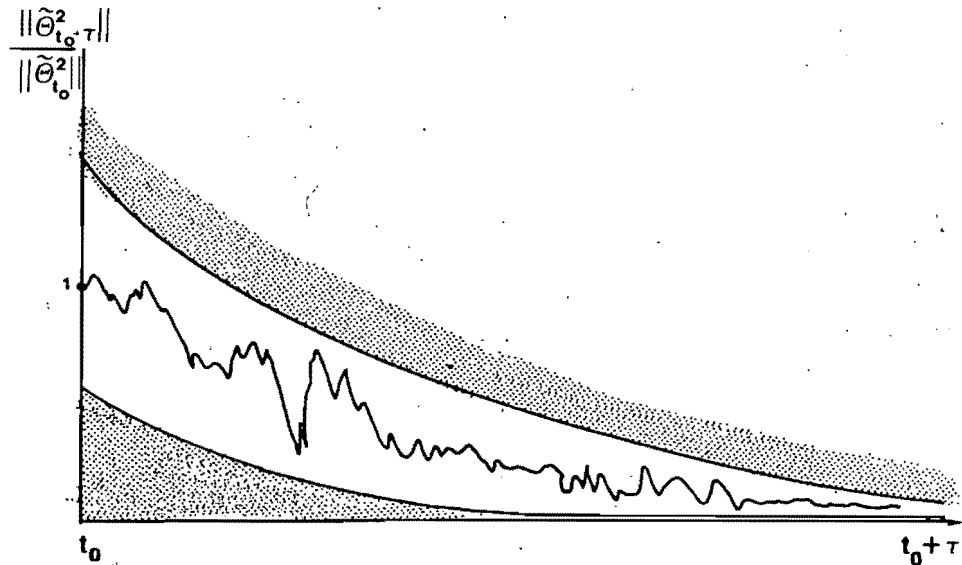


Fig. (3.12).

Sobre esto vale la pena detenerse un poco. Debemos notar que la amplitud de dicha franja está directamente vinculada al número de condición M/m que define en (3.50) la excitación persistente.

Si el número de condición es muy alto, la franja es muy ancha

y pueden esperarse transitorios fuertes, o fenómenos de "bursting" reportados frecuentemente en la literatura.

Estos fenómenos han sido asociados a crecimientos grandes ("blow-up") de la matriz de covarianza, (ver p.ej: Aström/Witternmark [46], o a errores numéricos de redondeo..." asociados a matrices de covarianza muy grandes ó rápidamente crecientes", como afirma Ydstie [41], por lo que este autor sugiere el uso de algoritmos más robustos numericamente.

Si M/m es alto, el número de condición de la matriz P_t puede ser alto también (ver ec. (3.57)), pero esto no es forzoso. Aún más, pueden ocurrir fuertes transitorios en $\tilde{\theta}_t$, aunque el número de condición instantáneo de P_t no crezca mucho, como se verá poco más adelante.

Las ecuaciones (3.71) y (3.72) señalan algo claro: delimitan una franja para la evolución de $\|\tilde{\theta}_{t_0+\tau}^2\|$. Los extremos (bordes), están fijados por la caracterización de x_t (3.50). Pero la realización particular determina la curva que $\|\tilde{\theta}_{t_0+\tau}^2\|$ habrá de describir.

Es decir que aún cuando la sucesión de vectores de observación x_t satisfaga la misma condición de excitación persistente (3.50), se podrían encontrar realizaciones que provocaran grandes oscilaciones dentro de esa franja, y otras que no.

Por eso, un mal condicionamiento ($\frac{M}{m} \gg 1$) permite la posibilidad de un fenómeno transitorio fuerte pero no lo obliga. A es to se debe que en algunos experimentos exista un problema de oscilación en $\|\tilde{\theta}_{t_{0+\tau}}^2\|$, y en otros no.

Aquí es donde la observación de Ydstie tiene más sentido. En efecto, la oscilación puede desencadenarse como resultado de un cambio en el parámetro, según lo acabamos de discutir. Pero un error de redondeo puede considerarse como una condición inicial a partir de la cual evoluciona $\tilde{\theta}_t^1$, que tiene un comportamiento análogo, según las ecuaciones (3.61) y (3.62).

Puede suceder entonces, que errores de redondeo, acarreo, o cualquier otra perturbación numérica, desencadene un fenómeno transitorio fuerte bajo un número de condición $\frac{M}{m}$ alto.

Observaciones

- i) Las cotas encontradas son independientes de la velocidad con que cambien los autovalores y los autovectores de P_t , puesto que simplemente toman en cuenta sólo los valores

extremos posibles.

- ii) Debe notarse que para alcanzar la cota superior, es necesario (ver ec. 3.60), que el salto en el parámetro se produzca en la dirección del autovector asociado al autovalor mínimo de P_{t_0} .

Además el autovalor mencionado debe alcanzar el valor inferior dado por las cotas de (3.57).

Cumplida esta condición, para que alcance la banda superior en un instante $t_0 + \tau$, también es necesario que $P_{t_0}^{-1} \Delta\theta$ (que es colineal con el salto en el parámetro: $\Delta\theta$, si se cumple la condición anterior) esté en la dirección del autovector asociado al máximo autovalor de $P_{t_0+\tau}$, y dicho autovalor debe alcanzar la cota (3.57).

Sobra decir que es bastante difícil que se cumplan ambas condiciones, pero en principio, es posible.

- iii) Como corolario de la observación anterior se desprende que si bien las cotas dadas por (3.68) y (3.69) no son posibles de alcanzar en un número bajo de instantes de muestreo, puesto que no es posible que P_t cambie abruptamente una misma dirección, de estar asociada al autovalor mínimo, a estar asociada al autovalor máximo.

Se deduce entonces que dichas cotas no son buenas (están holgadas) para los primeros instantes luego del salto en el parámetro.

- iv) De la observación ii) también se desprende que es posible tener situaciones donde exista un pésimo mal condicionamiento numérico, y sin embargo no haya "bursting". Más aún, que el comportamiento frente a un salto abrupto en el parámetro sea una rápida convergencia, si p.ej: el salto es dado en t_0 en la dirección asociada al autovalor máximo de P_{t_0} , y esa misma dirección tiende a ser la asociada al autovalor mínimo de $P_{t_0+\tau}$ con $\tau > 0$.
- v) De la observación ii) se desprende que no es necesario que instantáneamente P_t tenga un número de condición alto para que ocurra un transitorio fuerte en $\|\tilde{\theta}_{t_0+\tau}^2\|$, ya que importa $\lambda_{\text{máx}} P_{t_0+\tau} / \lambda_{\text{mín}} P_{t_0}$ y no su valor instantáneo.
- vi) Aunque toda la derivación de resultados se hizo para el algoritmo de mínimos cuadrados con olvido variable, es posible extender dichos resultados para una amplia clase de algoritmos de identificación de estructura similar, e incluso a algoritmos de control autoadaptivo.
- vii) En cuanto a los esquemas conocidos de elección del factor

de olvido λ_t , por lo general se hace hincapié en comprobar si son robustos o no al problema de bursting.

A la luz de lo discutido hasta ahora, vale la pena observar que:

- a) El algoritmo de traza constante, aunque al "mantener despierto" al identificador, acota el numerador de (3.66) ya que acota superiormente $t_r\{P_{t_0+\tau}\}$, no acota inferiormente el denominador, y no ejerce por lo tanto ninguna acción para evitar el fenómeno de "bursting".
- b) Algoritmos de elección de λ_t en función de errores de predicción pueden tener problemas serios si no efectúan al mismo tiempo alguna clase de protección contra la pérdida de la persistencia de la excitación.

Por ejemplo, el algoritmo de cantidad de Información constante (Fortescue et.al. '21), si se produce un error de predicción alto, baja el factor de olvido. Esto provoca, (ver ec. 3.56), que las cotas de los autovalores se abran abruptamente si $\frac{M}{m} \gg 1$, y de acuerdo a (3.66) la cota superior para el error de estimación crece muchísimo.

viii) La última observación vale para todos los esquemas de elección de λ_t , que baje λ_t (para ganar adaptabilidad) frente a un salto en el parámetro, si ese salto se produce cuando se pierde la persistencia de la excitación.

Por eso es necesario mantener un mecanismo paralelo que evite que se esté ejecutando la baja de λ_t frente a una situación de crecimiento del número de condición M/m más allá de márgenes preestablecidos.

Simulaciones

- 1) Para ejemplificar el comportamiento de $\|\tilde{\theta}_t^2\|$ frente a un salto en el parámetro se presentan dos simulaciones. La primer simulación trata del sistema

$$y_t = 0.9 y_{t-1} + 0.72 u_t \quad 1 \leq t \leq 300$$

$$y_t = 0.4 y_{t-1} + 0.85 u_t \quad 301 \leq t \leq 500$$

Se eligió una entrada que mantuviera $m = 1.310065$,
 $M = 2.456146 \psi_t$.

Se ejecutó la identificación con f.o. cte = 0.96, y en la figura (3.13 a) y b) se presenta la evolución del error $\|\tilde{\theta}_t^2\|$, contra los extremos dados por las ecs.

(3.71) y (3.72).

La figura (3.13 (b)) presenta el detalle para $350 \leq t \leq 450$.

Puede verse como el error tiende a alcanzar la cota superior, aunque esto no es abrupto sino que es necesario una transición como se mencionó en la observación ii).

2) La figura (3.14) muestra un caso similar para el sistema:

$$y_t = 0.9 y_{t-1} + u_t \quad 1 \leq t \leq 300$$

$$y_t = 0.7 y_{t-1} + 1.278137 u_t \quad 301 \leq t \leq 500$$

En este caso la evolución del error es próxima a su cota inferior.

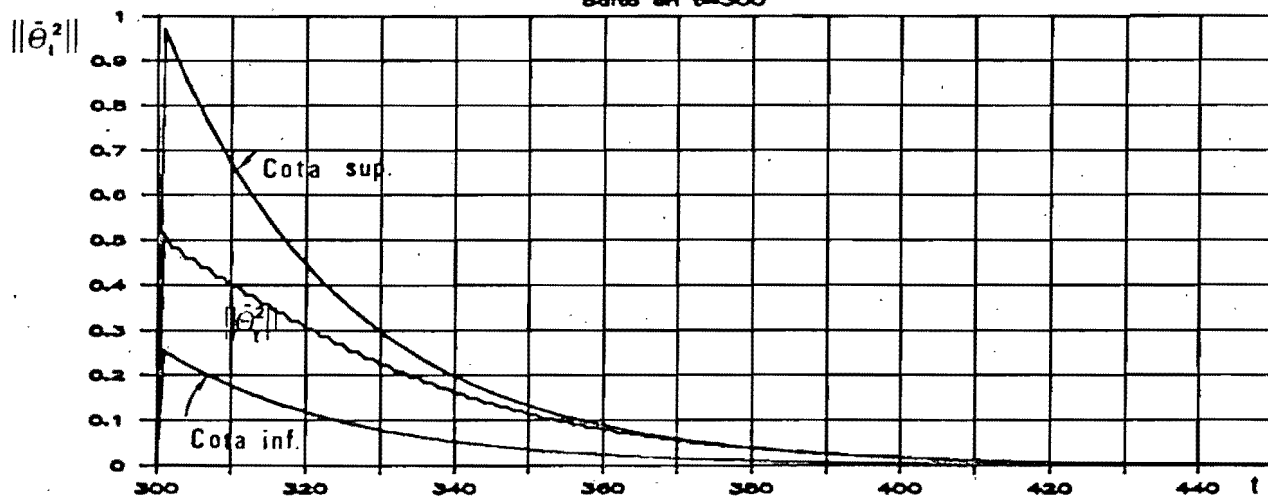
3) Para apreciar la influencia del número de condición se simuló el sistema:

$$y_t = 0.6 y_{t-1} + u_t$$

con $u_t = 1 + (-1)^t \cdot 0.001$

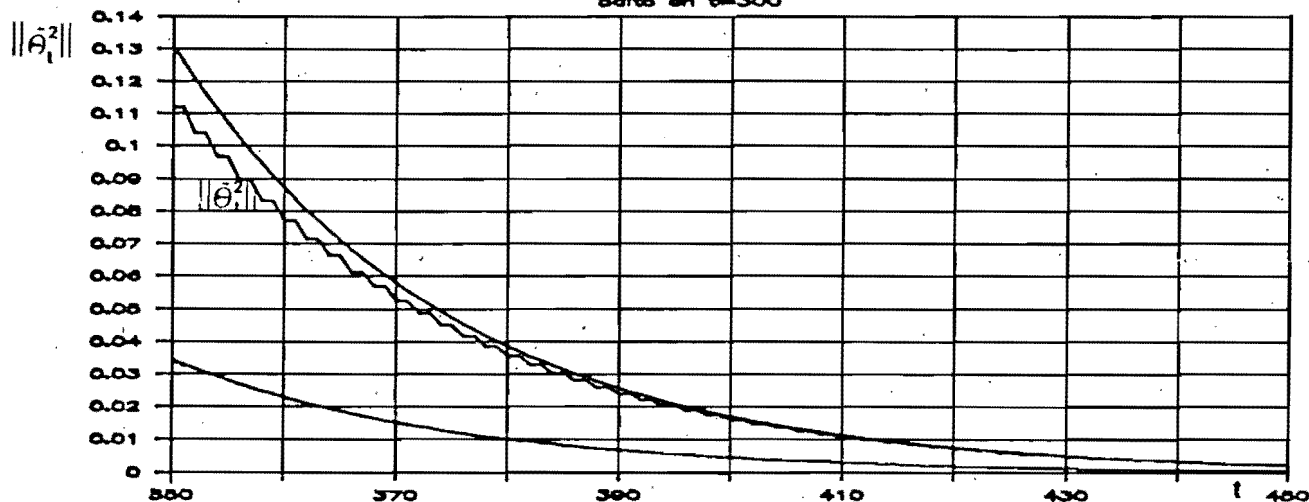
El número de condición M/m es extremadamente malo ($\approx 10^5$).

Error de estimacion
Salto en $t=300$



(a)

Error de estimacion
Salto en $t=300$



(b)

Figura 3.13

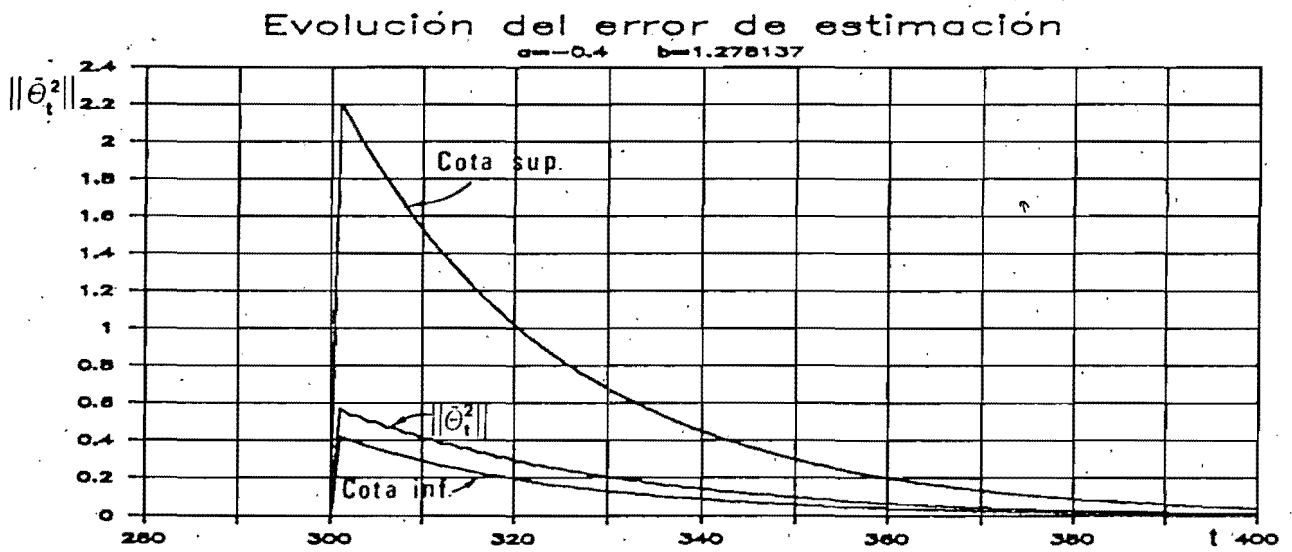


Figura 3.14

Se llevó a cabo un proceso de identificación sobre este sistema, y dos pequeños errores de identificación fueron inducidos en $t = 200$ y 300 , ambos con la misma magnitud $\|\tilde{\theta}_{t_0}^2\| = 0.01$, pero con diferente orientación. El factor de olvido empleado fue $\lambda = 0.96$.

La figura (3.15) muestra la evolución de la norma del error de estimación.

Se puede apreciar que a pesar de estar el sistema en estado estacionario, y por lo tanto P_t oscila ligeramente alrededor de un valor, el mal número de condición puede dar lugar a transitorios fuertes: También se aprecia que el número de condición "habilita" el transitorio pero no es obligatorio, puesto que la respuesta al error de $t = 200$ decae rápidamente. Mientras que la respuesta al error en $t = 300$ tiene amplitudes 10 veces más altas que el error original.

La única diferencia en ambas respuestas estriba en la dirección en la que se indujo el error.

Se repitió el mismo ejemplo simulado, pero ahora con

$$u_t = 1 + 0.001 \times r_t$$

donde r_t es una señal pseudo aleatoria de distribución unifor-

Error de estimacion

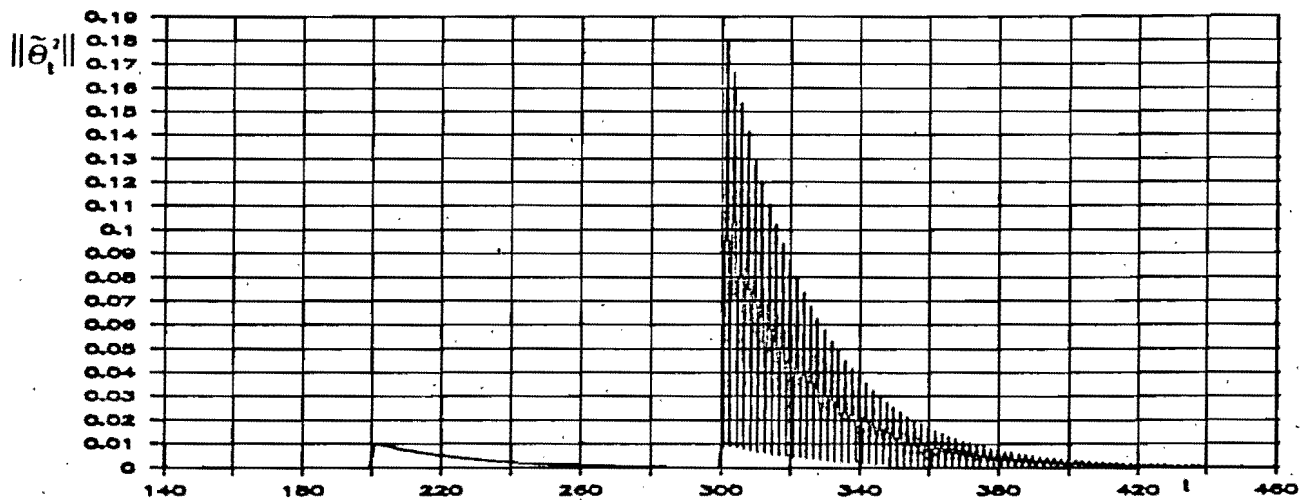


Figura 3.15

Error de estimacion

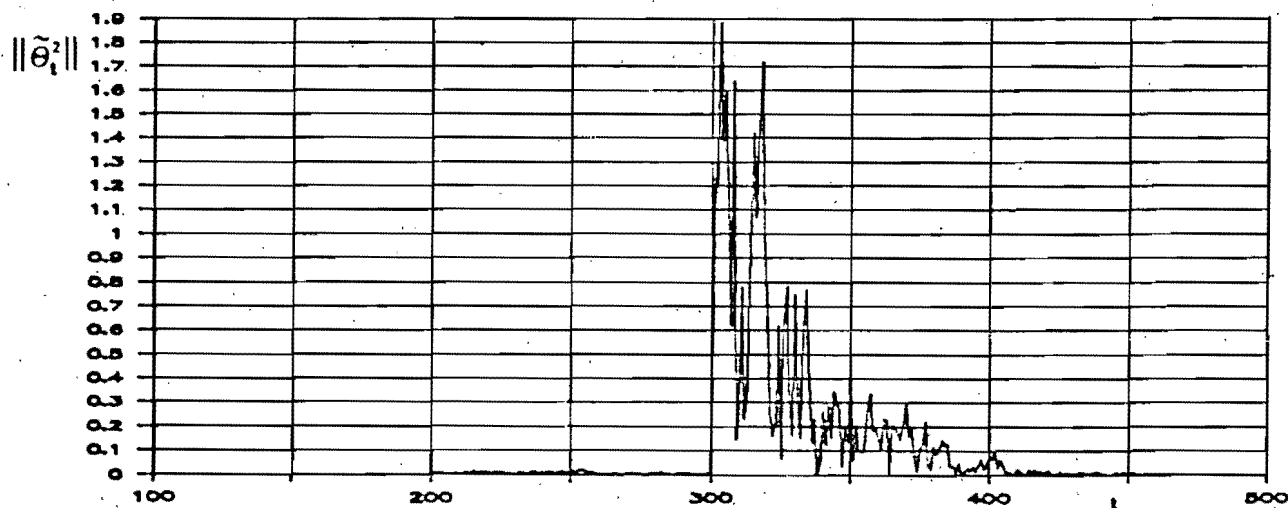


Figura 3.16

me $[-1,1]$.

En $t = 200$ y 300 se indujeron errores $\|\tilde{\theta}_{t_0}^2\| = 0.01$, con diferentes direcciones.

La figura (3.16) muestra la evolución de $\|\tilde{\theta}_t^2\|$, que confirma las observaciones anteriores. Mientras que se responde al primer error con ligeras oscilaciones, al segundo se responde con una excursión 190 veces más grande que el error original.

Si el número de condiciones M/m es muy alto (pérdida de la excitación persistente), ligeros errores de redondeo acumulados pueden dar lugar a un comportamiento muy malo del estimador.

b) Variaciones arbitrarias del parámetro θ_t

Si ahora se permite que $\Delta\theta_{t_0+\tau} \neq 0 \quad \forall \tau > 0$, es obvio que no se puede esperar convergencia puesto que el parámetro del sistema está cambiando permanentemente.

Sin embargo, si la variación del parámetro es acotada se puede demostrar que el error de estimación producido por esta variación también es acotado bajo las mismas hipótesis que el caso anterior.

Proposición 4

Si x_t es persistentemente excitante; el factor de olvido λ_t es tá acotado como antes por λ_m y $\lambda_M < 1$, como en (3.63), y

$\Delta\theta_{t_{o+i}} = 0 \quad \forall_i \leq 0$ y $\Delta\theta_{t_{o+i}}$ está acotada en $0 < i \leq \tau$, entonces:

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_{o+\tau}}^2\| < \frac{M}{m} \frac{1 - \lambda_m^p}{\lambda_m^p} \frac{\lambda_M^2 (1 - \lambda_M^\tau)}{(1 - \lambda_M^p)(1 - \lambda_M)} \max_{1 \leq i \leq \tau} \|\Delta\theta_{t_{o+i}}\| \quad \text{c.s.} \quad (3.73) \end{array} \right.$$

$$\left\{ \begin{array}{l} \|\tilde{\theta}_{t_{o+\tau}}^2\| < \frac{M}{m} \frac{1 - \lambda_m^p}{\lambda_m^p} \frac{\lambda_M^2}{(1 - \lambda_M^p)(1 - \lambda_M)} \max_{1 \leq i \leq \tau} \|\Delta\theta_{t_{o+i}}\| \quad \text{c.s.} \quad (3.74) \end{array} \right.$$

La demostración se encuentra en el apéndice C. La proposición está basada en un resultado análogo de R. Lozano [39], para un algoritmo de identificación normalizada con $\|x_t\| < 1$.

Si el factor de olvido es constante, (3.74) queda como:

$$\|\tilde{\theta}_{t_{o+\tau}}^2\| < \frac{M}{m} \frac{\lambda^{2-p}}{1 - \lambda} \max_{1 \leq i \leq \tau} \|\Delta\theta_{t_{o+i}}\| \quad \text{c.s.} \quad (3.75)$$

También en este caso, el número de condición $\frac{M}{m}$ incide proporcionalmente sobre la cota del error de estimación.

3.3.6 Influencia de las perturbaciones

Lo último que resta por analizar es el término $\tilde{\theta}_t^3$, resultado de la contribución del proceso del ruido e_t al error de estimación. El siguiente resultado acota dicha contribución.

Proposición 5

Si x_t es persistentemente excitante; λ_t está acotado como en (3.63), y e_t es una sucesión de ruido blanco centrado con $E\{e_t^2\} = \sigma^2$, entonces:

$$i) \quad \limsup_{t \rightarrow \infty} E\{\|\tilde{\theta}_t^3\|^2\} \leq \sigma^2 \frac{\lambda_M(1 - \lambda_m^P)}{m \lambda_m^P (1 - \lambda_M)} \quad \text{c.s.} \quad (3.76)$$

ii) Si $\lambda_m = \lambda_M < 1$ (factor de olvido constante)

$$\limsup_{t \rightarrow \infty} E\{\|\tilde{\theta}_t^3\|^2\} \leq \sigma^2 \frac{M}{m^2} \frac{(1 - \lambda^P)^2}{\lambda^{2P} \ln(1/\lambda)} \quad \text{c.s.} \quad (3.77)$$

La demostración se encuentra en el apéndice C.

Observaciones:

La varianza de la parte del error de estimación debida a perturbaciones aleatorias está asintóticamente acotada.

Las cotas encontradas son proporcionales a la varianza

del proceso de perturbaciones.

Si λ_m y $\lambda_M \rightarrow 0$ (ó $\lambda \rightarrow 0$ en el segundo caso), ambas cotas divergen.

La primera cota es útil para valores de λ pequeños.

La segunda cota es útil para valores de λ próximos a la unidad.

Comportamiento de $\|\tilde{\theta}_t^3\|$ para $\lambda \rightarrow 1$

Consideremos la cota de la ec. 3.77:

$$\limsup_{t \rightarrow \infty} E\{\|\tilde{\theta}_t^3\|^2\} \leq C(\lambda) \quad \text{c.s.}$$

donde

$$C(\lambda) = \sigma^2 \frac{M}{m^2} \frac{(1-\lambda^p)^2}{\lambda^{2p} \ln(1/\lambda)}$$

Si se repite el proceso de identificación para una sucesión de factores de olvido $\lambda \rightarrow 1^-$, el comportamiento de la cota $C(\lambda)$ se puede estudiar haciendo $\lambda = 1 - \epsilon$, ϵ pequeño.

$$C(\lambda) = C(1 - \epsilon) = \sigma^2 \frac{M}{m^2} \frac{(1-\lambda^p)^2}{\lambda^{2p} \ln(1/\lambda)}$$

Expandiendo $C(1-\epsilon)$ para $\epsilon \rightarrow 0$:

$$C(1 - \epsilon) \approx \sigma^2 \frac{M}{m^2} \frac{p^2 \epsilon^2}{\epsilon} = \sigma^2 \frac{M}{m^2} p^2 \epsilon \quad (\epsilon \rightarrow 0)$$

Con lo que se demuestra que la varianza de $\|\tilde{\theta}_t^3\|$ tiende a cero para $\lambda \rightarrow 1$ salvo en un conjunto de realizaciones de medida nula. Este resultado es consistente con los resultados de mínimos cuadrados ordinarios ampliamente conocidos. Las cotas análogas obtenidas hasta ahora fallaban precisamente en el caso $\lambda \rightarrow 1$.

se puede concluir que

$$\left\{ \begin{array}{l} C(\lambda) \rightarrow \sigma^2 \frac{M}{m^2} p^2 (1 - \lambda) \\ \text{para } \lambda \rightarrow 1^- \end{array} \right.$$

Comentarios

Llegados a este punto podemos establecer un resultado esperado. Existe en la selección de λ un compromiso entre velocidad de seguimiento de los cambios en el parámetro y precisión debida a la influencia de las perturbaciones.

En efecto: supongamos $\lambda_t = \lambda$ constante. De acuerdo a lo que acabamos de ver, para valores crecientes de λ , el error de es-

timación debido al ruido tiende a disminuir, pero de acuerdo a (3.71), la cota superior de convergencia del error debido a cambios en el parámetro aumenta. Naturalmente lo inverso ocurre para valores de λ decrecientes.

Este compromiso obliga a encontrar esquemas apropiados de selección de λ_t que tomen en cuenta al mismo tiempo la influencia de las perturbaciones y de los cambios en el parámetro del sistema.

3.4 Selección del factor de olvido λ_t .

Como ya se mencionó, existen grandes dificultades para encontrar una forma de determinación del factor de olvido λ_t . En cada instante de muestreo se dispone de información del error de estimación a través del error de predicción

$$e_{t+1} = x_{t+1}^T (\tilde{\theta}_t + \Delta\theta_{t+1}) + e_{t+1} \quad (3.78)$$

Las preguntas básicas que se trata de responder a partir de la información disponible son:

¿Cómo saber si hubo un cambio en el parámetro para forzar una disminución del factor de olvido?

¿Cómo realizar dicha disminución?

De (3.78) y suponiendo que el proceso e_t es independiente de los saltos en el parámetro:

$$E \{ \epsilon_{t+1}^2 / x_{t+1} \} = x_{t+1}^T E \{ (\tilde{\theta}_t^T + \Delta\theta_{t+1}^T)(\tilde{\theta}_t + \Delta\theta_{t+1}) / x_{t+1} \} x_{t+1} + \sigma^2 \quad (3.79)$$

La mayor parte de los métodos de selección del factor de olvido se basan en comparar ϵ_{t+1}^2 con alguna cantidad que estima su esperanza. De (3.79) se puede estimar σ^2 , pero la estimación del primer sumando del lado derecho de (3.79) tropieza con grandes dificultades.

Por lo general, estos métodos de selección recurren a uno de dos esquemas:

- i) Variar λ_t en forma continua como función de ϵ_{t+1} , x_{t+1} , P_t .
- ii) Mantener un nivel constante $\lambda_t = \lambda < 1$ que aborde variaciones lentas del parámetro mientras se corre un "detector" de grandes cambios del parámetro que pone en funcionamiento un mecanismo de adaptación que disminuye el valor de λ_t .

Veremos algunos de estos métodos.

3.4.1 Algoritmo de información constante

Una forma de elegir el factor de olvido fue propuesta por For-
tescue, Kershenbaum e Ydstie [2].

Se define recursivamente la cantidad de información de las ob-
servaciones como

$$I_t = \lambda_{t-1} I_{t-1} + \frac{|1 - x_t^T P_{t-1} x_t|}{1 + x_t^T P_{t-1} x_t} \epsilon_t^2 \quad (3.80)$$

poniendo 3.80 en t+1

$$I_{t+1} = \lambda_t I_t + \frac{|1 - x_{t+1}^T P_t x_{t+1}|}{1 + x_{t+1}^T P_t x_{t+1}} \epsilon_{t+1}^2$$

$$I_{t+1} = \lambda_t I_t + \frac{\epsilon_{t+1}^2}{1 + x_{t+1}^T P_t x_{t+1}} \quad (3.81)$$

Comparando las ecuaciones (3.81) y (3.9), queda claro que con
la simple elección adecuada de I_0 (ver inicialización), la
cantidad de información no es otra cosa que la función de cos-
to evaluada en $\hat{\theta}_{t+1}$.

Es decir:

$$I_{t+1} = \sum_{i=1}^{t+1} \alpha_i (y_i - x_i^T \hat{\theta}_{t+1})^2 \quad (3.82)$$

Lo que se propone es elegir la sucesión λ_t de una manera que la cantidad de información I_{t+1} se mantenga en un valor constante Σ_0 . Lo que se consigue con esto es que si el error de predicción ϵ_{t+1} es muy grande, el valor de λ_t se reduce.

La interpretación es la siguiente: un cambio en el modelo dará mayores errores de predicción, esto reduce λ_t y se "pesan" menos las observaciones pasadas.

Inversamente, si ϵ_{t+1} es pequeño, esto se interpreta como que el modelo actual es bueno, y con $\lambda \rightarrow 1$ se siguen tomando en cuenta los datos con los que fue obtenido (se prolonga la "ventana efectiva").

El valor Σ_0 pasa a ser un parámetro de diseño.

Se elige Σ_0 considerando el caso estacionario. Supongamos que el sistema considerado no varía, o varía muy lentamente.

Los errores de predicción serán variables aleatorias de una misma distribución. Si e_t es no correlacionado, centrado y varianza σ^2 entonces se espera que los errores de predicción ϵ_{t+1} tengan una varianza parecida a σ^2 .

Se elige entonces el valor de $\Sigma_0 = N_0 \sigma^2$, donde N_0 es la "ventana efectiva" que se desea tener en el caso estacionario,

con el sistema bien identificado. Es necesario por lo tanto, tener una idea del valor de la varianza del ruido para seleccionar "a priori" el valor de la cantidad de información Σ_0 deseada.

Una variante bastante empleada en la práctica consiste en elegir la ventana efectiva del caso estacionario, y elegir una cantidad de información:

$$\Sigma_t = N_0 \hat{\sigma}_t^2 \quad (3.83)$$

donde $\hat{\sigma}_t^2$ es una estimación instantánea de σ^2 , usualmente computada como

$$\hat{\sigma}_t^2 = \alpha \hat{\sigma}_{t-1}^2 + (1-\alpha) \epsilon_t^2 \quad (3.84)$$

donde α es un factor de olvido (o constante de tiempo) predeterminada.

Trataremos sólo el caso Σ_0 constante.

Cálculo recursivo

El cálculo de λ_t se hace igualando en (3.81) $I_{t+1} = \Sigma_0$.

De la ecuación (3.11) se obtiene que:

$$P_t = \frac{1}{\lambda_t} P^*_t \quad (3.85)$$

donde P^*_t es una matriz colineal con P_t que depende de las observaciones hasta el instante t .

Sustituyendo en (3.81):

$$\Sigma_0 = \lambda_t \Sigma_0 + \frac{\epsilon_{t+1}^2}{1 + x_{t+1}^T P^*_t x_{t+1} / \lambda_t}$$

$$\Sigma_0 = \lambda_t \Sigma_0 + \frac{\lambda_t \epsilon_{t+1}^2}{\lambda_t + x_{t+1}^T P^*_t x_{t+1}}$$

donde la incógnita es λ_t .

Esto conduce a una ecuación cuadrática. Para simplificar el cálculo recursivo, se calcula λ_{t+1} (un instante adelante) como el valor de λ_t suponiendo ya fija la matriz P_t , con:

$$\Sigma_0 = \lambda_{t+1} \Sigma_0 + \frac{\epsilon_{t+1}^2}{1 + x_{t+1}^T P_t x_{t+1}}$$

$$\therefore \lambda_{t+1} = 1 - \frac{\epsilon_{t+1}^2}{\Sigma_0 (1 + x_{t+1}^T P_t x_{t+1})} \quad (3.86)$$

En realidad es una aproximación, y esto puede dar lugar a un error, en general pequeño. Pero λ_{t+1} puede llegar a ser muy

chico, incluso negativo. Para evitar esto se introduce la condición

$$\lambda_{t+1} \geq \lambda_{\text{mín}} \quad (3.87)$$

finalmente el algoritmo queda como sigue:

- i) Cálculo de $\hat{\theta}_{t+1}$, P_t con las ecuaciones (3.9) y (3.10)
- ii) Cálculo de λ_{t+1} con las ecuaciones (3.86), (3.87).

En Osorio/Mayne [5], se puede encontrar una prueba de la convergencia del algoritmo en el caso determinístico.

Inicialización

Hay varias formas de inicializar el algoritmo de Fortescue.

Dos posibles son: 1) inicializar con $I_0 = 0$, y mantener un factor de olvido constante $\lambda < 1$ hasta que $I_1 \geq \Sigma_0$, a partir de $t = 1+1$ forzar la condición $I_t = \Sigma_0$.

2) inicializar con $I_0 = \Sigma_0$.

De las expresiones (3.9) y (3.81) se deduce que el valor de la función de costo J_t , converge a I_t con $t \rightarrow \infty$.

Esto es: $J_t \rightarrow I_t = \Sigma_0$ si $t \rightarrow \infty$

En efecto: $(J_t - I_t) = (J_0 - I_0)(\lambda_1 \dots \lambda_{t-1})$, con $\lambda_i < 1$
 $\lambda_i \geq 0$

Simulación

Sobre el mismo sistema simulado en la sección 3.2 de factor de olvido constante, se llevaron a cabo una serie de identificaciones con el algoritmo de información constante.

La figura (3.17) muestra la estimación de los parámetros con una relación señal a ruido de $S/N = 25$, y una cantidad de información $\Sigma_0 = 4$. De esta manera $N_0 = 100$ en estado estacionario, para lo cual $\lambda_t \rightarrow 0.99$ en estado estacionario.

Puede observarse que los resultados si $t < 300$ son semejantes a los de la figura 3.4.a, es decir la simulación con $\lambda = 0.99$. A partir de $t = 300$ el cambio en el parámetro se detecta, reduciendo el factor de olvido y obteniendo una velocidad de adaptación ligeramente inferior a la que se obtendría con un factor de olvido de 0.96 (con ventana efectiva asociada de 25 pasos). Comparar con la fig. 3.4.c.

La segunda parte de la figura muestra una cantidad que hemos convenido en denominar "ventana efectiva instantánea" N_t , que

es la suma de los pesos de la función de costo:

$$N_t = \sum_{i=1}^t \left(\prod_{j=i}^t \lambda_j \right) \quad (3.88)$$

Esta cantidad es una buena medida de la "ventana efectiva" real que toma en cuenta el algoritmo, y es más clara de visualizar que la sucesión λ_t que en general es bastante oscilante.

Puede verse que a partir de $t = 1$, N_t es creciente, se estabiliza entre $140 \leq t \leq 300$ en valores N_t entre 70 y 95 pasos.

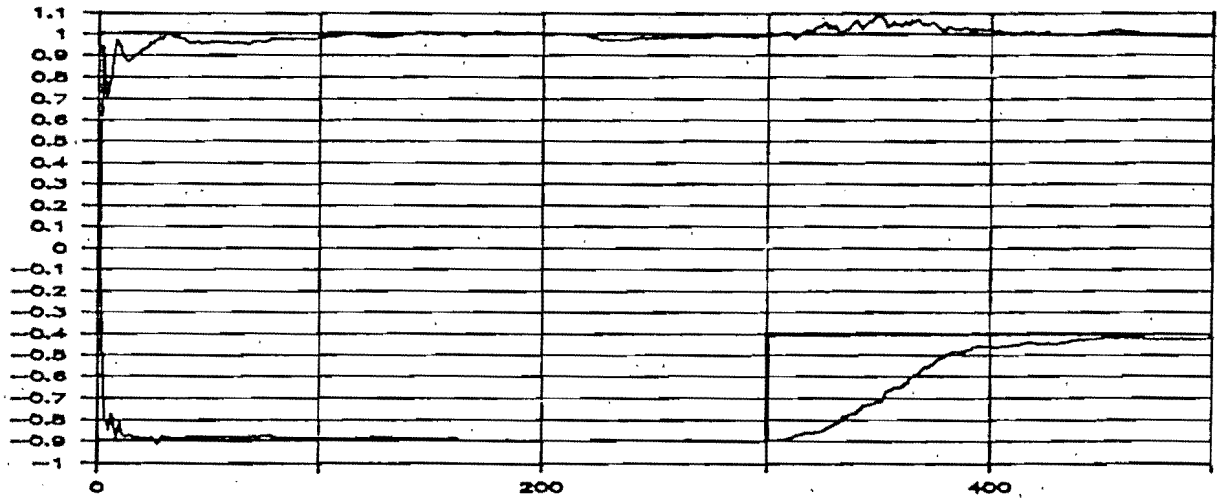
Pero a partir de $t = 300$, N_t decrece hasta cerca de 25 pasos, para volver a crecer nuevamente.

Esto puede interpretarse como que el algoritmo funcionó con un factor de olvido de alrededor de $\lambda = 0.988$ si $t \leq 300$, luego del salto en el parámetro con $\lambda \approx 0.96$, para volver a crecer λ al valor original luego de la adaptación.

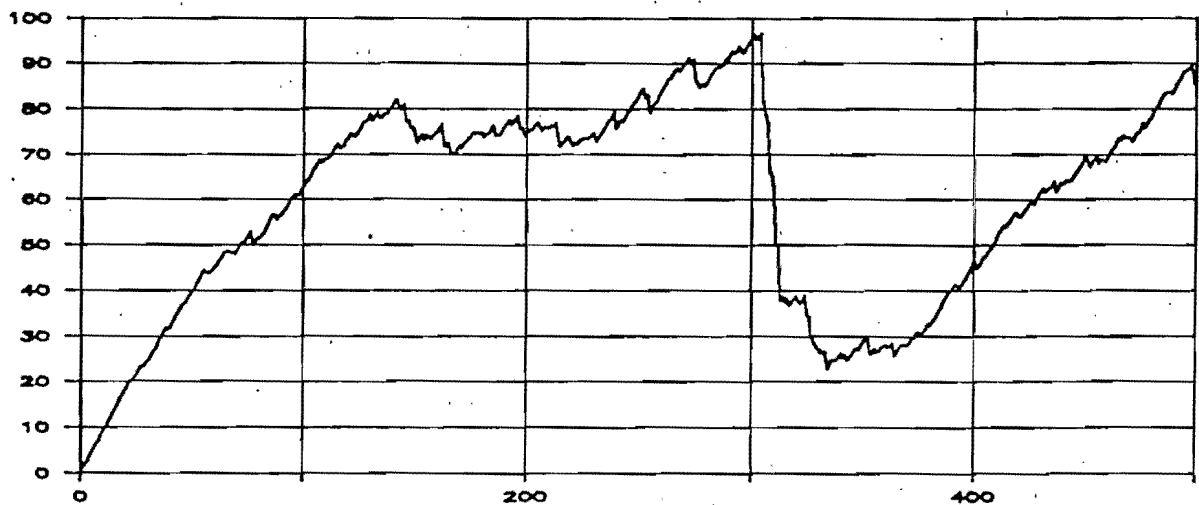
En este caso el algoritmo funcionó como se deseaba, con valores altos de λ_t si no hay cambios en el parámetro, y bajando λ_t cuando se detectaran dichos cambios.

La figura 3.18 muestra una simulación análoga, con $S/N = 2$ y $\Sigma_0 = 50$.

Info=4 S/N=25



Info=4 S/N=25, Ventana efectiva.



Estimación con el algoritmo de información constante
Relación señal a ruido=25, Cantidad de Información=4

Figura 3.17

El valor de Σ_0 se eligió para volver a tener una ventana efectiva en estado estacionario $N_0 = 100$.

El resultado es distinto al caso anterior. Los estimadores se comportan como en el caso de $\lambda = 0.99$ (ver fig 3.6.a). La gráfica de la ventana efectiva instantánea muestra ligeras oscilaciones alrededor de $N_t = 85$ (que corresponde a $\lambda = 0.988$) confirmando la similitud de los resultados de la figura 3.6.a.

¿Cuál fué el problema? ¿Porqué hay una diferencia cualitativa tan grande respecto al ejemplo anterior?

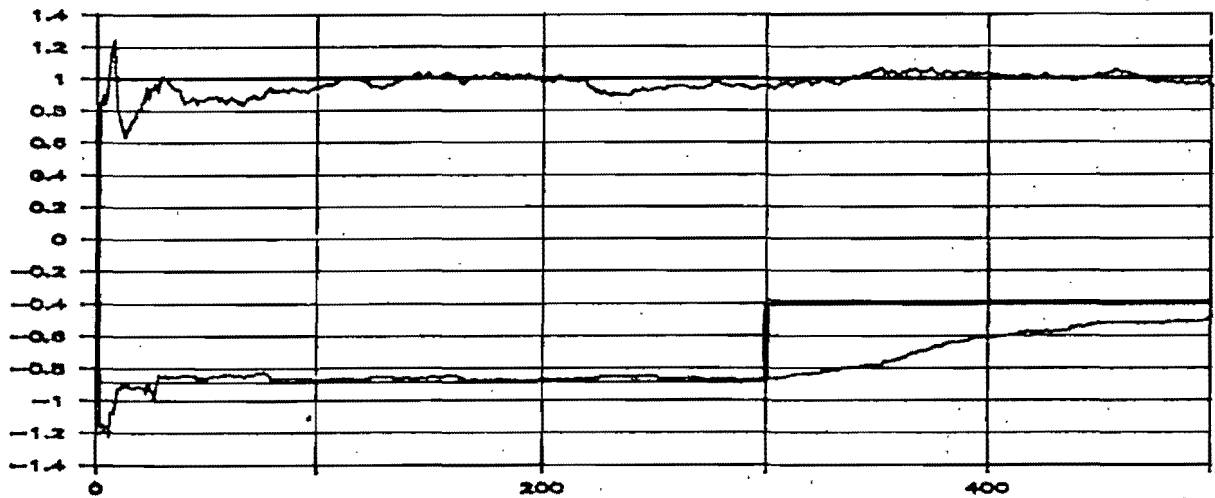
La respuesta radica en el nivel de la relación señal a ruido.

Recordemos la ec. (3.79). Si el valor σ^2 es muy grande, será difícil distinguir a partir de ϵ_{t+1}^2 que hay error en la estimación del parámetro pues σ^2 predominaría frente al primer sumando del lado derecho de (3.79).

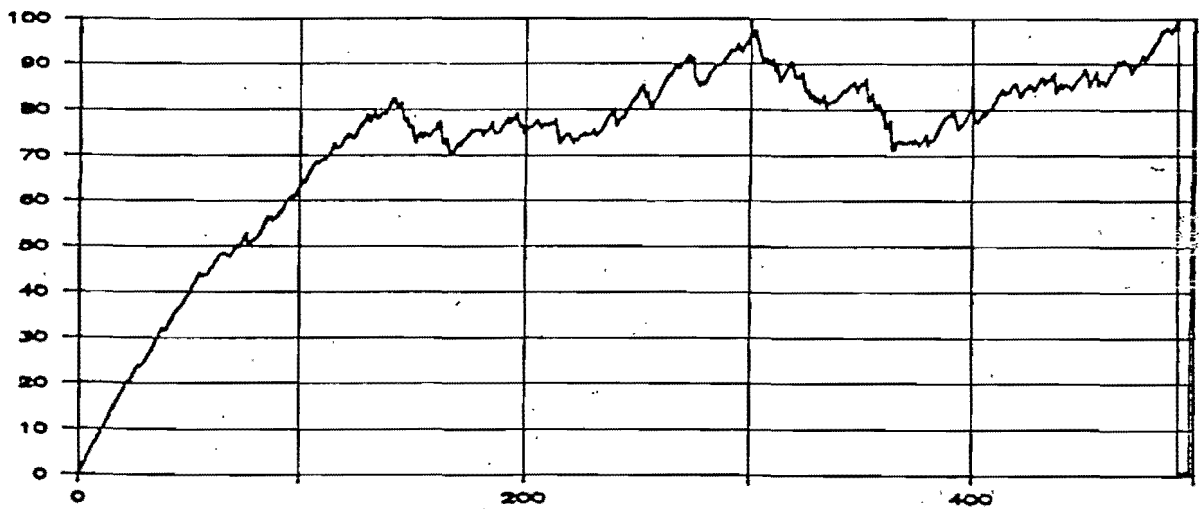
Tanto el algoritmo de cantidad de información constante como cualquier otro algoritmo que use ϵ_{t+1}^2 para modular la sucesión λ_t fallarán en la detección de saltos en los parámetros cuando el nivel de la relación señal a ruido sea bajo.

De (3.79) se puede concluir que además de la condición de excitación persistente, es necesario que

Estimados ,Info=50 S/N=2



Info=50 S/N=2, Ventana efectiva.



Identificación con el algoritmo de Información constante
Relación señal a ruido=2, Cantidad de Información = 50.

Figura 3.18

$$x_i^T x_i / \sigma^2 \quad (\text{o } m / \sigma^2)$$

tenga un valor alto, para que el algoritmo de información constante sea sensible a los cambios en el parámetro.

3.4.2. Algoritmo de traza constante

El algoritmo de mínimos cuadrados ordinario ($\lambda = 1$), tiene la propiedad de converger en media cuadrática al parámetro real, cuando este es constante (sistema invariante) y bajo algunas condiciones de regularidad (perturbaciones no correlacionadas, y excitación persistente).

En términos numéricos esto se traduce en que $P_t \rightarrow 0$ si $t \rightarrow \infty$ y $\hat{\theta}_t \rightarrow \theta^0$ (real) si $t \rightarrow \infty$. La convergencia será con probabilidad 1 o en media cuadrática según las restricciones que se impongan.

Observando la ecuación (3.11), con $\lambda = 1$, puede verse que siendo

$$P_t^{-1} = \sum_{i=1}^t x_i x_i^T + P_0^{-1} \quad (3.89)$$

(P_0 elegida como se vió anteriormente)

Luego P_t es definida positiva bajo condiciones de excitación

persistente, y además decreciente.

Si $P_t \rightarrow 0$, y existe un cambio en el valor del parámetro θ , el algoritmo no tendrá capacidad de seguimiento: P_t sigue siendo decreciente pues depende solo del vector de observaciones y no del error de predicción.

El efecto de $\lambda_t < 1$ se traduce en impedir el decrecimiento continuo de P_t . Bajo condiciones de excitación persistente y con x_t estacionario y ergódico, puede verse que $\lambda_t = \text{cte} < 1$ dá lugar a la convergencia de P_t a una matriz definida positiva $P_t \rightarrow P > 0$.

Esa es una forma de abordar el problema. Otra es asegurar que los autovalores de P_t no decrezcan a cero indefinidamente. Esto se puede hacer de varias maneras, una de ellas (ver p.ej. Landau, Lozano [27]; Lozano, Landau [15]) es mantener la traza de P_t constante.

La forma recursiva de hacerlo es muy sencilla: de las ecuaciones (3.9) y (3.10), calcular el estimador con la primera y sustituir la segunda por:

$$P_{t+1}^* = P_t - \frac{P_t x_{t+1} x_{t+1}^T P_t}{1 + x_{t+1}^T P_t x_{t+1}} \quad (3.90)$$

$$\lambda_{t+1} = \frac{t_r P_{t+1}^*}{t_r P} \quad (3.91)$$

$$P_{t+1} = \frac{1}{\lambda_{t+1}} P_{t+1}^* \quad (3.92)$$

donde $t_r P$ es el valor elegido para la traza de P_i , $i=1,2,\dots$

Este algoritmo fue utilizado en forma exitosa para el diseño de estrategias de control autoadaptivo de sistemas determinísticos.

Debe notarse que no toma en cuenta el error de predicción como medida de que el sistema está cambiando para variar su velocidad de adaptación, sino que mantiene su capacidad de adaptación por la vía de mantener por encima de cierto valor la traza de la matriz P_t .

Un inconveniente de este método es que para elegir la velocidad de adaptación del algoritmo, y elegir $t_r P$, hay que conocer previamente en que orden de magnitud estará el vector x_{t+1} .

3.4.3 Detección basada en el error de predicción

Otra manera de elegir λ_t es a través de un proceso de detección del cambio en el parámetro. La idea es actualizar en

cada paso una cantidad d_t que funcione como detector de cambios en el parámetro. Esta cantidad debe de crecer por encima de cierto umbral cuando ocurre un cambio en el parámetro.

El factor de olvido se puede calcular de dos maneras:

- i) Como una función de la cantidad d_t
- ii) Mantener $\lambda_t = \lambda_1$ (alto) si $d_t < U$, y $\lambda_t = \lambda_2$ (bajo) si $d_t \geq U$, donde U es cierto umbral de detección.

Una forma bastante común para d_t es la comparación de dos promedios de los errores cuadráticos de predicción.

$$\text{Llamando } \begin{cases} \hat{\sigma}_{\ell_t}^2 = \sum_{i=1}^t \alpha_1^{t-i} \epsilon_i^2 & (1 - \alpha_1) \\ \hat{\sigma}_{c_t}^2 = \sum_{i=1}^t \alpha_2^{t-i} \epsilon_i^2 \cdot (1 - \alpha_2), & \text{con } 1 > \alpha_1 > \alpha_2 \end{cases} \quad (3.93)$$

$\hat{\sigma}_{\ell_t}^2$ es un "promedio largo", que intenta estimar el valor de σ^2 , mientras que $\hat{\sigma}_{c_t}^2$ es un "promedio corto" que intenta estimar la varianza actual de ϵ_{t+1} .

La idea es construir d_t como:

$$d_t = \hat{\sigma}_{c_t}^2 / \hat{\sigma}_{\ell_t}^2 \quad (3.94)$$

Se espera que si hay cambios en el parámetro, crezcan los errores de predicción y por lo tanto d_t . Luego se utiliza d_t para computar λ_t de alguna de las dos maneras mencionadas.

Esta clase de mecanismos de detección fallan cuando el nivel del ruido es muy alto por las mismas razones ya expuestas en el caso del algoritmo de información constante.

A modo de ejemplo, se realizó el cómputo de d_t para el mismo caso simulado de la sección 3.2. Los valores de α_1 y α_2 fueron 0.99 y 0.95 respectivamente.

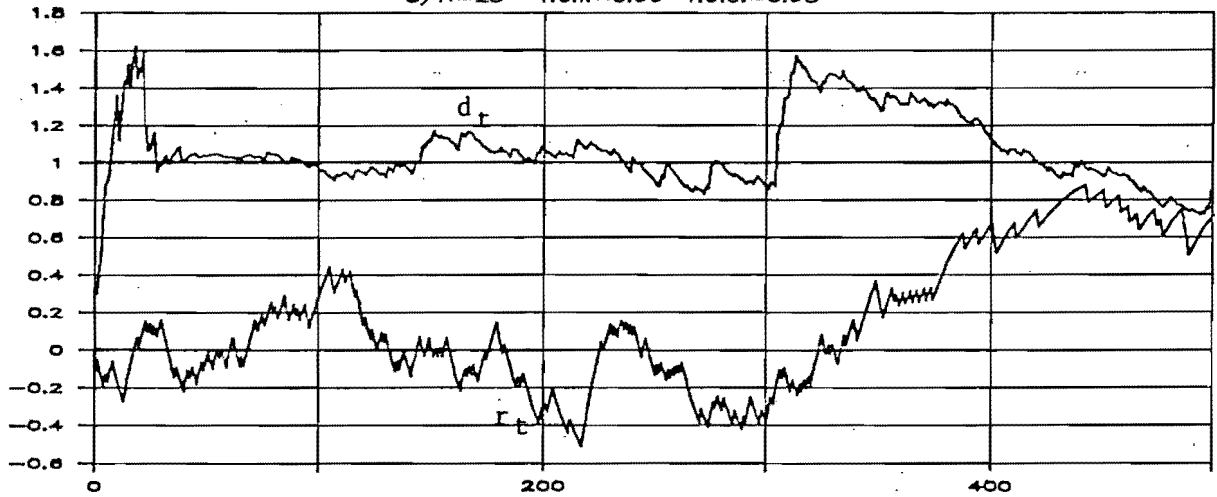
En la figura (3.19), las curvas superiores muestran en a), b), c) la evolución de d_t para relaciones señal a ruido de 25, 6.25 y 2, respectivamente. Se desprende de su observación, que para una relación señal a ruido de 25 se podría obtener a partir de d_t , información sobre un salto en el parámetro. Si la relación señal a ruido es 2, no es posible hacer esto.

3.4.4 Mecanismos robustos de detección

Tratando de evitar el problema de la influencia del nivel del ruido, T. Hägglund [47], propone un mecanismo robusto de detección de saltos aislados en el parámetro. Siguiendo su razonamiento, definimos

Razon de errores y test de salto

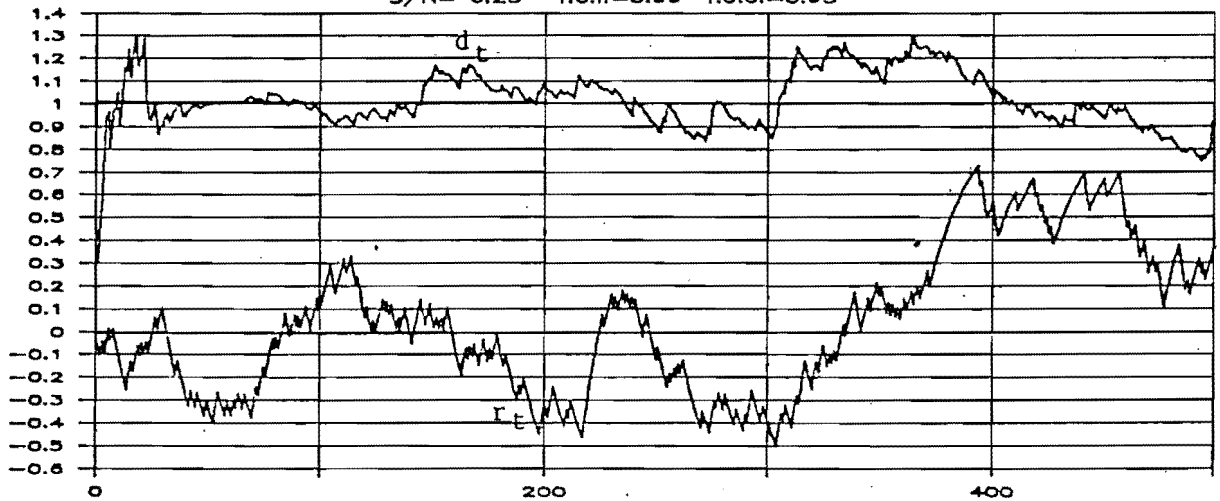
$S/N=25$ f.o.l.=0.99 f.o.c.=0.95



(a)

Razon de errores y test de salto

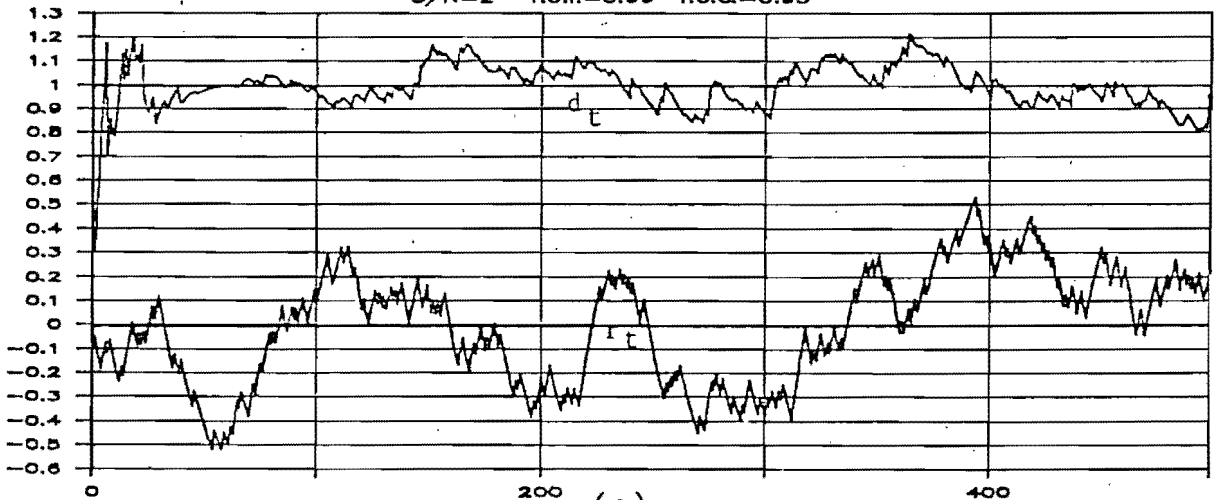
$S/N=6.25$ f.o.l.=0.99 f.o.c.=0.95



(b)

Razon de errores y test de salto

$S/N=2$ f.o.l.=0.99 f.o.c.=0.95



(c)

Figura 3.19

$$\Delta \hat{\theta}_t = \hat{\theta}_t - \hat{\theta}_{t-1} \quad (3.95)$$

nótese que por (3.9)

$$\Delta \hat{\theta}_{t+1} = \frac{P_t x_{t+1}}{1 + x_{t+1}^T P_t x_{t+1}} \epsilon_{t+1} \quad (3.96)$$

Esta cantidad es siempre computada para actualizar el estimador $\hat{\theta}_{t+1}$.

Si el sistema está relativamente bien identificado, se espera que el cambio en la estimación $\Delta \hat{\theta}_{t+1}$ sea en sentido aleatorio, luego las probabilidades:

$$P\{\Delta \hat{\theta}_t^T \Delta \hat{\theta}_{t-1} < 0\} \approx P\{\Delta \hat{\theta}_t^T \Delta \hat{\theta}_{t-1} > 0\} \quad (3.97)$$

Pero luego de cambio abrupto en el parámetro, se espera que el estimador se actualice (en tendencia) en dicha dirección, luego:

$$P\{\Delta \hat{\theta}_t^T \Delta \hat{\theta}_{t-1} < 0\} < P\{\Delta \hat{\theta}_t^T \Delta \hat{\theta}_{t-1} > 0\} \quad (3.98)$$

Lo que se propone es estudiar el signo del producto escalar de el último incremento con un promedio de los incrementos in mediatos anteriores.

Definiendo:

$$\left\{ \begin{array}{l} W_t \stackrel{\Delta}{=} \gamma_1 W_{t-1} + \Delta \hat{\theta}_t \quad 0 \leq \gamma_1 < 1 \end{array} \right. \quad (3.99)$$

$$\left\{ \begin{array}{l} S_t \stackrel{\Delta}{=} \text{signo}\{\Delta \hat{\theta}_t^T W_{t-1}\} \end{array} \right. \quad (3.100)$$

$$\left\{ \begin{array}{l} r_t \stackrel{\Delta}{=} \gamma_2 r_{t-1} + (1 - \gamma_2) S_t \quad 0 \leq \gamma_2 < 1 \end{array} \right. \quad (3.101)$$

Si no hay cambios en el parámetro, se espera que S_t tenga probabilidad parecida de valer +1 o -1, por lo que r_t debería oscilar alrededor del valor 0.

Si hay un cambio abrupto en el parámetro, se espera que S_t tenga más probabilidad de ser +1 que -1, luego r_t deberá crecer por encima de cero.

Si se supone que no hay cambios en el parámetro, y γ_2 es próximo a 1, Hägglund concluye que r_t es aproximadamente gaussiana, con media cero y varianza:

$$\sigma_r^2 = \frac{1 - \gamma_2}{1 + \gamma_2} \quad (3.102)$$

La prueba consiste en comparar r_t contra un umbral predeterminado r_0 .

Si $r_t > r_0$ se decide que hubo un cambio en el parámetro. Por

ejemplo, si $\gamma_2 = 0.95$, se puede elegir $r_0 = 0.5$ para una probabilidad de error de un instante de muestreo cada 1000, de acuerdo a (3.102).

Lo importante de este método es que se basa en la orientación de los incrementos del estimador, no hace suposiciones sobre la distribución ni magnitud del ruido. Se espera que se comporte en forma más robusta que los métodos descritos anteriormente.

Simulaciones

Este método fue empleado en el mismo ejemplo simulado de la sección 3.2. Los valores utilizados fueron $\gamma_1 = \gamma_2 = 0.95$, junto con una identificación donde se mantuvo un factor de olvido constante $\lambda = 0.99$ a lo largo del experimento.

La figura (3.19) a) b) c) muestra en sus curvas inferiores la evolución del "test de salto" r_t para valores de la relación señal a ruido $S/N = 25, 6.25$ y 2 respectivamente.

Respecto al mecanismo de detección de la sección anterior se notan dos hechos:

- 1) Es más robusto respecto al nivel del ruido. Ver por ejemplo el caso en que $S/N = 6.25$, fig 3.19.b.

- 2) La desventaja es que es más lento. Para valores de $S/N = 25$ y 6.25 , tarda entre 70 y 80 instantes de muestreo en detectar el cambio en el parámetro.

Finalmente hay que notar que ya para valores bajos de S/N , como es el caso $S/N = 2$ la detección resulta más imprecisa.

Siguiendo el mismo razonamiento que lleva a Hägglund a emplear W_t en lugar de $\Delta\hat{\theta}_t$, para estudiar tendencias, ensayamos una modificación del algoritmo, empleando

$$S_t = \text{signo} \{ z_t^T W_{t-r} \} \quad (3.103)$$

donde

$$z_t = \alpha_1 \Delta\hat{\theta}_t + \dots + \alpha_r \Delta\hat{\theta}_{t-r+1} \quad (3.104)$$

Se hizo para dos casos:

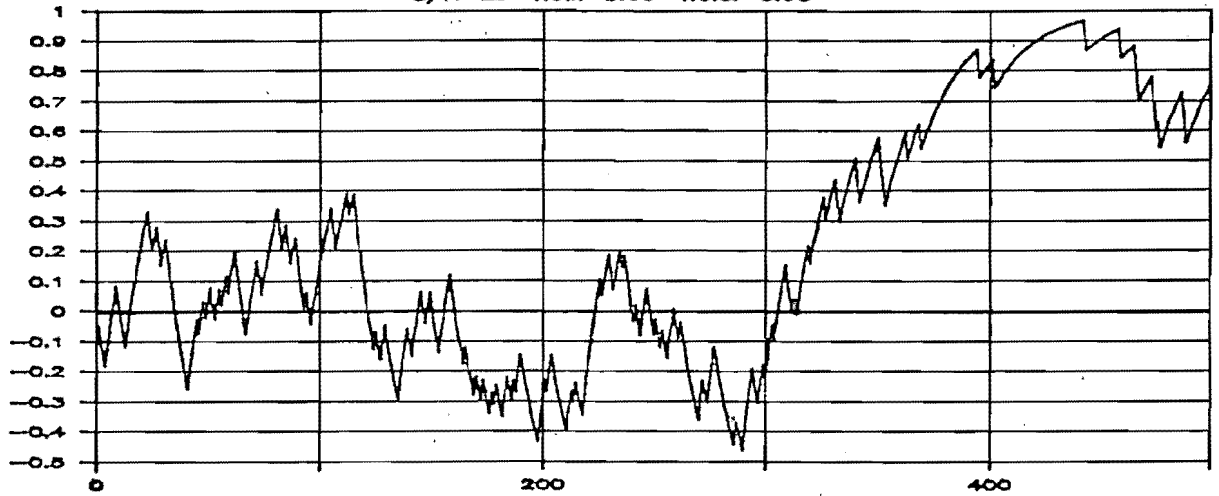
$$\text{i) } r = 2, \quad z_t = 0.5 \Delta\hat{\theta}_t + 0.5 \Delta\hat{\theta}_{t-1} \quad (3.105)$$

$$\text{ii) } r = 5, \quad z_t = 0.2 \sum_{i=0}^4 \Delta\hat{\theta}_{t-1} \quad (3.106)$$

Las figuras 3.20 y 3.21 muestran los resultados obtenidos en las simulaciones para $r = 2$ y $r = 5$ respectivamente. Los valores de γ_1 y γ_2 fueron 0.95 y las relaciones señal a ruido fueron de 25, 6.25 y 2.

Test de salto ($r=2$)

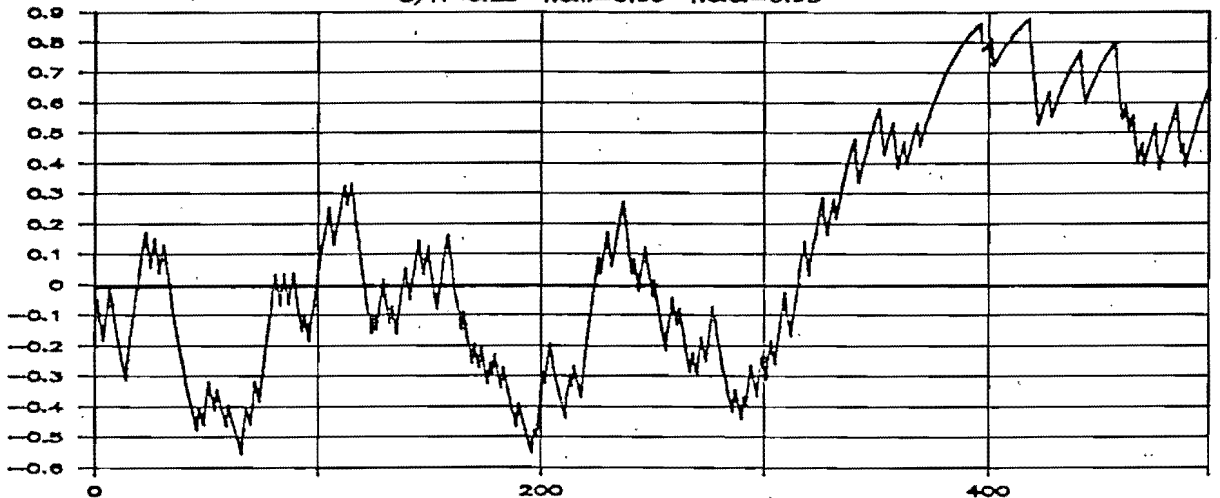
$S/N=25$ f.o.l.=0.99 f.o.c.=0.95



(a)

Test de salto ($r=2$)

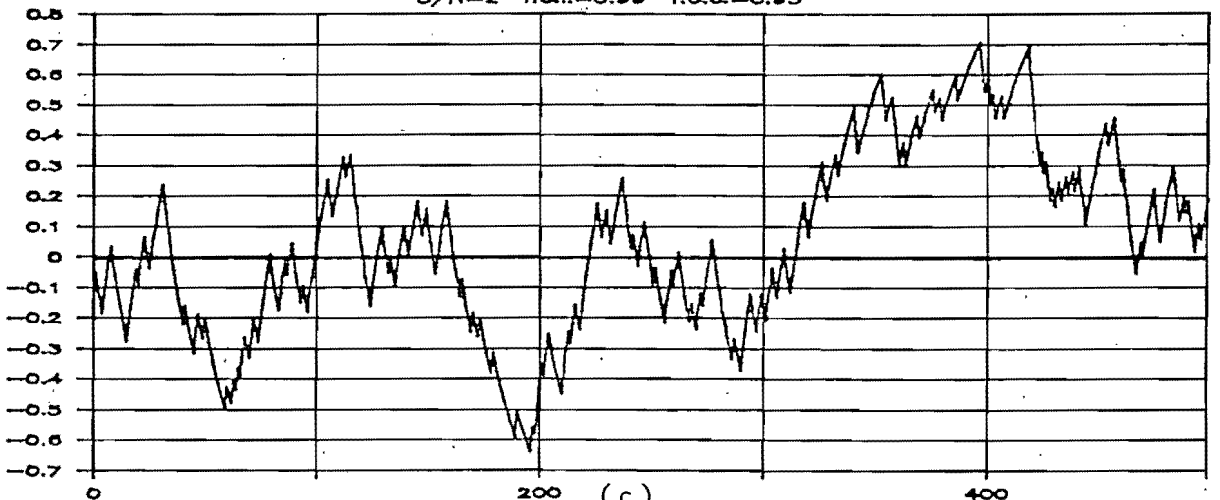
$S/N=6.25$ f.o.l.=0.99 f.o.c.=0.95



(b)

Test de salto ($r=2$)

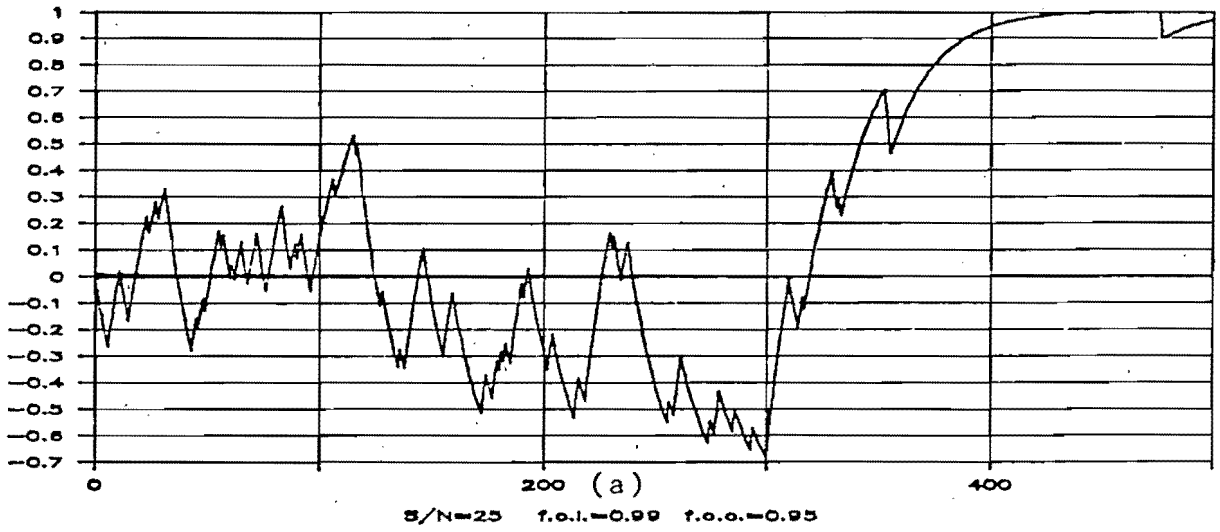
$S/N=2$ f.o.l.=0.99 f.o.c.=0.95



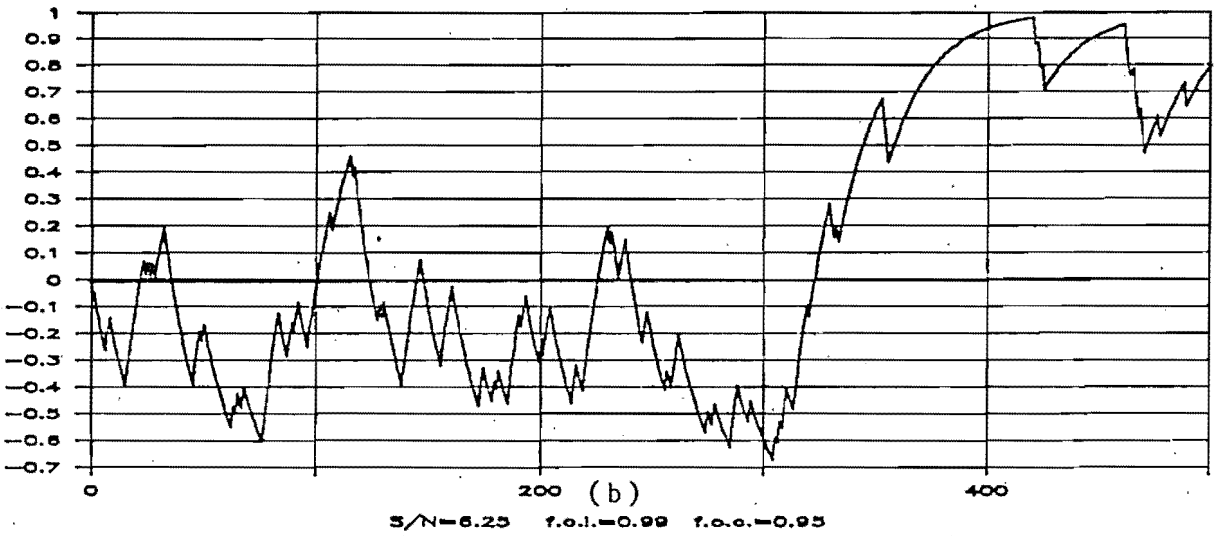
(c)

Figura 3.20

Test de salto (r=5)



Test de salto (r=5)



Test de salto (r=5)

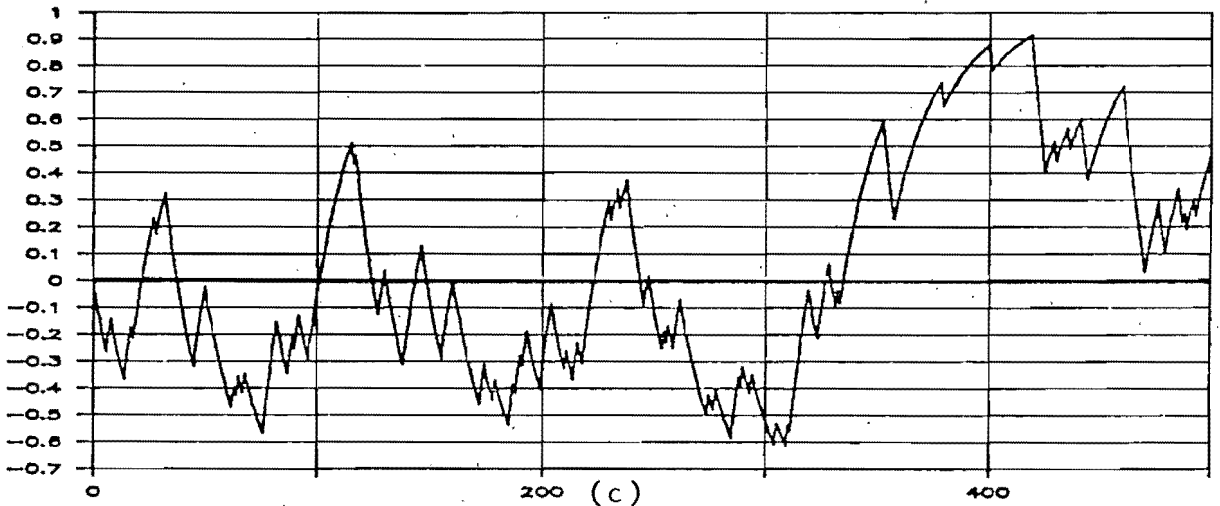


Figura 3.21

Se nota una muchísimo mayor robustez respecto al nivel del ruido. De hecho, manteniendo el umbral de 0.5 son netamente detectables los saltos en el parámetro.

Aunque se introdujo un retardo adicional de r pasos en la detección (ver (3.103) y (3.104)), se redujo a cerca de la mitad el tiempo de detección del salto en el parámetro.

Se ensayaron valores más altos de r sin conseguir ventajas evidentes.

Observaciones:

De los experimentos realizados podemos concluir que:

- . Una modificación al método propuesto por Hägglund, da lugar a un mecanismo bastante más rápido de detección del salto en el parámetro y robusto respecto el nivel de ruido presente.
- . Este método se puede complementar con esquemas basados en el error de predicción como por ejemplo el algoritmo de información constante. Se puede llevar a cabo el cómputo de λ_t mediante el algoritmo de información constante, que detectará los saltos grandes en el parámetro y será eficiente cuando el nivel de ruido no sea muy alto.

En forma paralela se puede computar r_t y detectar saltos más pequeños en el parámetro ó saltos en presencia de altos niveles de ruido. De esta manera se aprovecharían las ventajas de ambos procedimientos.

3.5 Algunas aplicaciones reportadas

A modo de ejemplo de la variedad de procesos en que ha sido empleado, enumeraremos algunas de las aplicaciones del algoritmo de mínimos cuadrados con factor de olvido que han sido reportadas.

El control autoadaptivo de una planta de acondicionamiento de aire es reportado por R. Shumann [8]. Se controla temperatura y humedad relativa de aire a un gasto promedio de 400 a 500 m^3/h .

Se utiliza un controlador adaptivo tipo dead-beat con identificación a factor de olvido constante.

En el Instituto de Ingeniería (Canales/Canetti [18]), se desarrolló un controlador autoadaptivo en un microcomputador de un solo chip (8748-Intel). El factor de olvido es constante; el controlador puede ser de dos tipos: dead-beat o por asignación de polos. Fue ensayado en el control de pequeños motores de corriente directa.

Dumont [1] reporta la utilización de un regulador autoadaptivo utilizando el esquema de cantidad de información constante para el control de una planta de fabricación de pulpa de madera por fricción de placas.

Fortescue et al. [2] reportan la utilización del mismo algoritmo de cantidad de información constante para el control de una planta experimental de absorción y desorción de anhídrido carbónico.

3.6 Conclusiones

En este capítulo se ha discutido en forma bastante extensa el algoritmo de mínimos cuadrados con factor de olvido variable. De los métodos estudiados para la identificación de sistemas variables en el tiempo es el que hemos elegido para las aplicaciones que nos ocuparán en los capítulos que siguen.

La razón de esta elección se encuentra en su sencillez numérica, versatilidad en el sentido de requerir poca información apriorística sobre el proceso a identificar, y porque encaja dentro del esquema de predicción lineal multipasos que emplearemos en el resto del trabajo.

Varios autores han estudiado este algoritmo. Johnstone y Anderson [48] han estudiado la convergencia para el caso deter-

minístico dentro de un esquema de control en lazo cerrado, y con factor de olvido constante.

Osorio Cordero y Mayne [5] establecieron la convergencia para el caso determinístico de un regulador basado en este algoritmo con una forma particular de elección del factor de olvido (información constante).

R. Lozano [39] estudió la convergencia para el caso general: factor de olvido variable y perturbaciones estocásticas. Los resultados obtenidos tienen el inconveniente de fallar en la convergencia cuando el factor de olvido tiende a la unidad.

En las secciones anteriores se estudió la convergencia del algoritmo, llegando a resultados más fuertes que los existentes en la literatura en el sentido que se contempla la situación $\lambda \rightarrow 1$, para el caso general de factor de olvido variable y perturbaciones aleatorias.

Finalmente se discutieron distintas maneras de elegir el factor de olvido y detectar cambios en los parámetros. Estos métodos así como los resultados generales de convergencia fueron ilustrados con simulaciones.

El análisis general del algoritmo así como los ejemplos simulados muestran la posibilidad de utilizar el algoritmo de mí-

nimos cuadrados con factor de olvido variable con bastante ro
bustez frente a condiciones de alto nivel de ruido presente,
así como asegurando la convergencia (y evitando problemas de
"bursting" y "blow-up") por la vía de mantener condiciones de
excitación persistente.

4. APLICACION A PREDICCIÓN AUTOAJUSTABLE

4.1 Introducción

En la práctica se presenta con frecuencia la necesidad de predecir el comportamiento de la salida de un sistema, dentro de un horizonte de tiempo futuro de longitud pre-definida. Un ejemplo, que será discutido en el capítulo siguiente, es la predicción del caudal de un río en una sección determinada, con base en las observaciones de las precipitaciones hasta el instante presente.

Diferentes autores han ido dando soluciones cada vez más sencillas a este problema.

Cuando se conocen los parámetros del sistema, el esquema propuesto por Aström [29] permite predecir la salida y_{t+k} en el

intervalo: $1 \leq k \leq \ell$, implementando ℓ predictores independientes en paralelo. Cada uno de ellos predice a un intervalo de k instantes de muestreo $1 \leq k \leq \ell$. Wittenmark [31], desarrolla una versión autoadaptativa para el caso de parámetros desconocidos aplicable a cada uno de estos predictores.

Sin embargo el problema se vuelve complejo si ℓ es grande, por la cantidad de predictores y parámetros a identificar, en el caso autoadaptativo.

Una solución más sencilla fue propuesta por De Keyser, Van Cawenberghe [34]. Allí se muestra que el número de predictores autoadaptativos en paralelo que es necesario implementar es: $\max(n, \ell)$, donde n es el orden del sistema y ℓ la longitud del horizonte de predicción.

Más adelante España [25], muestra que cualquier predicción óptima puede obtenerse a partir de la aplicación recursiva del predictor a 1 paso. En el caso de parámetros desconocidos sólo es necesario identificar los parámetros de este predictor para obtener asintóticamente la predicción óptima k pasos adelante, para cualquier valor de $k \geq 1$.

En las siguientes secciones se muestra esta idea.

4.2 Predictor óptimo multipasos, parámetros conocidos

Consideremos un sistema dinámico lineal en tiempo discreto, descrito desde el punto de vista de entrada-salida por:

$$(1 - A(q^{-1}))y_t = B(q^{-1})u_t + C(q^{-1})e_t \quad (4.1)$$

donde

$$\begin{cases} A(q^{-1}) = a_1 q^{-1} + \dots + a_n q^{-n} \\ B(q^{-1}) = b_0 + b_1 q^{-1} + \dots + b_m q^{-m} \\ C(q^{-1}) = 1 + c_1 q^{-1} + \dots + c_r q^{-r} \end{cases} \quad (4.2)$$

e_t es una sucesión de variables aleatorias independientes y centradas.

Supongamos conocidos los parámetros del sistema (coeficientes de A, B, C).

Haciendo uso de la identidad:

$$C(q^{-1}) = (1 - A(q^{-1})) + q^{-1}N(q^{-1}) \quad (4.3)$$

donde orden $\{N\} = \max\{n, r\} - 1$

la ecuación (4.1) puede ser escrita (ver Aström [29]) como:

$$y_{t+1} = \frac{B}{C} u_{t+1} + \frac{N}{C} y_t + e_{t+1} \quad (4.4)$$

Definimos el predictor óptimo a un paso $\hat{y}(t+1/t)$ de y_{t+1} dadas las observaciones hasta el instante t , al que minimiza la siguiente expresión:

$$E \{ [(y_{t+1} - \hat{y}(t+1/t))^2] / t \} \quad (4.5)$$

donde la esperanza E está condicionada a la σ -álgebra generada por las salidas observadas hasta el tiempo t , conocidas las entradas pasadas y futuras.

La solución (Aström [29]) está dada por los primeros dos sumandos del lado derecho de (4.4). Veamos por otro lado el enfoque de Wittenwark [31], que consiste en transformar el problema en otro de regulación con criterio de varianza mínima del error de predicción:

Sea el error de predicción en $t+1$:

$$\epsilon(t+1/t) = y_{t+1} - \hat{y}(t+1/t) \quad (4.6)$$

Por (4.4) este es:

$$\begin{aligned} \epsilon(t+1/t) &= \frac{N}{C} y_t + \frac{B}{C} u_{t+1} + e_{t+1} - \hat{y}(t+1/t) = \\ &= \frac{N}{C} \hat{y}(t/t-1) - \hat{y}(t+1/t) + \frac{B}{C} u_{t+1} + \frac{N}{C} \epsilon(t/t-1) + e_{t+1} = \\ &= - \frac{1-A}{C} \hat{y}(t+1/t) + \frac{B}{C} u_{t+1} + \frac{N}{C} \epsilon(t/t-1) + e_{t+1} \end{aligned} \quad (4.7)$$

Donde en la última igualdad se hizo uso de la identidad (4.3) y la notación

$$q^{-1}\hat{y}(t+1/t) = \hat{y}(t/t-1)$$

Como se sabe (ver Papoulis [30]), el predictor que optima (4.5) es:

$$\hat{y}(t+1/t) = E\{y_{t+1}/t\}$$

Por lo tanto, tomando esperanzas condicionales en (4.7) y teniendo en cuenta la definición (4.6), y las características de e_t resulta:

$$0 = -\frac{1-A}{C} \hat{y}(t+1/t) + \frac{B}{C} u_{t+1} + \frac{N}{C} \epsilon(t/t-1) \quad (4.8)$$

luego

$$\hat{y}(t+1/t) = \frac{B}{1-A} u_{t+1} + \frac{N}{1-A} \epsilon(t/t-1) \quad (4.9)$$

La ecuación (4.9) es la expresión recursiva para el predictor óptimo a 1 paso en función de las entradas, errores de predicción y predicciones anteriores.

Deduciremos ahora una expresión recursiva para el predictor de k pasos adelante. Reemplazando en (4.9) el índice temporal t por $t+k-1$ (con $k>1$), condicionando la expresión al tiempo t y usando la conocida relación:

$$E\{E[(\cdot)/t+k-1]/t\} = E\{(\cdot)/t\}$$

se puede demostrar que el predictor óptimo en $t+k$ está dado por:

$$\begin{aligned} \hat{y}(t+k/t) &\triangleq E(y_{t+k}/t) = E(y(t+k/t+k-1)/t) = \\ &= \frac{B}{1-A} u_{t+k} + \frac{N}{1-A} \bar{\epsilon}_{t+k-1} \end{aligned} \quad (4.10)$$

Donde $\bar{\epsilon}_{k+k} \triangleq \begin{cases} 0 & \text{si } k > 0 \\ \epsilon(t+k/t+k-1) & \text{si } k \leq 0 \end{cases}$ (4.11)

para pasar de (4.9) a (4.10) es necesario observar que:

$$\begin{aligned} E\{\epsilon(t+k/t+k-1)/t\} &= E\{[y_{t+k} - \hat{y}(t+k/t+k-1)]/t\} = \\ &= \begin{cases} = \hat{y}(t+k/t) - \hat{y}(t+k/t) = 0 & \forall k > 0 \\ = y_{t+k} - \hat{y}(t+k/t+k-1) = \epsilon(t+k/t+k-1) & \forall k \leq 0 \end{cases} \end{aligned}$$

Obsérvese que la predicción de k pasos adelante se realiza aplicando en forma recursiva el mismo predictor de 1 paso k veces. Nótese también que los errores de predicción futuros (no medidos aún en el tiempo t) son sustituidos por ceros.

Se puede reescribir la ecuación (4.10) en su forma recursiva explícita:

$$(1-A) \hat{y}(t+k/t) = B u_{t+k} + N \bar{\epsilon}_{t+k-1} \quad (4.12)$$

La notación convenida (que proviene de la deducción de (4.10)), requiere tener cuidado al calcular, en (4.12), la cantidad

A $\hat{y}(t+k/t)$. En efecto, el operador q^{-1} (retardo unitario) opera sobre las predicciones como sigue:

$$\begin{cases} q^{-j} \hat{y}(t+k/t) = \hat{y}(t+k-j/t) & \text{si } j < k \\ q^{-j} \hat{y}(t+k/t) = \hat{y}(t+k-j/t+k-j-1) & \text{si } j \geq k \end{cases} \quad (4.13)$$

La regla (4.13) puede obtenerse fácilmente aplicando el operador q^{-j} a la ecuación (4.10).

4.3 Predicador multipasos autosintonizable

Si los parámetros del sistema a predecir son desconocidos, el problema de predicción también puede ser reducido al regulador de varianza mínima aplicado al modelo (4.7).

Veremos el predicador de un paso:

Considérese el modelo (4.7) con perturbación e_{t+1} desconocida, y observaciones $\hat{y}(i+1/i)$, u_{j+1} , $e(j/j-1)$.

Se propone el algoritmo:

- a) Imponiendo la condición $C=1$, estimar los polinomios: $(1-A)$, B , N con mínimos cuadrados recursivos y observaciones hasta el tiempo t . Esto da lugar a las estimaciones $(1-\hat{A}_t)$, \hat{B}_t , \hat{N}_t .

- b) Calcular el predictor $\hat{y}(t+1/t)$ con la ecuación (4.9) y los polinomios estimados en el paso a).

Dado que en realidad el polinomio: $C \neq 1; (1-\hat{A}_t), \hat{B}_t, \hat{N}_t$ son estimadores sesgados de $(1-A), B$ y N respectivamente. Sin embargo el algoritmo tiene la propiedad de autosintonía, es decir que (ver Aström, Wittenmark [32], o Wittenmark [31]) si los órdenes de los polinomios $(1-\hat{A}_t), \hat{B}_t, \hat{N}_t$ son correctos y si sus coeficientes convergen a valores tales que $(1-\hat{A})$ no tiene ceros en común con \hat{N} y \hat{B} , entonces el predictor a un paso converge asintóticamente al óptimo. O sea:

$$\frac{\hat{B}_t}{1-\hat{A}_t} \rightarrow \frac{B}{1-A} \quad (4.14)$$

$$\frac{\hat{N}_t}{1-\hat{A}_t} \rightarrow \frac{N}{1-A}$$

Recordemos que

ver (4.3)

$$\begin{cases} \text{gr } A = n \\ \text{gr } B = m \\ \text{gr } N = \max(n, r) - 1 \end{cases}$$

Para el predictor multipasos autosintonizable faltan agregar a los pasos a) y b) los siguientes:

- c) Reemplazar $\hat{y}(t+1/t)$ calculado en b) y el valor y_{t+1} observado para calcular $e(t+1/t)$ con (4.6).

d) Usar la expresión (4.12) con los polinomios estimados en a) para calcular $\hat{y}(t+2/t) \dots \hat{y}(t+k/t)$.

Observaciones:

- Como la condición (4.14) asegura la convergencia del predictor de un paso al óptimo y el predictor de k pasos adelante es la aplicación recursiva del predictor de un paso, k veces, entonces el predictor de k pasos adelante converge al óptimo también. De esta manera el predictor de k pasos adelante "hereda" la propiedad de autosintonía del predictor de un paso.
- Como ya se ha dicho, la predicción en un horizonte de longitud ℓ con este algoritmo tiene dos ventajas importantes sobre los algoritmos anteriores.
 1. No es necesario implementar ℓ predictores de k pasos adelante ($1 \leq k \leq \ell$) en paralelo. Alcanza con uno aplicado recursivamente.
 2. El número de parámetros a identificar es menor.

Veamos esto en más detalle:

Como se puede ver en [25], el predictor de k pasos adelante tiene la expresión recursiva:

$$F_k (1-A) \bar{y}(t+k/t) = BF_k u(t+k) + N_k \bar{e}(t/t-k) \quad (4.15)$$

donde F_k , N_k se deducen de la descomposición:

$$C(q^{-1}) = F_k(q^{-1}) (1-A(q^{-1})) + q^{-k} N_k(q^{-1})$$

Los órdenes de F_k y N_k son $k-1$ y $n-1$ respectivamente. Por lo tanto el predictor de k pasos adelante hace uso de i_k parámetros, donde:

$$i_k = 2n+m-1+2k$$

Al predictor multipasos que se obtiene utilizando ℓ (o n si $\ell > n$, ver introducción) predictores en paralelo del tipo dado por la ecuación (4.15) (para $k = 2, \dots, \ell$) lo llamaremos predictor multipasos en paralelo (PMP).

Si se va a predecir sobre un horizonte de largo $\ell > n$, se requerirán de n predictores en paralelo. El número total de parámetros a identificar resulta entonces:

$$\sum_{k=1}^n i_k = 3n^2 + nm \quad (4.16)$$

Esta cantidad crece con n^2 . Por otro lado en el predictor recursivo se necesita identificar i_1 términos:

$$i_1 = 2n+m+1 \quad (4.17)$$

Esto es independiente del largo l del horizonte de predicción.

- La estructura de adaptación recursiva del predictor recursivo, admite las mismas variantes que el algoritmo de mínimos cuadrados que se usa para la identificación. Esto lo habilita a incorporar esquemas con factor de olvido como los vistos en el capítulo 2 para que el predictor realice un "seguimiento" de la variación del sistema que se intenta predecir.
- La sencillez del cálculo es enorme cuando se lo programa en una computadora digital. En la ref (35), puede encontrarse el listado del programa que se usó para las simulaciones, escrito en PASCAL para la Bourroghs 8700 del P.U.C.

4.4 Simulaciones

Como se vió en la sección anterior, el predictor multipasos re cursivo (PMR) converge asintóticamente al óptimo. También lo hacen los predictores multipasos en paralelo (PMP).

Hay dos preguntas que quedan por resolver:

¿Cuál es la velocidad de convergencia?

¿Cuál de los dos esquemas converge más rápidamente?

El resultado más general de convergencia de mínimos cuadrados que se puede aplicar, es que cuando el instante t es suficiente-

mente grande,

$$\text{Var} [\hat{\theta}_t] \propto \frac{1}{\sqrt{t-p}} \quad (4.18)$$

donde $\hat{\theta}_t$ es el vector de parámetros estimados y p el número de estos ($p = \dim \hat{\theta}_t$), para factor de olvido constante. Si el factor de olvido es variable referirse a los resultados del Capítulo 3. Este resultado presupone que los datos del problema de identificación (las observaciones de las entradas y las salidas) son generados por el sistema a identificar. Sin embargo en nuestro caso (ver fig. 4.1) lo que se pretende identificar es el predictor óptimo que tiene como entradas además de $u(t)$, el error de predicción $\epsilon(t/t-1)$. Señal que es generada con los parámetros estimados del predictor óptimo al instante t y no por el verdadero predictor óptimo.

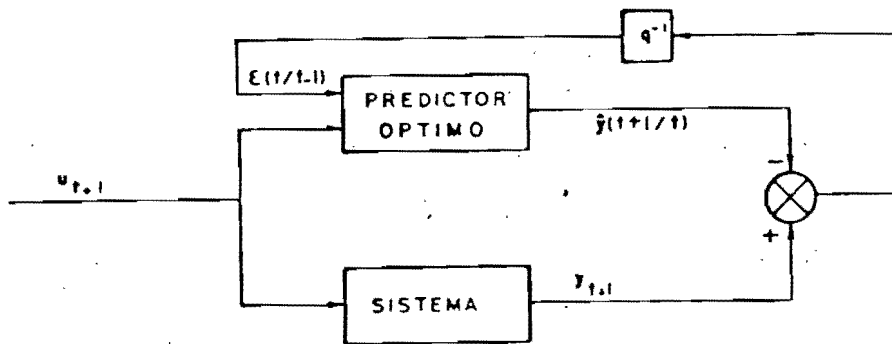


Fig. 4.1 Funcionamiento del predictor óptimo

Aunque por esta razón no podamos decir precisamente con qué velocidad converge el esquema PMR, si podemos esperar una mejora respecto al esquema PMP.

Vagamente hablando, el hecho de tener menos parámetros que identificar, significa que se necesita menos información para identificarlos. Esto es que se necesitan menos observaciones, en el esquema PMR. Se espera por lo tanto que PMR converja más rápidamente al predictor óptimo que PMP.

Para observar el comportamiento transitorio, y la capacidad de adaptabilidad de ambos esquemas se realizaron varias simulaciones. El tipo de sistemas a predecir fueron de dos clases:

- (i) Modelos ARMA: (sin entrada $u(t)$)
- (ii) Modelos ARMAX (con entrada $u(t)$), variables en el tiempo.

Se hicieron gran cantidad de simulaciones, pondremos dos ejemplos ilustrativos ya que los resultados fueron similares en todos los casos.

4.4.1 Simulaciones con modelos ARMA.

Se simularon cinco sistemas tipo ARMA (4.19) donde e_t es una sucesión de variables aleatorias independientes idénticamente distribuidas (v.a.i.i.d.), $e_t \sim N(0,1)$.

$$y_t = \frac{C(q^{-1})}{A(q^{-1})} e_t \quad (4.19)$$

El polinomio A se tomó de 4° orden en todos los casos. Cada uno de estos modelos se simuló con 10 realizaciones distintas de e_t .

Las llamaremos realización (j), $j=1\dots 10$.

Con las series $\{y_t^{(j)}\}_{t=1}^{250}$ así obtenidas, se procesaron los dos esquemas de predicción (PMR, PMP), 4 pasos adelante.

En los cinco sistemas simulados, y las diez realizaciones por sistema, se observó que:

- 1) Ambos predictores convergen a la misma predicción.
- 2) Convergen a la predicción óptima.
- 3) Converge más rápidamente el algoritmo recursivo (PMR).

Como ejemplo se describen los resultados obtenidos con la simulación del sistema:

$$y_t = \frac{1 - 0.5q^{-1}}{1 - 1.3q^{-1} + 1.7q^{-2} - 0.8q^{-3} + 0.4q^{-4}} e_t \quad (4.20)$$

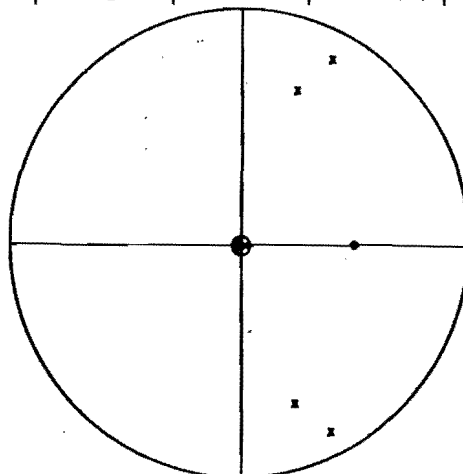


Fig. 4.2 Patrón de polos y ceros del sistema (4.20)

En los dos predictores se utilizó en la parte de identificación una matriz de covarianzas inicial $P_0 = 10 I$, $\hat{e}_0 = 0$ y un factor de olvido convergente a la unidad para evitar un fuerte transitorio inicial:

$$\lambda_{t+1} = 0.92 \lambda_t + 0.08$$

$$\text{con } \lambda_0 = 0.3$$

Sobre 10 realizaciones, se estimó la desviación cuadrática media (DCM) del error de predicción a 4 pasos como:

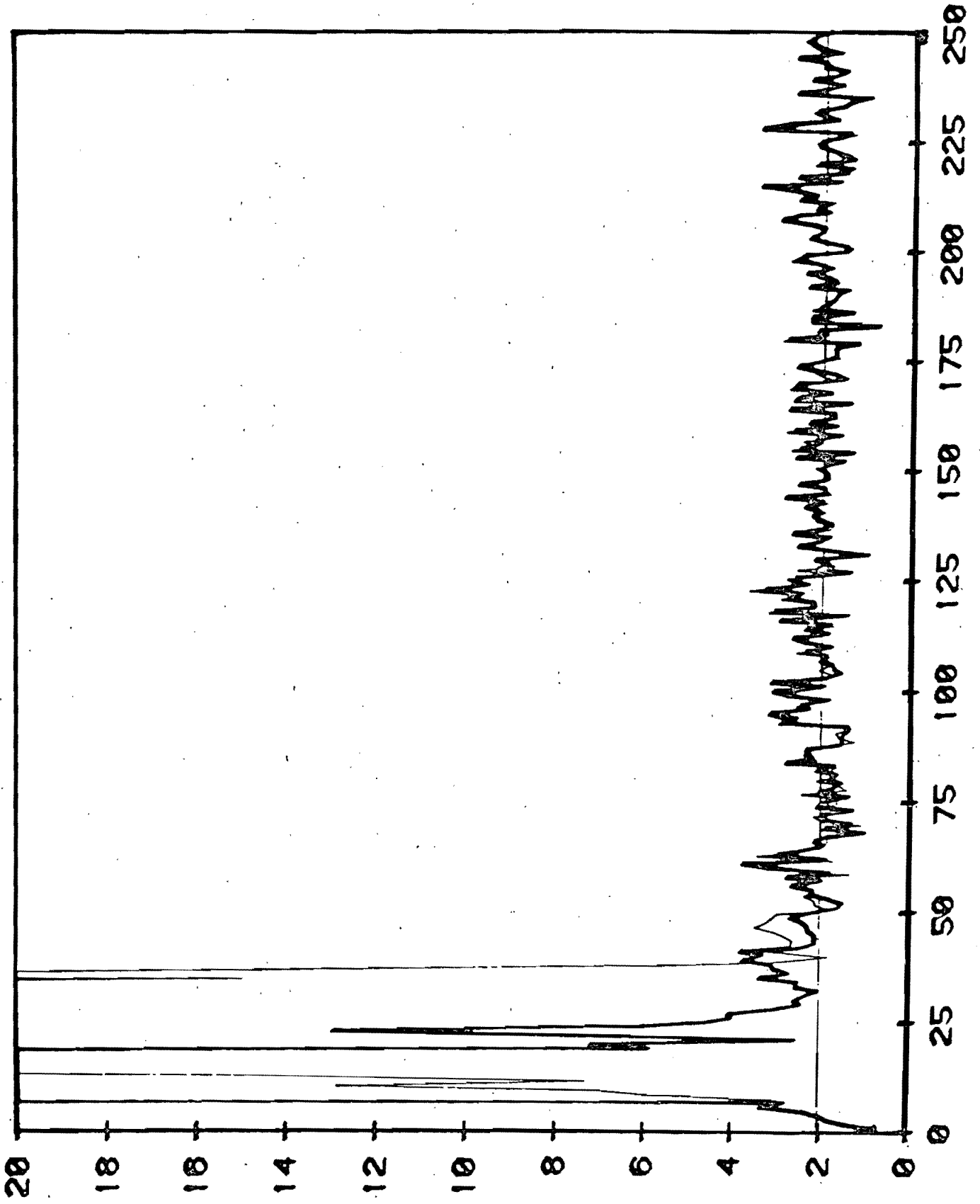
$$\text{DCM}[\hat{e}(t/t-4)] = \sqrt{\frac{\sum_{j=1}^{10} [y^{(j)}(t) - y^{(j)}(t/t-4)]^2}{10}} \quad (4.21)$$

Donde j es el índice de realización $1 \leq j \leq 10$.

En la figura 4.3 se grafica $\text{DCM}[\hat{e}(t/t-4)]$ de los dos esquemas. En línea gruesa PMR y en delgada PMP.

Se puede observar que se verifican las tres afirmaciones hechas al principio. Además es mucho menor la excursión del transitorio inicial en PMR que en PMP.

Se ha trazado una línea horizontal $\epsilon = 2.021$, que es la desviación cuadrática media esperada para el predictor óptimo.



ERROR CUAD. MEDIO DE LOS DOS PREDICTORES.

(Fig 4.3)

Los programas de predicción se encuentran en la ref (35).

Los algoritmos de simulación de señales aleatorias y sistemas dinámicos, que se utilizaron a lo largo de este trabajo, se discuten en [17].

Observación - Vale la pena comentar en este momento un detalle de carácter práctico notado en las simulaciones. Este se refiere a la parte de identificación de los predictores.

Se observará que se eligió la matriz de covarianzas inicial $P_0 = 10 I$, un valor relativamente pequeño comparado con los que usualmente se emplean para inicializar el algoritmo de mínimos cuadrados. ($P_0 \cong 100 I$).

La matriz P_0 que puede interpretarse como una medida de

$E \{ [\hat{\theta}_0 - \theta] [\hat{\theta}_0 - \theta]^T \}$ representa una estimación "apriorística" de la distancia a la que se puede hallar el valor real del parámetro θ del estimado inicial $\hat{\theta}_0$.

Algebraicamente esto se expresa en que la corrección para la actualización del parámetro es proporcional a dicha matriz; de (2.41) resulta:

$$\hat{\theta}_1 = \hat{\theta}_0 + \frac{P_0 x_1}{1 + x_1^T P_0 x_1} (y_1 - x_1^T \hat{\theta}_0)$$

Si el valor de P_0 inicial es muy pequeño la estimación del parámetro se realiza lentamente. Si P_0 es muy grande, los transitorios también lo serán.

En el caso de identificación del predictor esto es delicado porque las siguientes observaciones son generadas por las estimaciones y no por un modelo invariante. Para darle al predictor características de rápida adaptabilidad (transitorio corto) y oscilaciones menos fuertes en el transitorio inicial, se combinó una matriz de covarianzas $P_0 > I$ pero relativamente pequeña, con un factor de olvido que originalmente es bajo pero que tiende a 1.

En las simulaciones se ensayaron distintos valores de $P_0 = \delta I$, con $10 \leq \delta \leq 10^3$. Con valores altos ($\delta = 10^3$) se observó una velocidad de convergencia similar a la de las simulaciones presentadas bajo las condiciones expuestas anteriormente, pero excursiones de 1 o 2 órdenes de magnitud mayor en el transitorio.

De las observaciones de las simulaciones se concluyó que se obtenían buenos resultados (desde un punto de vista práctico), con una inicialización.

$$\begin{cases} P_0 = \delta I \\ \lambda_t = \lambda_{t-1}^\alpha + (1-\alpha) \end{cases} \quad (4.22)$$

con valores

$$\begin{cases} 10 \leq \delta \leq 30 \\ 0.9 \leq \alpha \leq 0.97 \\ \lambda_0 < 0.5 \end{cases}$$

4.4.2. Pruebas con modelos ARMAX

Se probaron los algoritmos de predicción con datos de un modelo ARMAX variable en el tiempo. El modelo fue construido de la siguiente manera:

Se simularon tres sistemas:

$$1) [1-1.7q^{-1}+1.8q^{-2}-q^{-3}+0.4q^{-4}]y_t=[1-q^{-1}+0.6q^{-2}]u_t+[0.0477-0.0239q^{-1}]e_t$$

$$2) [1-2.78q^{-1}+2.98q^{-2}-1.42q^{-3}+0.248q^{-4}]y_t=[0.14-0.224q^{-1}+0.112q^{-2}]u_t+[0.0096-0.00384q^{-1}]e_t$$

$$3) [1-2.2q^{-1}+2q^{-2}-0.8q^{-3}+0.1q^{-4}]y_t=[1-1.5q^{-1}+0.6q^{-2}]u_t+[0.0386-0.027q^{-1}]e_t$$

Los coeficientes de la parte determinística se eligieron de manera que la ganancia estática G_{DC} cumpla: $1 \leq G_{DC} \leq 1.2$.

Los coeficientes de la parte estocástica se eligieron de modo que la D.C.M. del ruido a la salida sea 0.1 si $\text{var } e_t = 1$.

Así, con condiciones similares en la entrada y las perturbaciones, la relación señal a ruido es similar en los tres casos.

Los tres sistemas se simularon excitados con un pulso cuadrado de valor 1 $1 \leq t \leq 50$, o si $51 \leq t \leq 100$.

Se concatenaron en el tiempo los tres sistemas de modo que se obtuvo un sistema variable en el tiempo, a saltos. El modelo del sistema en 1) para $1 \leq t \leq 100$, 2) para $101 \leq t \leq 200$, 3) para $201 \leq t \leq 300$.

En la figura 4.4 se observa la respuesta del sistema al tren de pulsos cuadrados ya mencionado, si $e_t = 0$.

En la figura 4.5 se observa la respuesta del sistema al mismo tren de pulsos, con una realización de $e_t \sim N(0,1)$; e_t es una sucesión de variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.).

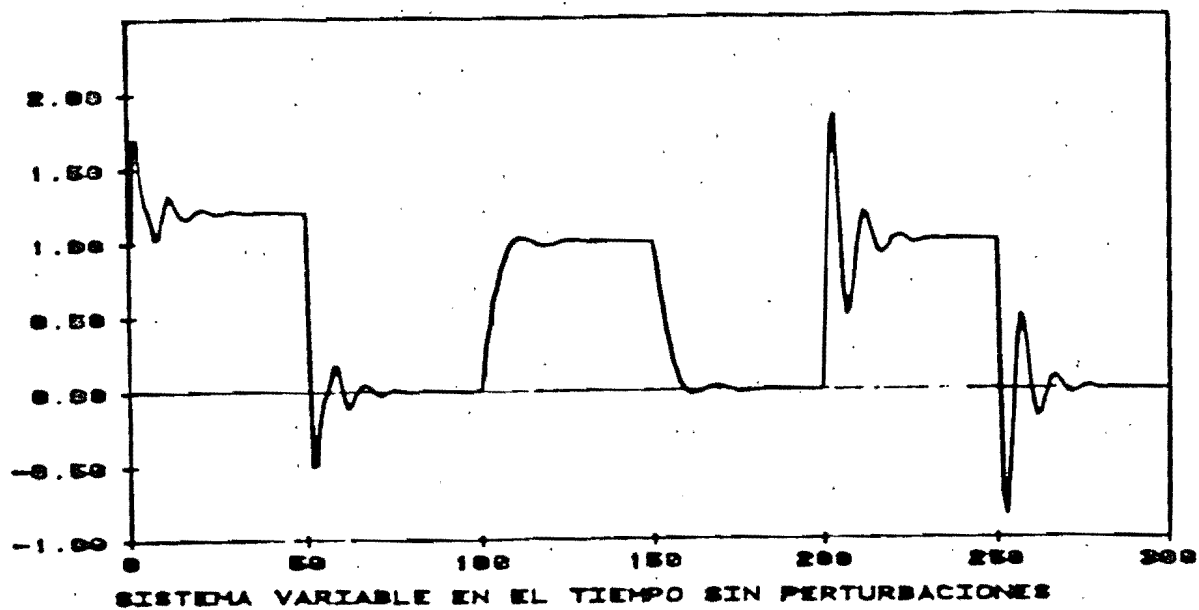


Figura 4.4.-

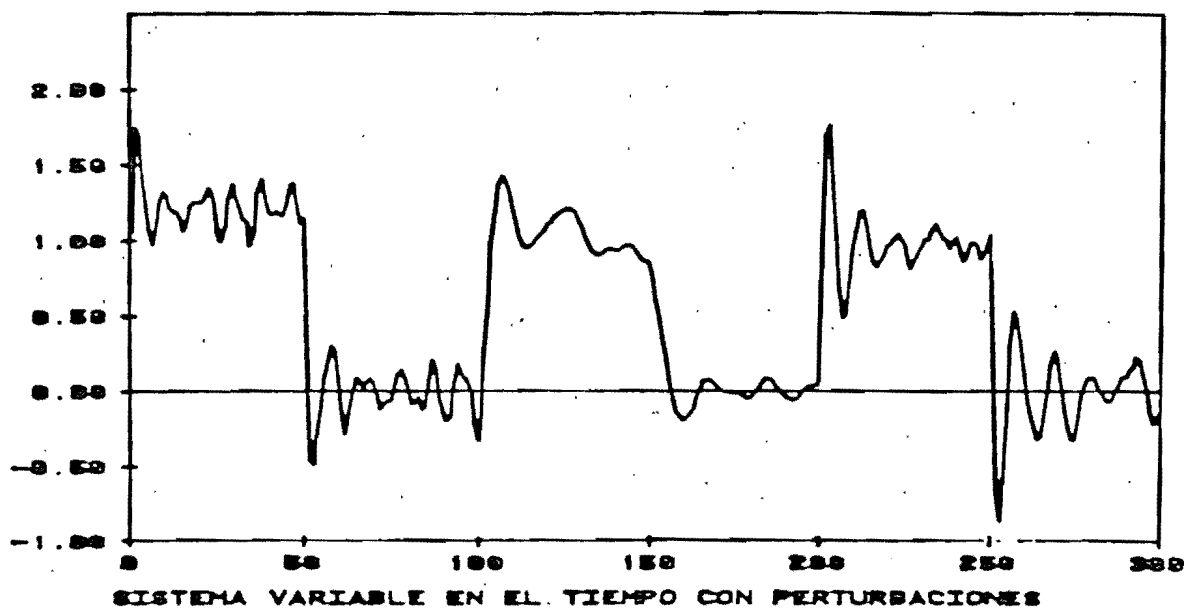


Figura 4.5 .-

Se hicieron 10 simulaciones de este sistema variable en el tiempo, con la misma entrada, y cada una con una realización independiente de e_t distribuida $N(0,1)$.

Se sometió cada simulación a los dos predictores (PMR, PMP), comparándose sólo las predicciones a 4 pasos: $\hat{y}(t/t-4)$ y suponiendo conocidas las entradas hasta el instante t .

La identificación de ambos predictores se hizo para tres valores del factor de olvido: $\lambda=0.975$, $\lambda=0.98$, $\lambda=0.985$.

Presentaremos algunos resultados de las predicciones obtenidas con $\lambda=0.98$. Los dos predictores se inicializaron de idéntica manera, con $\hat{\theta}_0=0$, $P_0=10 I$, y un factor de olvido que crece exponencialmente desde $\lambda_0=0.3$ a $\lambda_\infty=0.98$.

$$\lambda_t = 0.92 \lambda_{t-1} + 0.0784 \quad (4.23)$$

La figura (4.6) muestra el comportamiento de PMR para una de las diez realizaciones, y la figura (4.7) el comportamiento de PMP para la misma realización, en las mismas condiciones. En ambas se dibuja la predicción a 4 pasos (línea delgada) contra la salida real (línea gruesa).

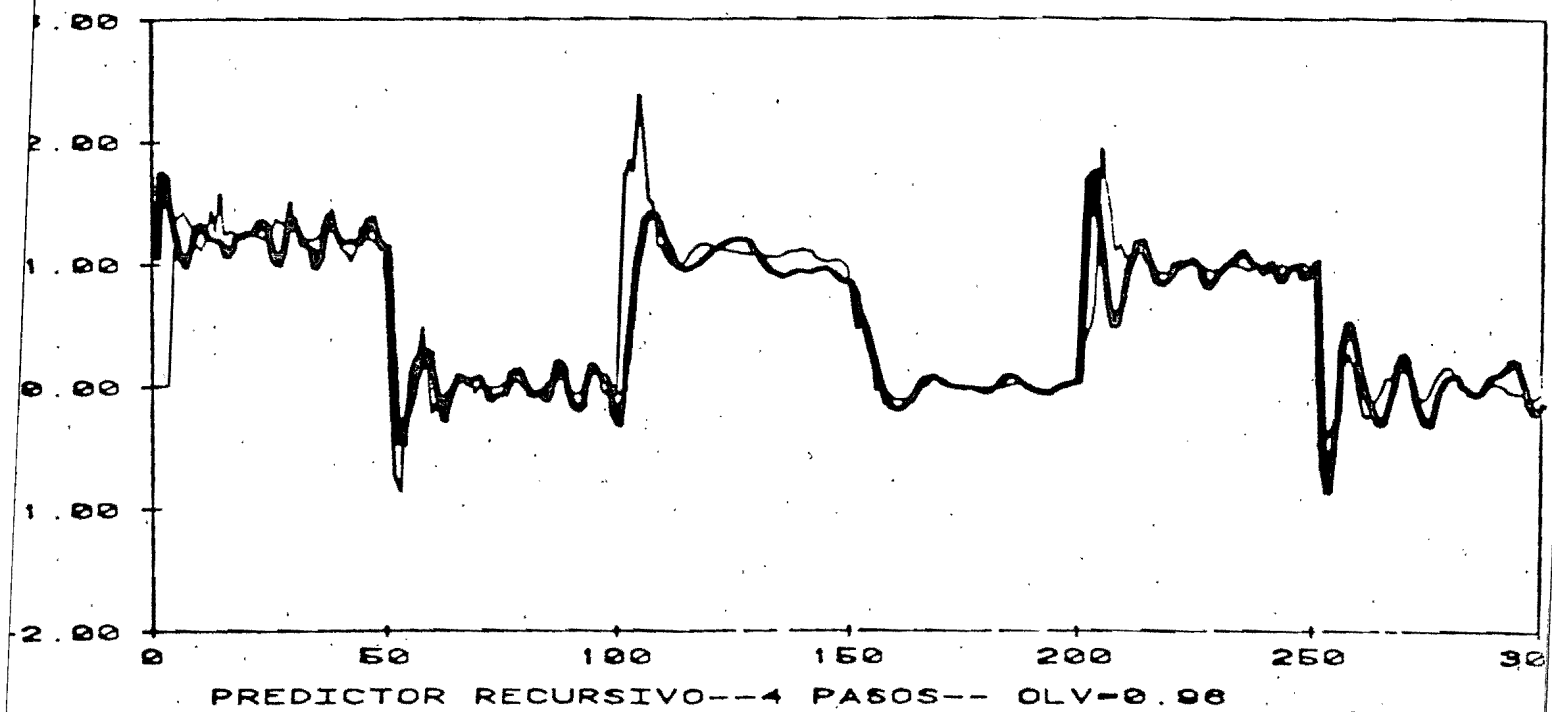


Figura 4.6

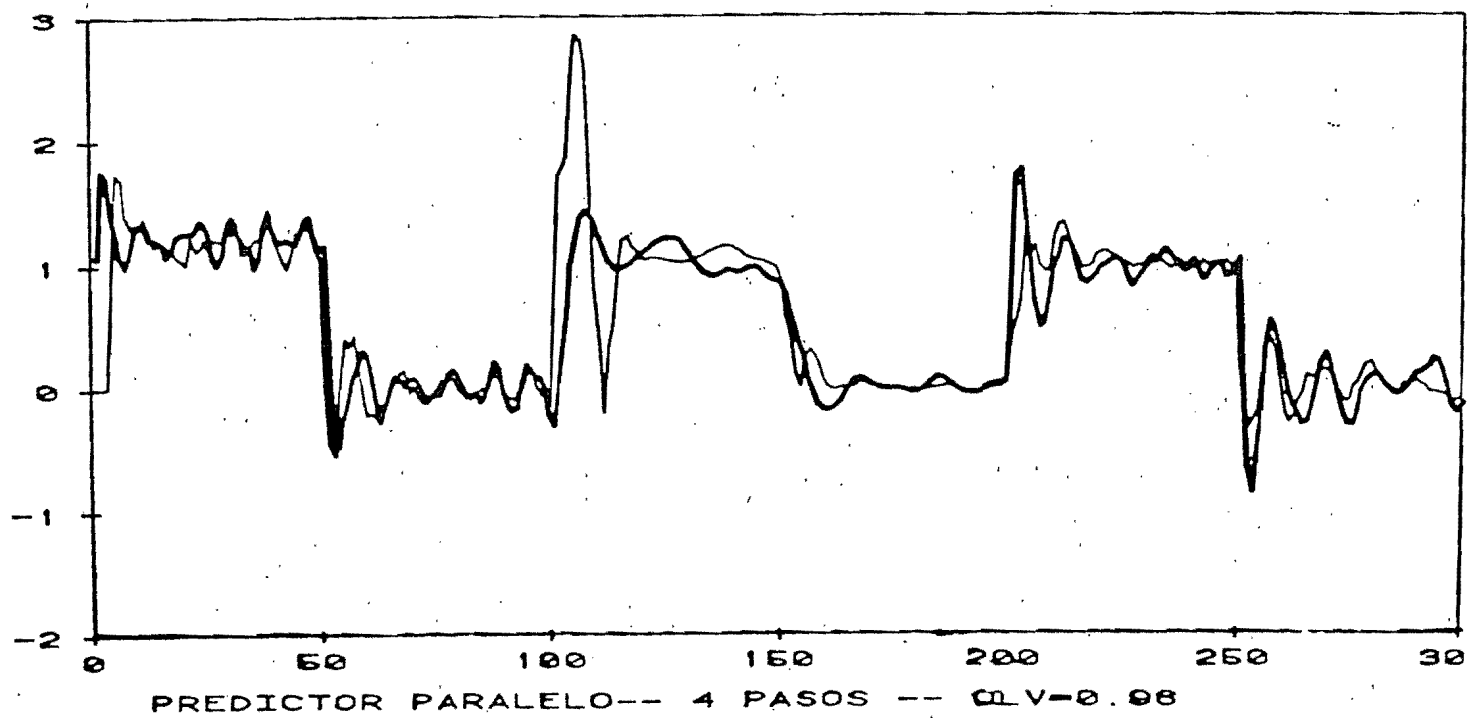


Figura 4.7

De la observación de estas figuras resulta aparente un mejor comportamiento transitorio de PMR sobre PMP.

Para tener un resultado más concluyente se hizo algo análogo a las pruebas sobre sistemas sin variables exógenas. Se estimó la desviación cuadrática media del predictor a cuatro pasos, promediando los errores de predicción sobre 10 realizaciones independientes. El cálculo se hizo con (4.21).

Los resultados se ilustran en las figuras (4.8) y (4.9) en distintas escalas.

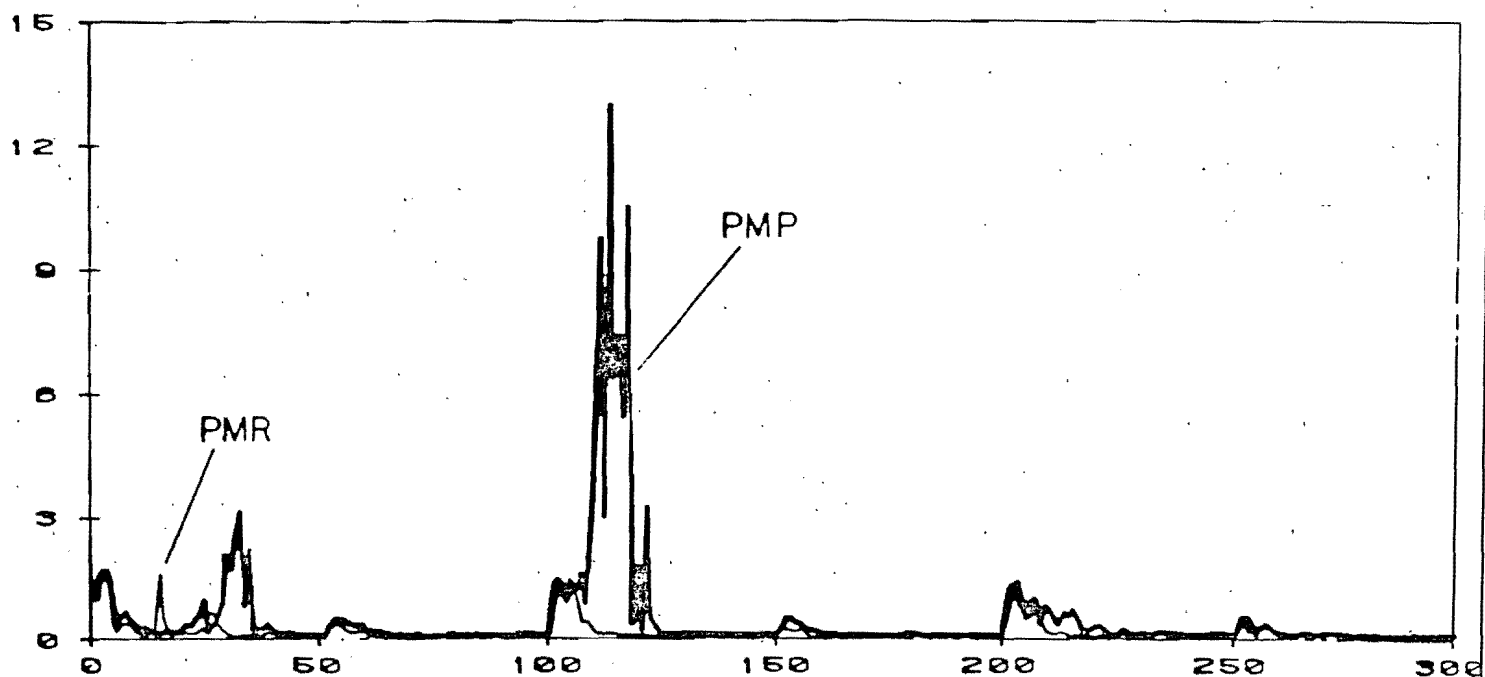


Figura 4.8 .-

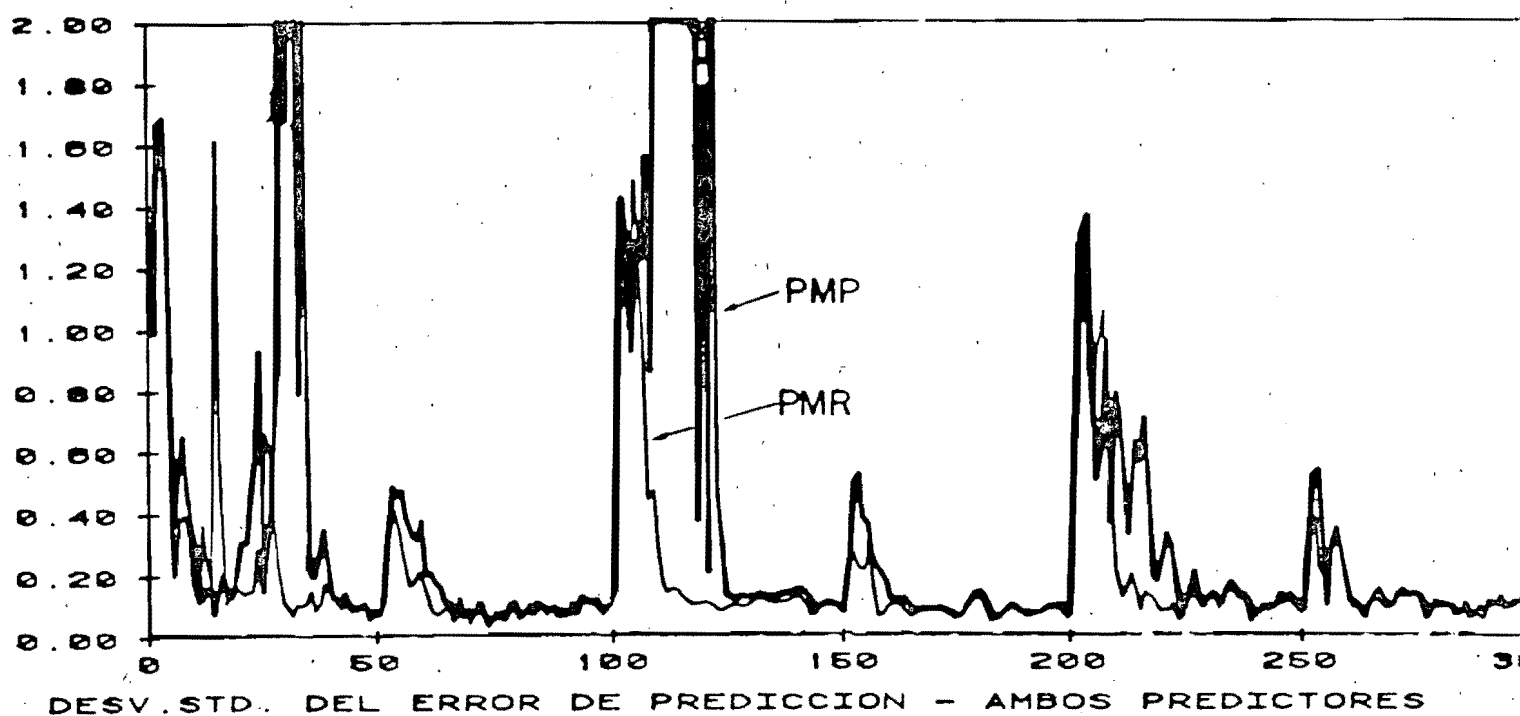


Figura 4.9

De ellos es posible extraer las siguientes observaciones:

- El predictor recursivo converge más rápido, tanto en lo que se refiere a la amplitud de los transitorios como a su duración.
- Ambos predictores tienden a dar la misma predicción si el sistema se mantiene invariable en el tiempo; este límite común coincide con el predictor óptimo en todas las pruebas.
- Como la entrada se mantuvo constante por intervalos, el aprendizaje sobre la parte del predictor que le corresponde, se produce alrededor de los saltos ($t = 0, 50, 100, 150 \dots$).

A esto se debe que ambos predictores tengan aumentos de los errores de predicción en los instantes de discontinuidad de la entrada.

- A cada modelo de los tres concatenados, le corresponden dos de estas discontinuidades. En la segunda discontinuidad ($t=50,150,250$) los errores de predicción son bastante menores que en la primer discontinuidad ($t=0,100,200$).

En términos intuitivos, se podría decir que al predictor le resta menos por "aprender" en la segunda discontinuidad que en la primera.

4.5 Adaptabilidad del predictor recursivo multipasos

4.5.1 Sistema sin perturbaciones

Como se mencionó al final de la sección 4.3, el predictor recursivo multipasos, en su versión autoadaptativa, admite distintas formas de "seguimiento" de sistemas variables en el tiempo mediante la incorporación de algoritmos de identificación con "olvido" para determinar los parámetros del predictor.

Esta propiedad fue verificada con algunas pruebas sobre datos simulados. En primer lugar veremos pruebas sobre sistemas determinísticos, sin perturbaciones.

Se simuló el mismo sistema variable en el tiempo que en la sección anterior, excitado por la misma entrada, pero sin perturbaciones.

Sobre las señales simuladas, se aplicó el predictor recursivo a cuatro pasos, con dos esquemas de adaptación en la identificación de los parámetros: el de cantidad de información de Fortescue et. al. y el de factor de olvido constante.

En las figuras (4.10), (4.11) y (4.12) se observa el predictor a cuatro pasos con el algoritmo de cantidad de información constante (en línea delgada), para tres valores distintos de la cantidad de información ($\Sigma_0 = 0.0001, 0.0003, 0.005$ respectivamente), contra el valor real de la salida, en línea gruesa. En los tres casos se eligió como valor mínimo del factor de olvido: $\lambda_{\text{mín}} = 0.92$. Los valores del factor de olvido instantáneos se grafican contra el tiempo en las mismas figuras.

Pueden observarse claramente los períodos de aprendizaje que siguen a los instantes $t=50, 100, 150, 200, 250$. En las gráficas de la predicción se destacan como períodos con fuertes transitorios, y en las gráficas del factor de olvido, como disminuciones drásticas de éste, hasta el valor mínimo 0.92.

Resulta evidente como el valor de Σ_0 modula la velocidad de convergencia.

PREDICTOR

OLVIDO

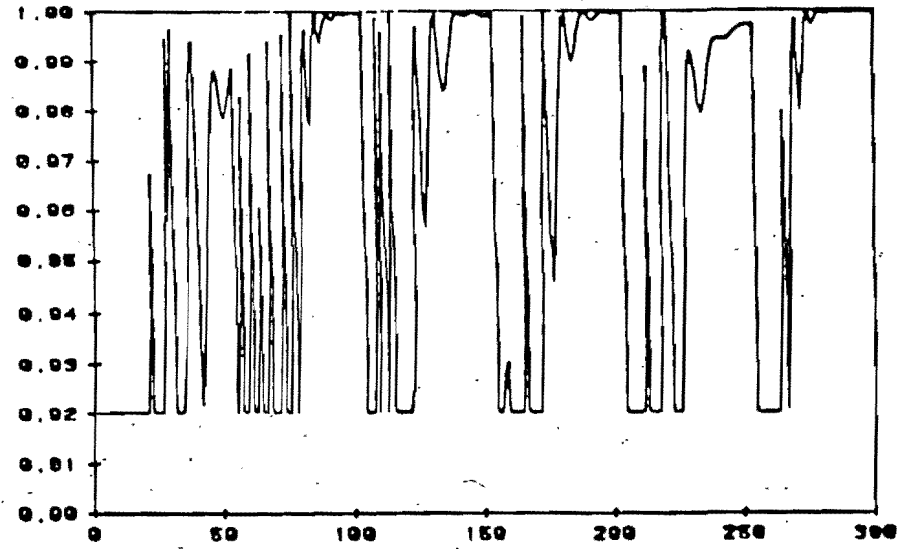
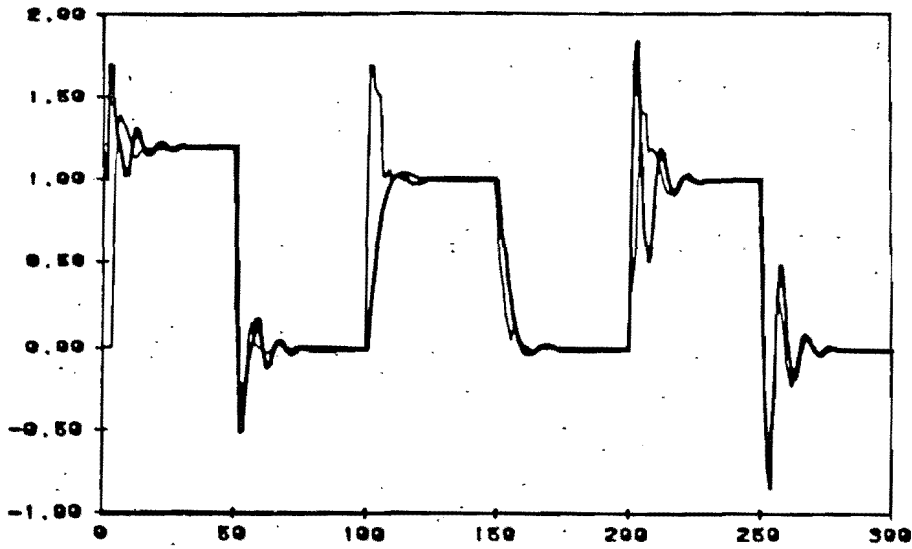


Figura 4.10-Predictor recursivo a 4 pasos, información constante $\Sigma_0=0.0001$.-

PREDICTOR

OLVIDO

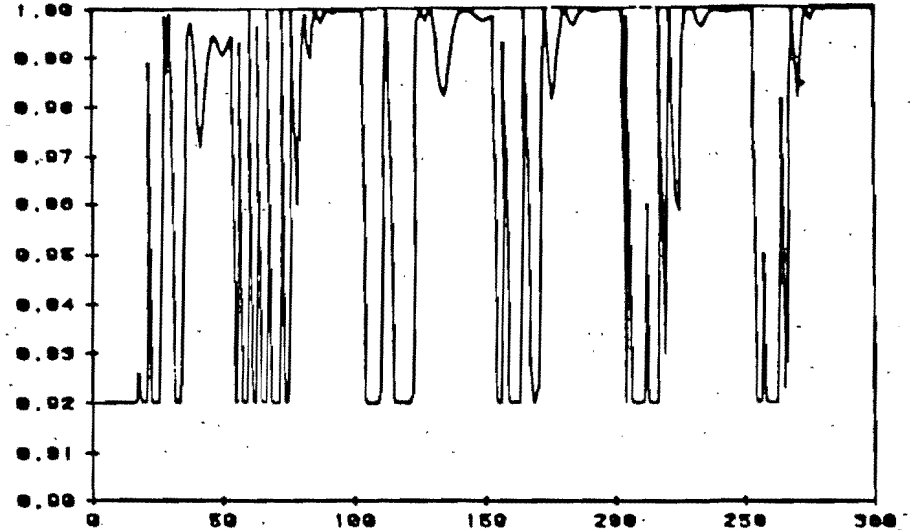
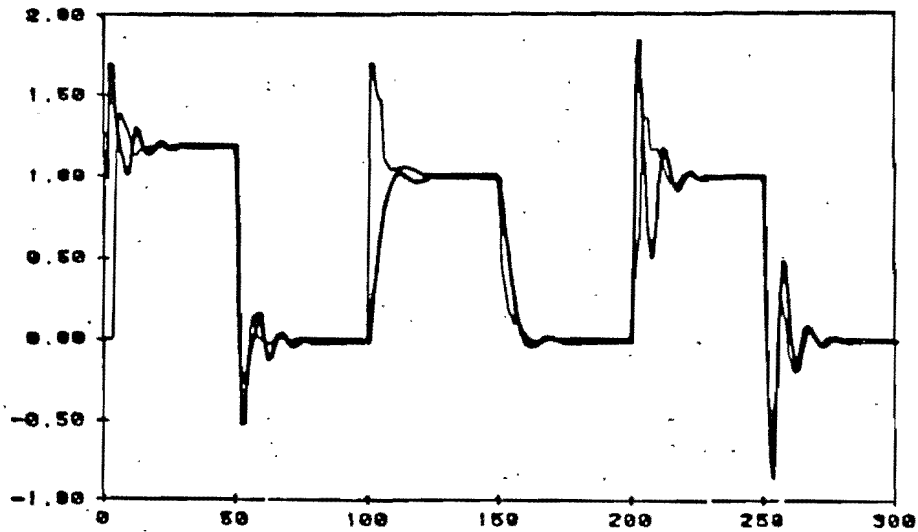


Figura 4.11-Predictor recursivo a 4 pasos, información constante $\Sigma_0=0.0003$.-

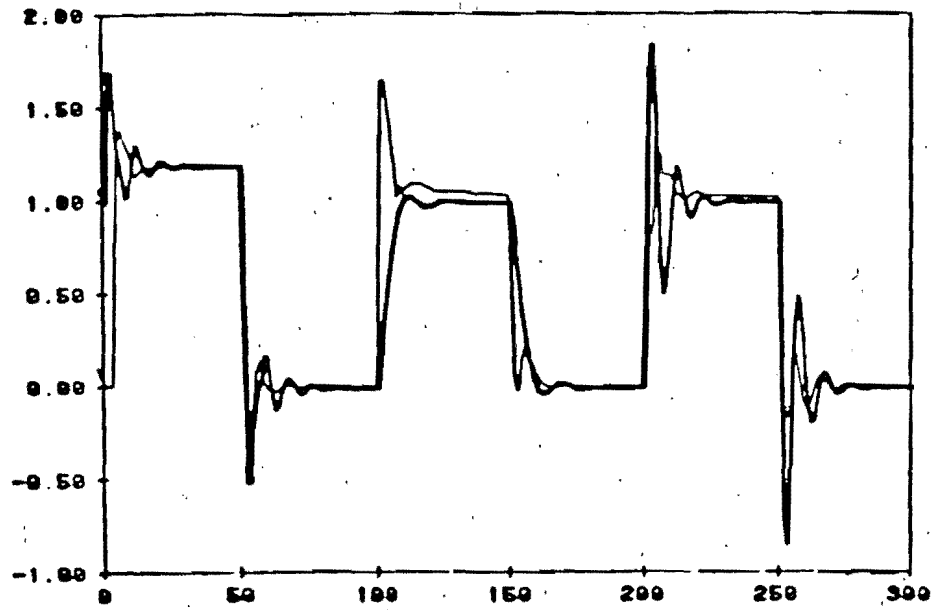
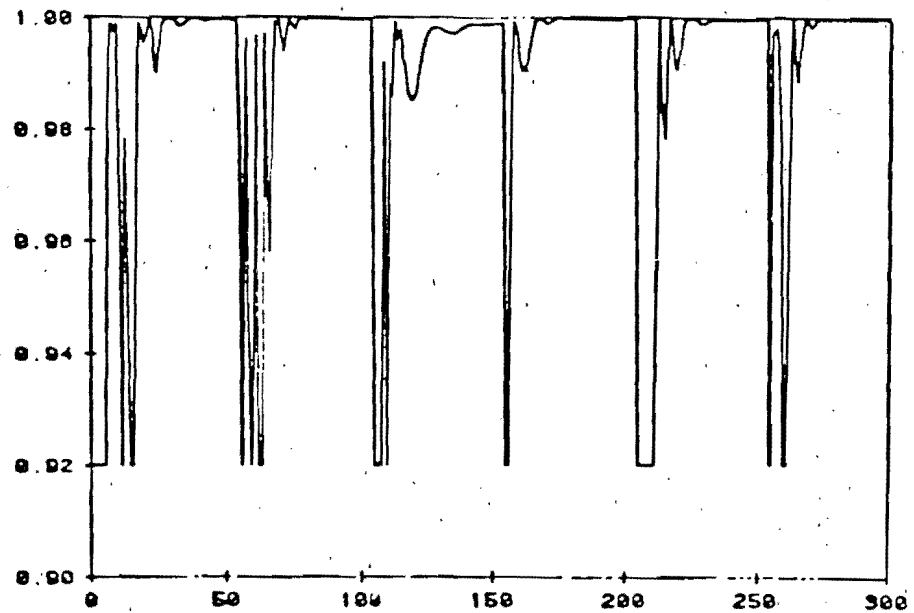
PREDICTORFACTOR DE OLVIDO

Figura 4.12-Predictor recursivo a 4 pasos; información constante. $\Sigma_0=0.005$

En efecto, se observa como valores más pequeños de Σ_0 fuerzan a λ_t a ser más pequeño reduciendo la "ventana efectiva", que es una buena medida del número de instantes de muestreo necesarios para la adaptación.

En el caso determinístico, un valor sumamente pequeño de Σ_0 es naturalmente lo indicado. En la figura (4.10) se puede ver la velocidad con la que se adapta el predictor, para $\Sigma_0 = 10^{-4}$.

A modo de comparación, en la figura (4.13) se grafica el resultado de aplicar el mismo predictor con factor de olvido constante $\lambda = 0.98$.

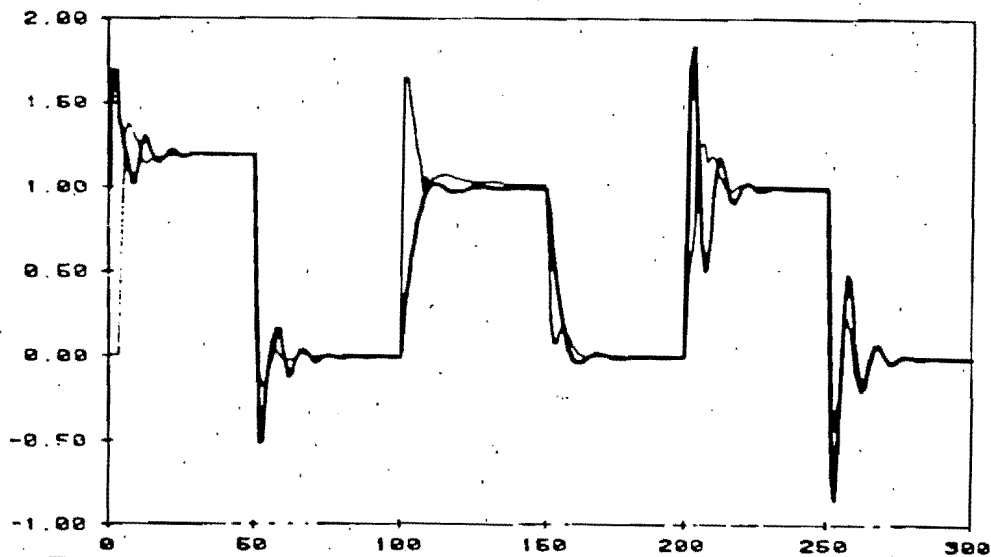


FIG.(4.13) Predictor recursivo a 4 pasos, con factor de olvido constante $\lambda = 0.98$.

La comparación de este resultado con las figuras (4.10) y (4.11), muestra las ventajas del algoritmo con olvido variable:

Este algoritmo puede combinar (seleccionando adecuadamente ϵ_0 y λ_{\min}), una rápida velocidad de adaptación cuando es necesario, con una tendencia de $\lambda_t \rightarrow 1$ cuando la adaptación se ha realizado.

Por ejemplo, los resultados ($\lambda = 0.98$) de la figura (4.13) son similares a los de la figura (4.11), con $\lambda = 3 \times 10^{-4}$. Pero en este último caso $\lambda_t \rightarrow 1$ luego de la adaptación, en lugar de permanecer constante como es el caso de $\lambda = 0.98$.

La propiedad del algoritmo de cantidad de información constante, de hacer $\lambda_t \rightarrow 1$ si $\epsilon_t \rightarrow \phi$, conduce a "extender" la ventana efectiva, conservando la información de las observaciones

pasadas.

Este hecho no le dá inmunidad absoluta pero sí mayor robustez a posibles problemas de "blow-up", en situaciones como la de este ejemplo.

4.5.2 Sistema con perturbaciones

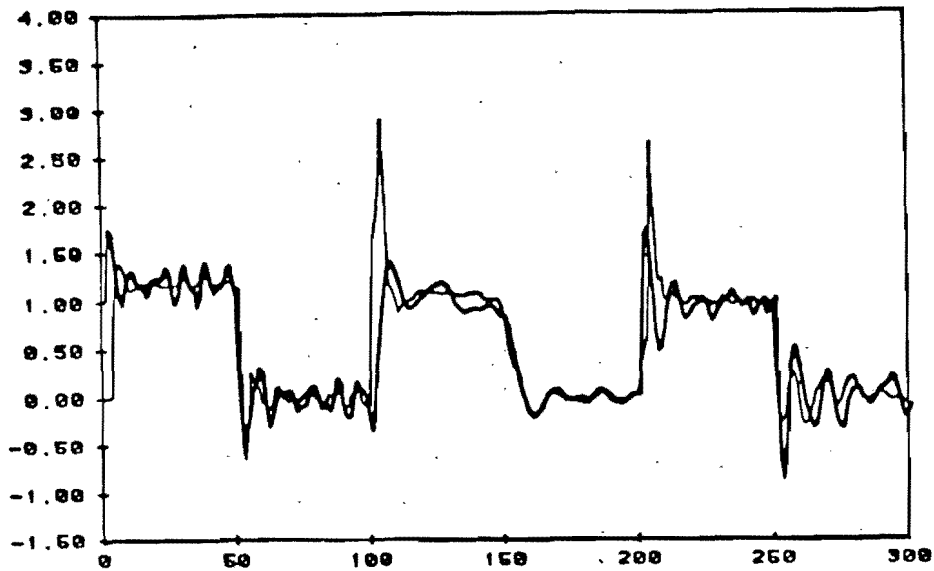
Se simuló el mismo sistema con perturbaciones aleatorias, en las mismas condiciones de la sección 4.4.

Se aplicó el predictor recursivo a 4 pasos. Las figuras (4.14) y (4.15) muestran el predictor con información constante ($\sigma_0 = 0.01$ y $\sigma_0 = 0.1$, respectivamente), en línea delgada, contra la salida real en línea gruesa. También se muestra el factor de olvido contra el tiempo. En ambos casos se eligió $\lambda_{\min} = 0.92$. Puede apreciarse como σ_0 modula la velocidad de convergencia. Por las características de la perturbación aleatoria (espectro) ésto es más apreciable en la sección de datos correspondiente a $t \in [100, 199]$.

En ese período el sistema actúa como un pasa-bajos, y la perturbación es esencialmente de baja frecuencia y por lo tanto es más predecible. La varianzá del error de predicción es por lo tanto menor, y son más apreciables los aumentos o disminuciones de dicha varianzá.

Se puede comparar otra vez este resultado contra el resultado

PREDICTOR



OLVIDO

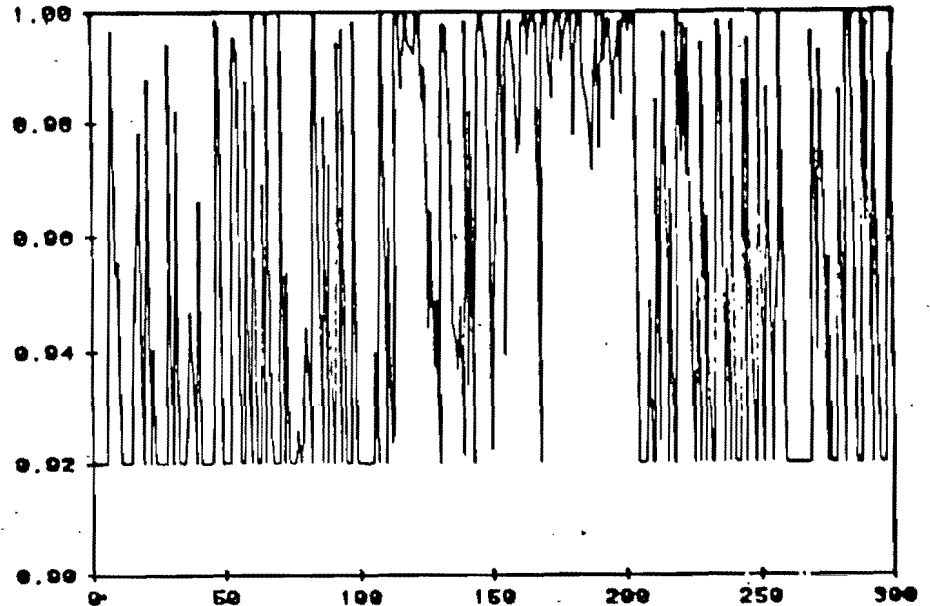
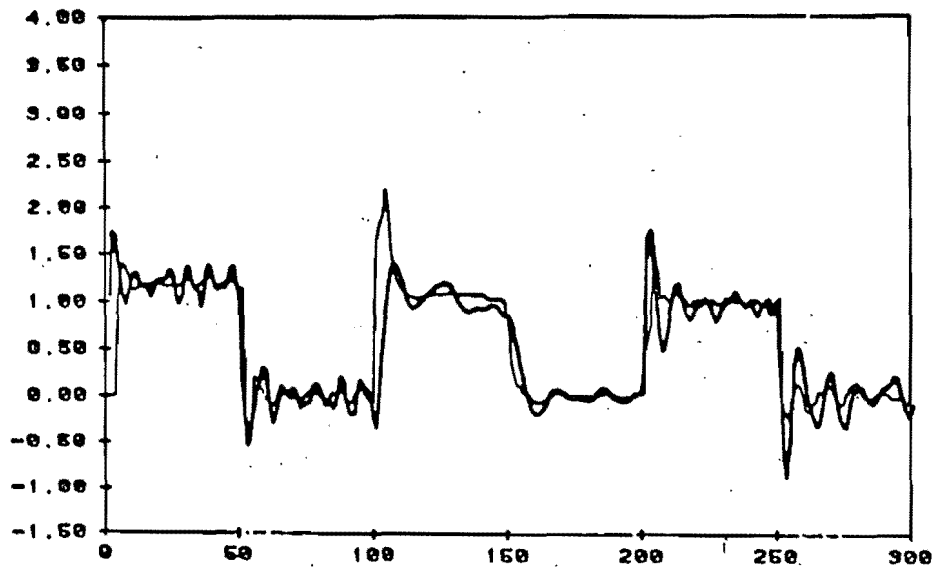


Figura 4.14-Predicador recursivo a 4 pasos, información constante $\Sigma_o=0.01$.

PREDICTOR



OLVIDO

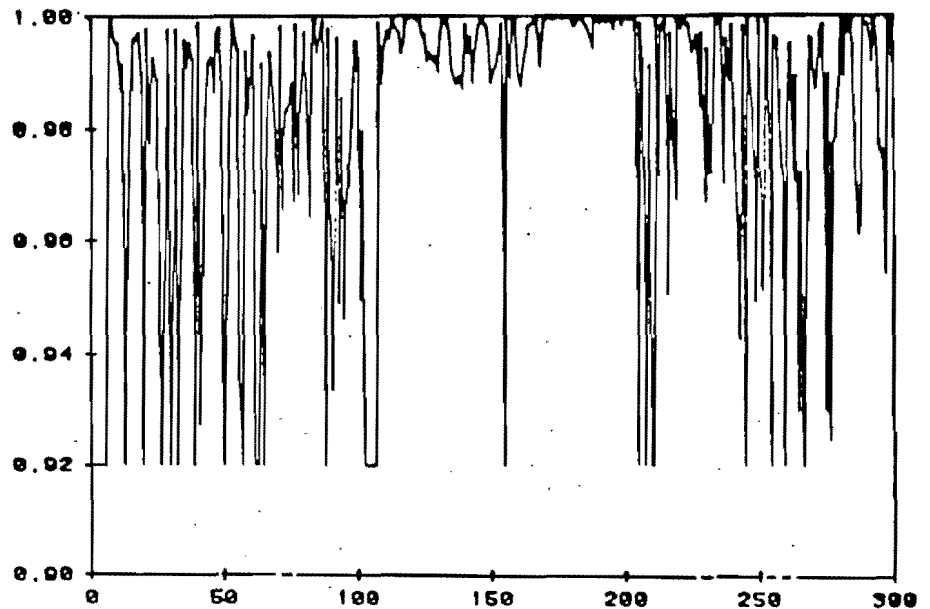


Figura 4.15-Predicador recursivo a 4 pasos, información constante $\Sigma_o=0.1$.

obtenido con factor de olvido constante.

La figura (4.16) muestra el predictor a 4 pasos (línea delgada) contra la salida real (línea gruesa) con $\lambda = \text{cte}$, para $\lambda = 0.92$ y $\lambda = 0.98$.

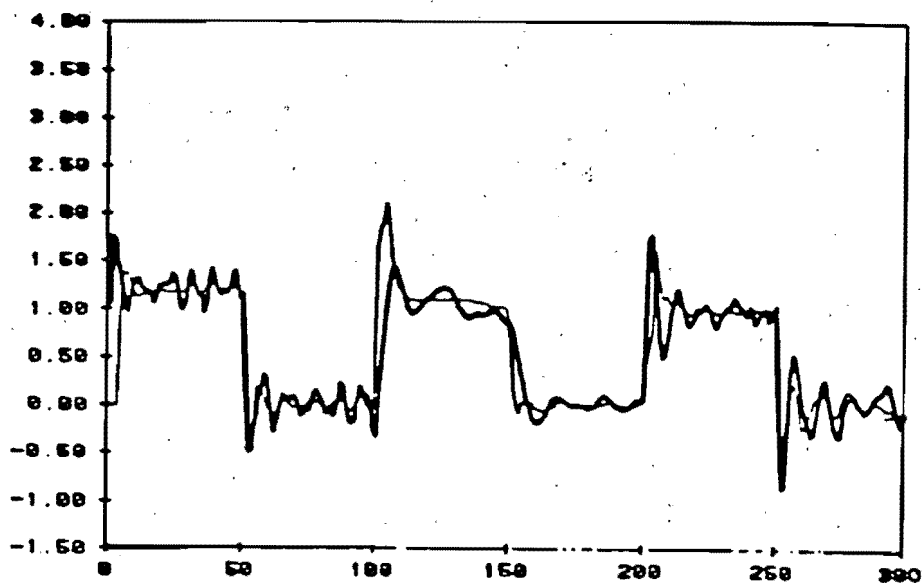
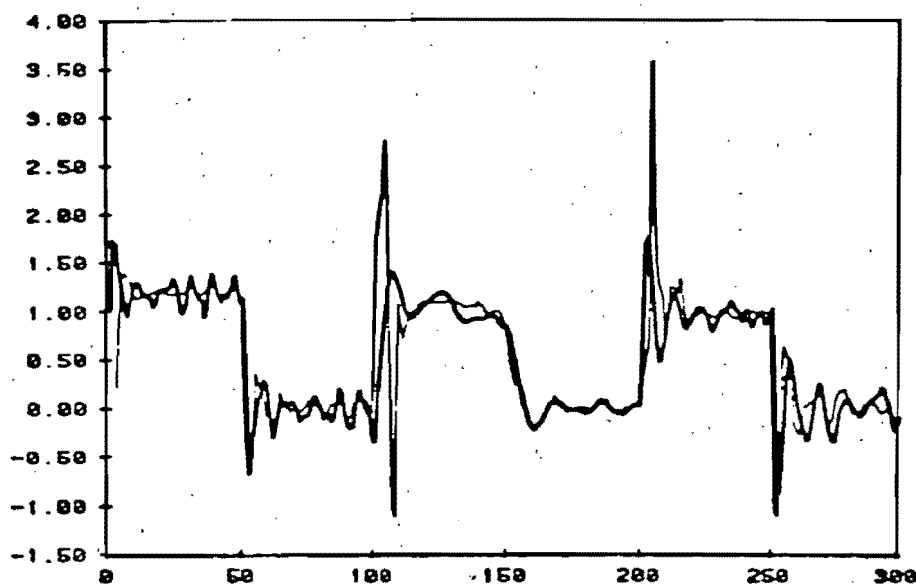
Se puede apreciar que a menor valor de λ , mayor es la velocidad de adaptación, pero también es mayor la inestabilidad (obsérvense los transitorios con $\lambda = 0.92$).

También en este caso se concluye la ventaja del algoritmo de información constante frente al de factor de olvido constante, en el sentido de que el primer algoritmo disminuye sustancialmente el factor de olvido (mejorando la velocidad de adaptación), sólo cuando se detectan grandes errores de predicción.

4.6 Conclusiones

Se ha mostrado que la estructura del predictor multipasos óptimo, se puede reducir a la más simple: el de un paso. También que la propiedad de autoadaptabilidad (en el caso de parámetros desconocidos) del predictor a un paso, se hereda al predictor multipasos recursivo.

La forma de adaptación: identificación recursiva por mínimos cuadrados, permite la incorporación del factor de olvido para realizar la predicción autoadaptativa de sistemas variables

a) $\lambda = 0.98$ b) $\lambda = 0.92$ Figura 4.16-Predicador recursivo a 4 pasos con factor de olvido λ constante.

en el tiempo.

Mediante extensas simulaciones, se verificó esta propiedad con el método descrito en la sección 3.2 (factor de olvido constante), así como también se comprobó la afirmación intuitiva de que el esquema de predicción recursiva no sólo posee sobre el predictor de largo fijo (PMP) las ventajas de mayor sencillez y rapidez de cálculo, sino también mayor velocidad de convergencia.

Finalmente se verificó a través de simulaciones la posibilidad de combinar el esquema de predicción con otros esquemas de identificación.

En particular con el esquema de cantidad de información constante. Las ventajas de este método de identificación se heredan al predictor autoajutable, como ha podido verse en comparación con el método de factor de olvido constante.

CAPITULO 5.

Aplicación a la predicción de caudales en cuencas hidrológicas.

La predicción del caudal de un río con base en las observaciones de las precipitaciones en su cuenca y el caudal de sus afluentes es un problema típico de predicción autoadaptiva porque:

- Existe un retardo implícito entre entradas y salidas, que permite la "predicción" de las últimas.
- Es sumamente difícil de modelar en forma precisa ya que se trata de un complejo sistema no lineal, de parámetros distribuidos, con entradas distribuidas.
- Es variable en el tiempo y por lo tanto es necesario un me-

canismo de adaptación.

Utilidad:

La importancia de la predicción de caudales resulta de su uso en:

* Control de otros sistemas:

- Diseño de estrategias óptimas de acumulación de aguas para riego.
- Diseño de estrategias óptimas de generación de energía eléctrica en centrales hidroeléctricas.

* Toma de decisiones:

- Diseño de políticas de embalse, protección contra la erosión.
- Previsión de catástrofes.

- * Las aplicaciones anteriores se jerarquizan en un país como México, que tiene largas y diferenciadas temporadas de lluvia y sequía.

5.1 Modelo simplificado de una cuenca

Consideraremos un modelo sùmamamente simplificado bajo las hipótesis:

- 1) Entradas y salidas concentradas.
- 2) Escurrimiento superficial e infiltración subterránea linea

les.

3) Pérdidas concentradas.

4) No hay afluentes.

El desarrollo que sigue está basado en trabajos realizados en el Instituto de Ingeniería, ver p. ej. ESPAÑA, DOMINGUEZ, FLORES [24].

La figura 5.1 ilustra un esquema físico y un diagrama de bloques del modelo.

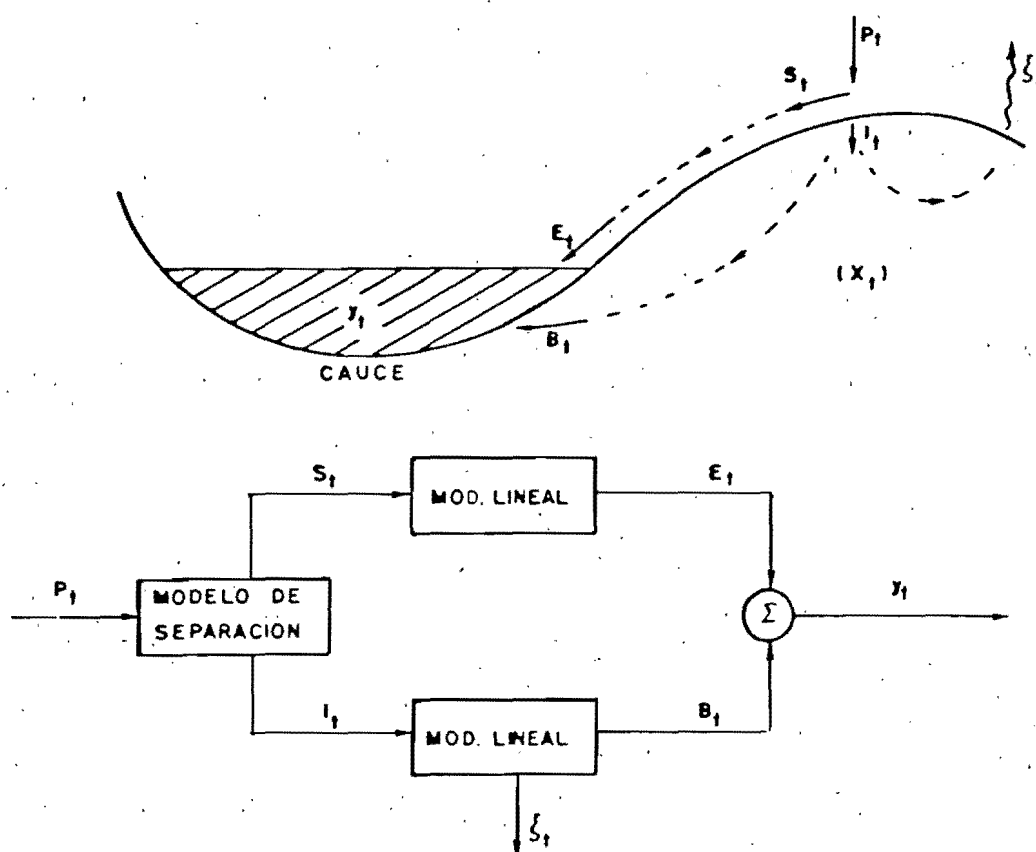


FIGURA 5.1 Modelo de la cuenca.

Las variables de interés son:

P_t : precipitación [mm]/ Δt

S_t : parte de la precipitación que escurre superficialmente (precipitación efectiva) [mm]/ Δt

I_t : parte de la precipitación que se infiltra [mm]/ Δt

ξ_t : pérdidas por evapo-transpiración [mm]/ Δt

X_t : estado de humedad de la cuenca en [mm] equivalentes

E_t : contribución al caudal por escurrimiento superficial [mm]/ Δt

B_t : contribución al caudal por infiltración (flujo base) [mm]/ Δt

Y_t : gasto total medido en una sección determinada, al final de la cuenca [mm]/ Δt

Debe notarse que todos los volúmenes están medidos en mm equivalentes cada Δt (período de muestreo). Cada mm equivalente, corresponde a un volumen:

$$V = 10^3 A$$

donde V se mide en m^3 , y A es el área de captación de la cuenca en km^2 .

Relación entre las variables de interés (sub-modelos discretos):

1. Separación

$$P_t = S_t + I_t \quad (5.1)$$

2. Flujo base

La relación entre I_t y B_t se modela linealmente, como un sistema de dos tanques con orificio, unidos mediante una línea de transporte con retardo. Ver fig 5.2 .

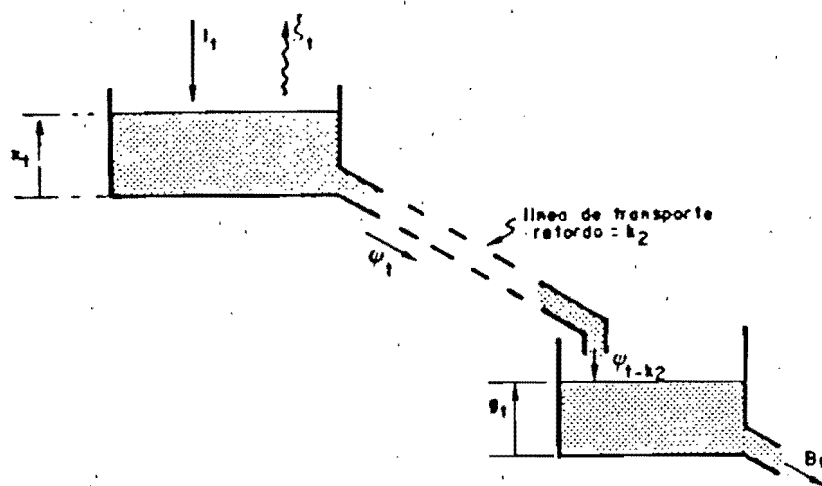


FIG 5.2 Modelo físico del gasto base

Se supone

$$\psi_t = \alpha X_t \text{ [mm}/\Delta t], \alpha > 0 \quad (5.2)$$

$$\xi_t = \beta X_t \text{ [mm}/\Delta t], \beta > 0 \quad (5.3)$$

balance de masa:

$$X_{t+1} = X_t - \psi_t - \xi_t + I_t \quad (5.4)$$

$$X_{t+1} = \gamma X_t + I_t \quad (5.5)$$

donde: $\gamma = 1 - \alpha - \beta$ es la fracción de X_t evapotranspirada o cedida al tanque 2 que supondremos de ahora en adelante $\gamma < 1$.

La variable X_t [mm] representa el estado de humedad de la cuenca en el instante t . La ecuación (5.5) describe la evolución del estado.

En el segundo tanque:

$$B_t = \delta g_t \quad \delta > 0 \quad (5.6)$$

$$g_{t+1} = g_t + \psi_{t-k_2} - B_t \quad (5.7)$$

$$g_{t+1} = \epsilon g_t + \alpha X_{t-k_2} \quad (5.8)$$

donde $\epsilon = 1 - \delta$

Entonces, de (5.5), (5.2), (5.6) y (5.8), la función de transferencia descrita entre I_t y B_t resulta:

$$B_t = \frac{q^{-k_3} c}{(1 - \gamma q^{-1})(1 - \epsilon q^{-1})} I_t \quad (5.9)$$

donde $k_3 = k_2 + 1$

$$c = \alpha \delta$$

3. Gasto superficial

También acá se supone una relación lineal entre E_t y S_t , de la forma:

$$E_t = \frac{B(q^{-1})}{A(q^{-1})} S_t \quad (5.10)$$

4. Modelo de separación

La separación de la precipitación en dos partes: una que se infiltra y otra que escurre superficialmente corresponde a un fenómeno que es bastante complejo y de características no lineales.

Cualquier modelo debe tener en cuenta que cuanto más húmeda está la cuenca, menor es la parte de la precipitación que se infiltra, y mayor la que escurre superficialmente.

Se han considerado varios modelos simplificados ver [24], el que adoptaremos es:

$$\begin{cases} I_t = P_t \phi_t & (0 \leq \phi_t \leq 1) & (5.11) \\ \phi_t = e^{-k X_t} & & (5.12) \end{cases}$$

donde k es una constante, y X_t es el estado de humedad de la cuenca. Nótese que si se aproxima la exponencial de (5.12) con una función lineal, el coeficiente k tiene la interpretación física de $\frac{1}{X_{\text{máx}}}$ donde $X_{\text{máx}}$ es la altura máxima del primer tanque de la fig 5.2. Esto es la humedad máxima de la cuenca. Entonces la infiltración sería proporcional al porcentaje de espacio que queda libre en ese tanque.

5. Gasto total

$$Y_t = E_t + B_t$$

de (5.13), (5.9), (5.10):

$$Y_t = \frac{B(q^{-1})}{A(q^{-1})} S_t + \frac{q^{-k_s} c}{(1-\gamma q^{-1})(1-\delta q^{-1})} I_t \quad (5.14)$$

Denominando $\frac{c(q^{-1})}{D(q^{-1})}$ a la función de transferencia entre I_t y B_t , y considerando el modelo (5.11) de separación resulta de (5.14):

$$Y_t = \underbrace{\frac{B(q^{-1})}{A(q^{-1})} P_t (1-\phi_t)}_1 + \underbrace{\frac{c(q^{-1})}{D(q^{-1})} P_t \phi_t}_2 \quad (5.15)$$

La estructura del modelo (5.15) permite concluir que:

(a) Para $\phi_t \rightarrow 0$ ($x_t \rightarrow \infty$) (grandes precipitaciones, en período prolongado), el modelo (5.15) tiende a la parte 1, lineal en la precipitación P_t .

Esta parte es la debida sólo al escurrimiento superficial. Se puede entonces trabajar con un modelo lineal para grandes avenidas, debidas a grandes precipitaciones.

a') $\phi(t) \rightarrow 1$ ($x(t) \rightarrow 0$), cuenca muy seca, sólo cuenta la parte 2 que es la contribución del flujo base al escurrimiento total.

(b) Si se pueden determinar las constantes k , γ fuera de lí-

nea, en un período de calibración del modelo; entonces con ellas se puede tener una estimación de ϕ_t suponiendo que k , γ se mantienen constantes luego del período de calibración. (Vease que supuesto $\gamma < 1$ en la ec 5.5 el modelo tiende a "olvidar" la condición inicial por lo que no resulta crítico conocer con exactitud x_0).

Conocida ϕ_t se puede plantear el problema como uno de predicción lineal de un sistema lineal de dos entradas: $P_t(1-\phi_t)$ y $P_t\phi_t$.

- (c) Se obtiene un enfoque intermedio, extendiendo el rango de aplicación de (a') como sigue:

Si se pudiera tener "a grosso modo" una estimación del valor de la parte 2 de (5.15) (gasto base), a la que llamaremos \bar{B}_t , entonces denominando

$$\bar{E}_t = Y_t - \bar{B}_t = \underbrace{\frac{B(q^{-1})}{A(q^{-1})} P_t(1-\phi_t)}_1 + \underbrace{\left[\frac{C(q^{-1})}{D(q^{-1})} P_t\phi_t - \bar{B}_t \right]}_0 \quad (5.16)$$

Se puede trabajar con el modelo lineal 1, y dejar a cargo de la parte autoadaptiva del predictor el ajuste de un modelo lineal que aproxime la relación $P_t(1-\phi_t) + \bar{E}_t$, por el error cometido al mal estimar B_t .

- (d) Se puede intentar ajustar un modelo lineal directamente a la relación precipitación-gasto. Esto es linealizar el modelo no-lineal y seguir entonces el "modelo tangente"

(5.17).

Es importante que el predictor tenga buena velocidad de adaptación para seguir las variaciones de las condiciones de la cuenca.

$$Y_t = \frac{F(q^{-1})}{G(q^{-1})} P_t \quad (5.17)$$

Es sin embargo previsible que existan errores de medición importantes, el efecto que esto tiene en un predictor con buena velocidad de adaptación es una fuerte incidencia de los errores de medición en el error de predicción. (Ver discusión capítulos 2 y 3 sobre compromiso entre velocidad de seguimiento y precisión en la identificación).

Los modelos encontrados en la literatura tienen todos los niveles de complejidad imaginables.

Los que tratan de tomar en cuenta los efectos de infiltración y evapotranspiración, necesitan mediciones o al menos estimaciones en línea de algunas características físicas del suelo.

En otros casos, las hipótesis necesarias para el tratamiento matemático y el conocimiento incompleto de las variables físicas involucradas, dan lugar a grandes divergencias entre los períodos de calibración y de predicción.

Eso lleva a la necesidad de recalibraciones periódicas del modelo para actualizar los modelos, esencialmente no estacionarios que se obtienen.

Todos estos inconvenientes hacen pensar que los modelos tipo "caja negra", más sencillos y fáciles de tratar deben ser al menos tan eficientes como los más complicados.

E. Todini y J. Wallis [33] utilizan un modelo lineal, invariante en el tiempo, descrito por la respuesta al impulso con una ligera modificación para tomar en cuenta el efecto de la humedad. Pero esencialmente el modelo global se define fuera de línea y permanece invariante en el tiempo. Ese modelo se usa como "modelo en paralelo" para estimar caudales futuros, es decir que no se toma en cuenta, para la predicción, mediciones en tiempo real del caudal.

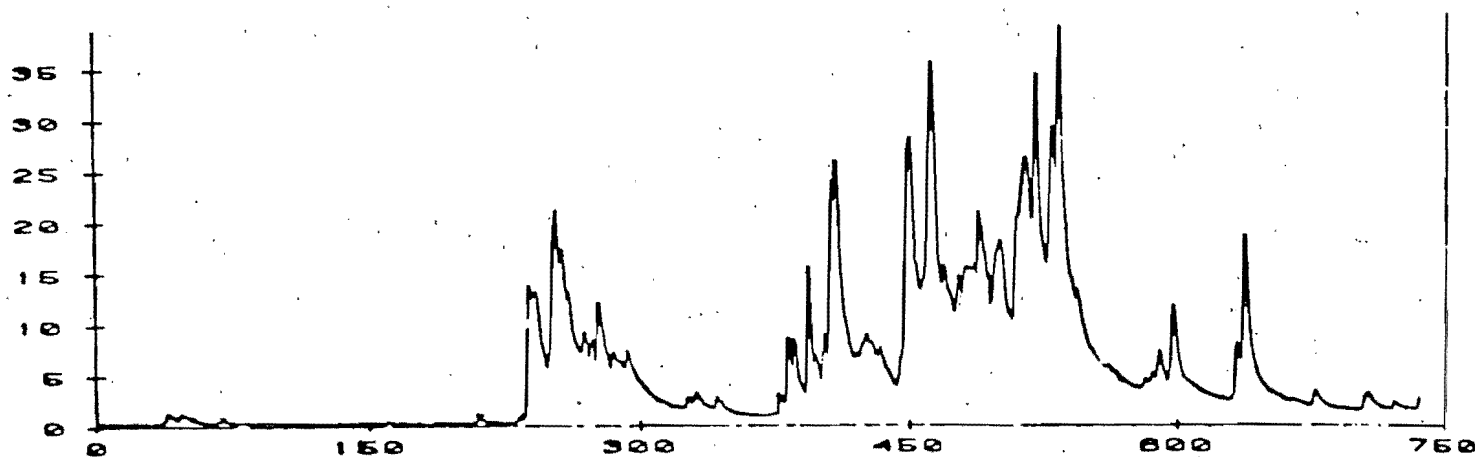
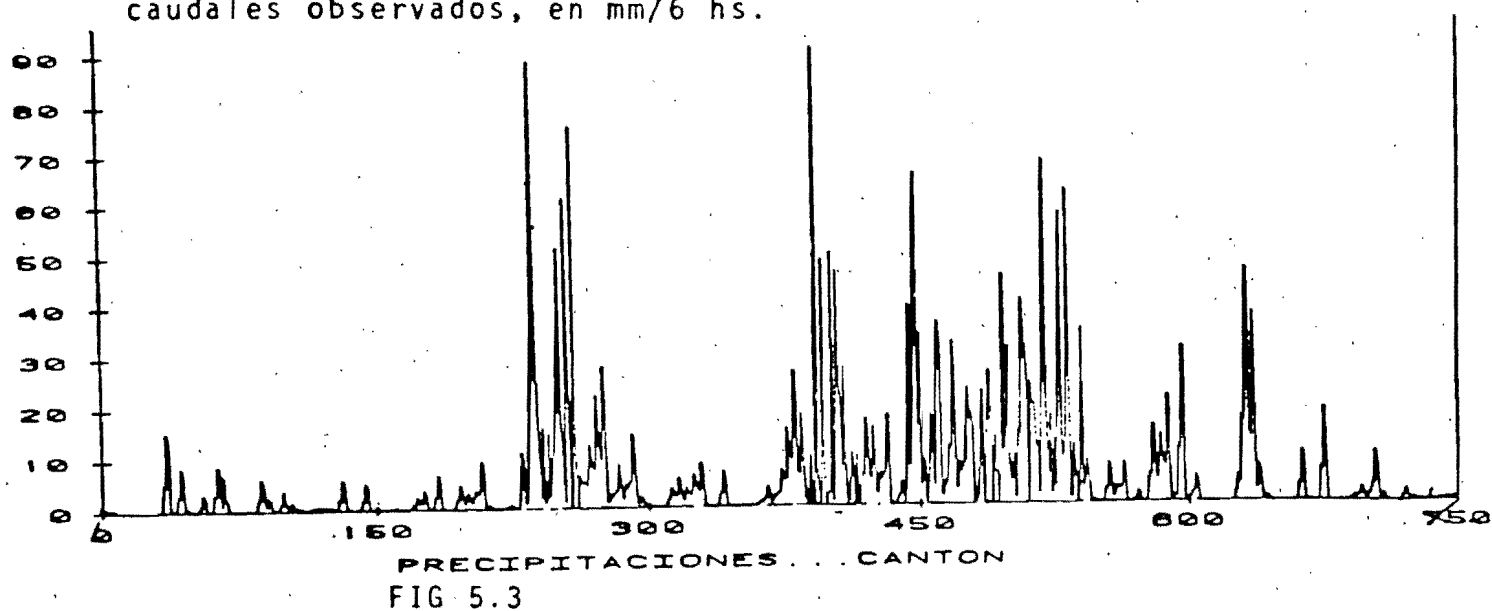
Es previsible entonces que el combinar la sencillez de las descripciones de entrada-salida de sistemas dinámicos lineales, con la capacidad de seguir un modelo variable en el tiempo, del predictor lineal autoadaptivo, dé lugar a formas más eficientes de predecir el escurrimiento de una cuenca.

5.2 Pruebas con datos reales. Cuenca de Cantón.

Se tuvieron disponibles para pruebas, observaciones a lo largo de 184 días, cada 6 horas (736 muestras), de precipitacio-

nes y caudal en una sección del río Cantón.

Las figuras (5.3) y (5.4) muestran las precipitaciones y los caudales observados, en mm/6 hs.



Veremos algunas pruebas realizadas según el enfoque (c) del inciso anterior.

Estimación del gasto base.

En este enfoque se trata de obtener una estimación del gasto base, a "grosso modo". Para ello hay que fijar un modelo de separación

$$\begin{cases} I_t = P_t \phi_t \\ \phi_t = e^{-k} x_t \end{cases}$$

y un modelo entre la precipitación infiltrada I_t y el gasto base B_t :

$$B_t = \frac{c q^{-k_3}}{(1 - \gamma q^{-1})(1 - \delta q^{-1})} I_t \quad (5.18)$$

Para lo que es necesario determinar los parámetros k , k_3 , γ , δ , c .

Para hallar γ y δ se consideraron algunos intervalos de medición en los que no había precipitaciones, por lo que $I_t = 0$.

Para ello se ajustó un modelo autoregresivo de segundo orden a las series observadas de recesión sin precipitaciones, dicho modelo corresponde al denominador de (5.18). (Ver esquema de la fig 5.5).

En todas las series observadas, el polo $\gamma=0.975$ aparece con bastante precisión, y está dentro del rango previsto por los

hidrólogos para el valor de la constante γ .

La constante de tiempo asociada es de 22.7 días. El otro polo no coincide en todas las series consideradas. Se eligió el modelo identificado en la serie que corresponde a las muestras [614,625] de la fig 5.4.

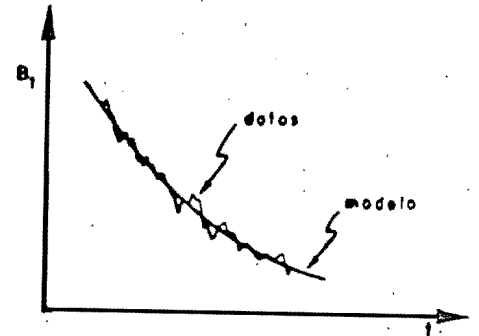


FIG 5.5

$$1 - 1.133 q^{-1} + 0.154 q^{-2} = (1 - \gamma q^{-1})(1 - \delta q^{-1})$$

En la temporada de sequía se supuso que toda la precipitación se infiltraba. Con los datos correspondientes a dicha temporada se estimó el retraso $k_3 = 4$.

Alternativamente al numerador de (5.18) propuesto: $c q^{-4}$ se probó ajustar a las observaciones otros numeradores de la forma

$$c_0 + c_1 q^{-1} + c_2 q^{-2} + c_3 q^{-3} + c_4 q^{-4} + c_5 q^{-5}$$

Con los datos de la temporada de sequía se observó que un numerador de la forma

$$c(0.4 q^{-3} + q^{-4}) \quad (5.19)$$

daba mejores resultados que cq^{-4} (el propuesto en el modelo). Queda por determinar el valor de c . Para calcularlo se estableció el balance de masa en la parte infiltrada de la precipitación. La precipitación total en la cuenca en el período observado es de 4146.8 mm. El gasto total en la cuenca es de 3885.8 mm.

Supusimos que: $\left\{ \begin{array}{l} 1) \text{ Todas las pérdidas (por evapo-transpiración) se producen en la parte infiltrada de la precipitación.} \\ 2) \text{ La precipitación infiltrada es estimativamente un 40\% del total de la precipitación.} \end{array} \right.$

Entonces:

$$\text{pérdidas} = 4146.8 - 3885.8 = 261 \text{ mm}$$

$$\text{prec. infiltrada} = 0.4 \times 4146.8 \cong 1659 \text{ mm}$$

$$\text{ganancia del modelo gasto base} \cong \frac{1659 - 261}{1659} = 84.3\%$$

$$\text{de donde } c \cong 0.0127$$

El modelo lineal para el gasto base quedó como:

$$B_t = \frac{0.005 q^{-3} + 0.0127 q^{-4}}{1 - 1.133 q^{-1} + 0.154 q^{-2}} I_t \quad (5.20)$$

Finalmente queda por determinar la constante k para el modelo de separación. Como γ se supone conocida ($\gamma=0.975$), se puede calcular x_t en función de I_t , y por lo tanto ϕ_t . Se procedió a calcular el valor de B_t que se obtendría para distintos valores de k . Como B_t no puede superar el gasto total,

se eligió el máximo valor de k cuyo gasto base asociado no superara nunca el gasto total.

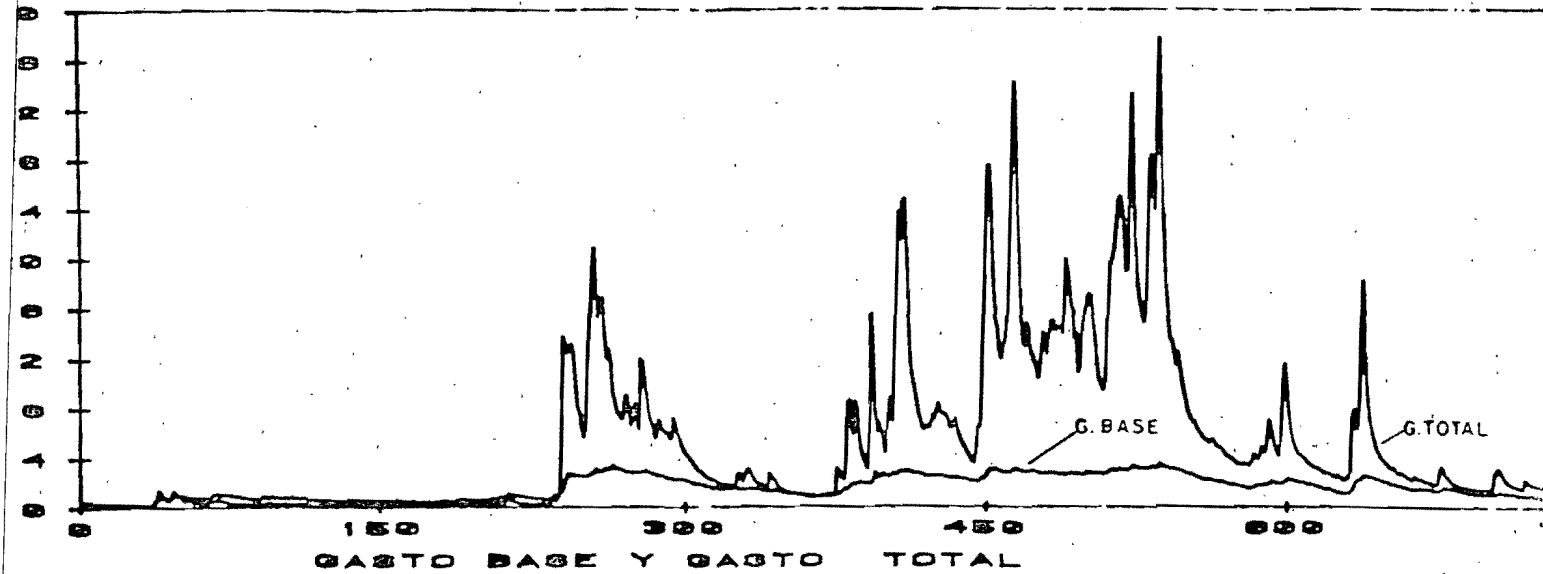


FIG 5.6

Determinación del modelo del escurrimiento superficial.

Se ajustó un modelo lineal al escurrimiento estimado

$$\bar{E}_t = \frac{B(q^{-1})}{A(q^{-1})} \bar{S}_t$$

donde

$$\begin{cases} \bar{E}_t = Y_t - \bar{B}_t & \text{(escurrimiento superficial estimado)} \\ \bar{S}_t = P_t - \bar{I}_t & \text{(precipitación efectiva estimada)} \end{cases}$$

\bar{B}_t y \bar{I}_t se estimaron según el párrafo anterior.

Una búsqueda sistemática del orden y retraso por mínimos cuadrados repetidos (ver [17]) de este modelo, arrojó como re-

sultado para $B(q^{-1})$ y $A(q^{-1})$ polinomios de la forma:

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + b_3 q^{-3}$$

$$A(q^{-1}) = 1 - a_1 q^{-1} - a_2 q^{-2}$$

Ya conocida la estructura del modelo de escurrimiento se aplicó el predictor multipasos autoadaptativo, prediciendo 4 pasos adelante el valor de \hat{E}_t con $\hat{E}_{t+4/t}$.

El esquema global del predictor de r pasos adelante que supone una estimación del modelo para el gasto base se representa en la fig 5.7.

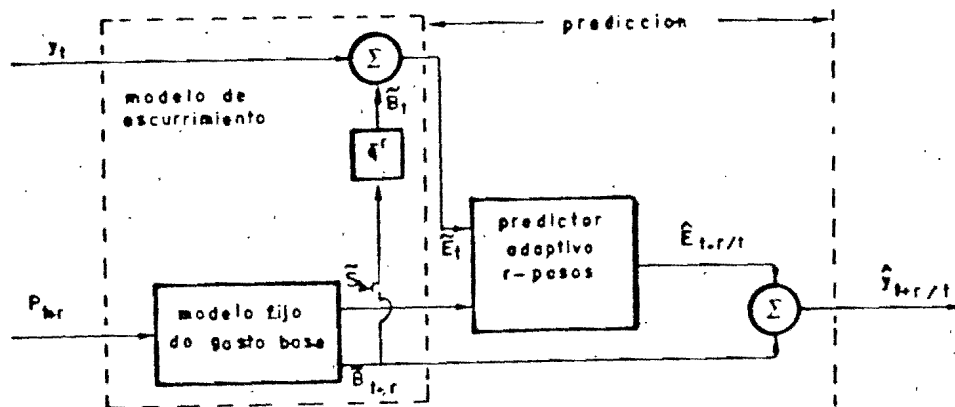


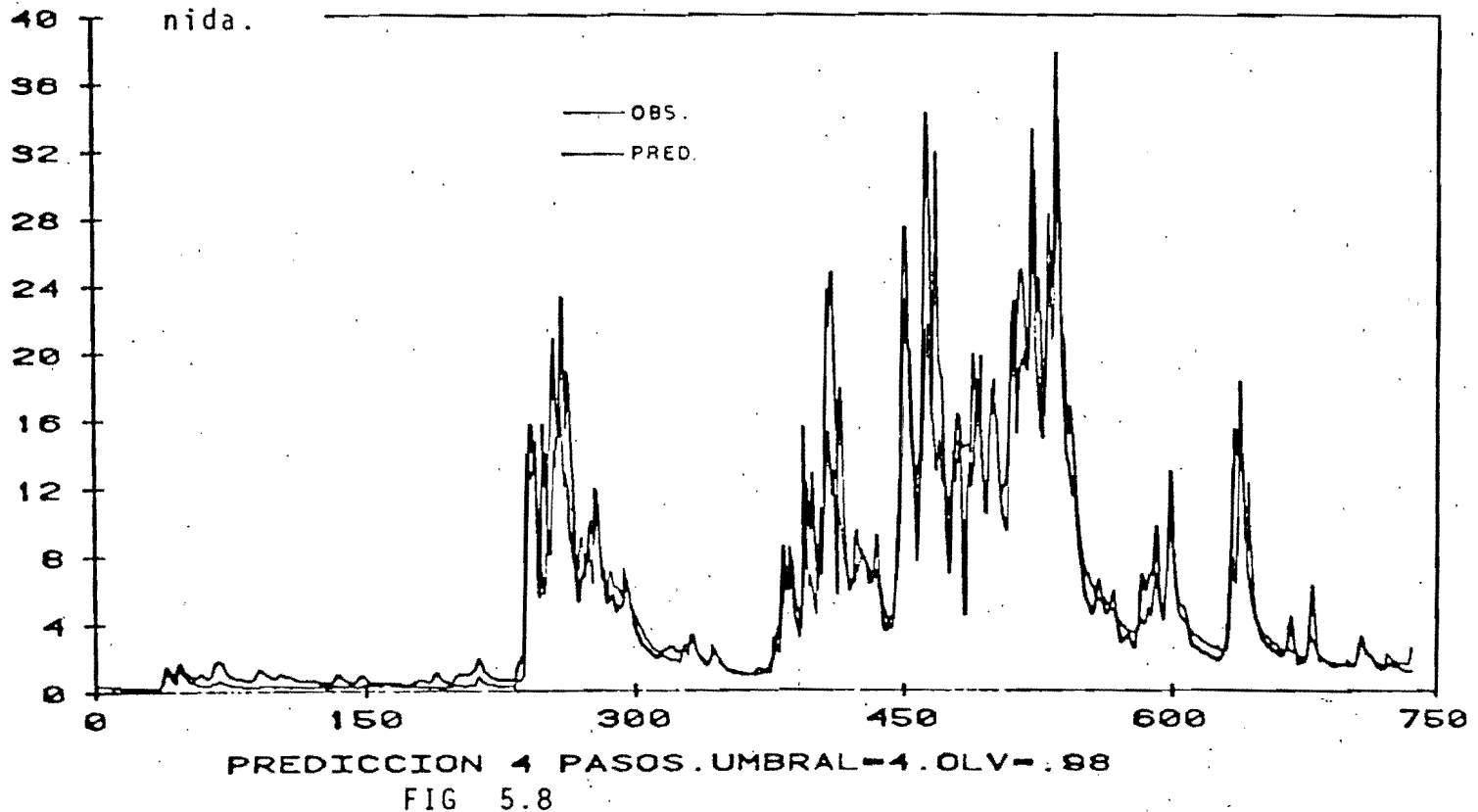
FIG 5.7 Esquema del predictor r pasos adelante, suponiendo conocidas las precipitaciones futuras.

Entonces
$$\hat{Y}_{t+4/t} = \hat{E}_{t+4/t} + \hat{B}_{t+4}$$

Debe notarse que para calcular $\hat{E}_{t+r/t}$ es necesario suponer conocidas las lluvias futuras (r -pasos adelante).

Para no tomar en cuenta períodos sin precipitaciones que aportarían poca información a la identificación, la adaptación del predictor se suspendía si el escurrimiento superficial estimado (\hat{E}_t) era menor que un umbral igual a 4 mm/6. h.

La figura (5.8) representa la predicción a 4 pasos así obtenida.



El predictor autoadaptivo se ejecutó con un factor de olvido constante $\lambda=0.98$.

Puede verse que hay zonas donde se predice muy bien, y puntos donde la predicción está francamente alejada del caudal observado.

Comparando los resultados con los obtenidos por R.DOMINGUEZ/M.

CHIN [26], quienes ajustaron un modelo no lineal fuera de línea con una optimización global, es notable la coincidencia que se observa en las dos predicciones del caudal, tanto en los lugares donde la predicción es buena, como donde la predicción es mala.

Creemos que los resultados fueron bastante satisfactorios considerando la poca confiabilidad de los datos, especialmente la medición de las precipitaciones.

5.3 Cuenca de Cantón, modelo lineal puro.

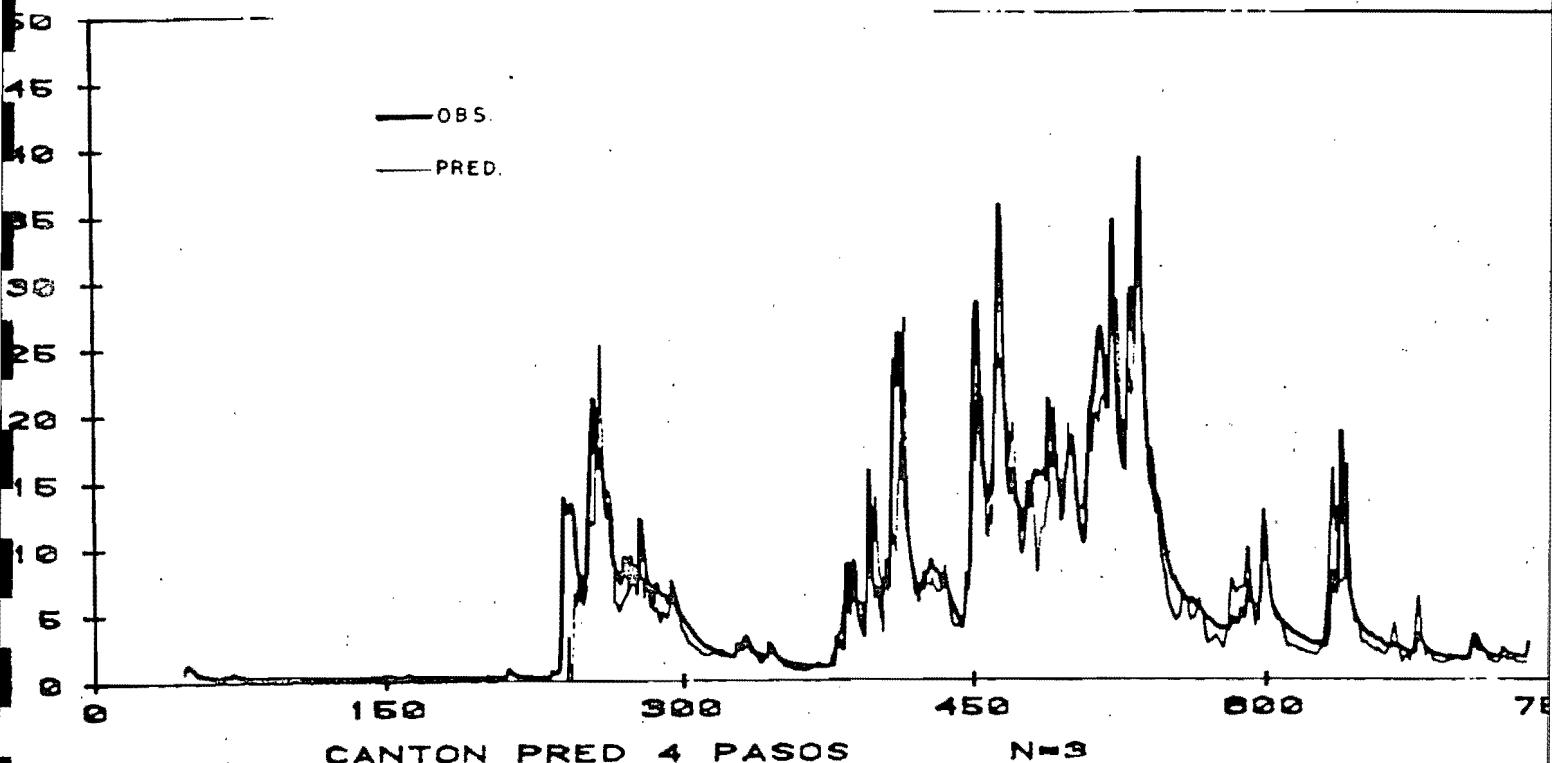
Vista la complejidad del cálculo anterior (cálculo fuera de línea de un modelo para el gasto base, estimación del escurrimiento superficial, y predicción sobre el gasto superficial estimado), se decidió ensayar el esquema más simple dado por el enfoque (d), es decir, ajustar un modelo lineal global, y adaptarlo a lo largo del tiempo.

Con los mismos datos, se puso a funcionar el predictor multi pasos autoadaptivo sobre las sucesiones de precipitaciones y gastos totales observados.

Otra vez se usó un factor de olvido de 0.98 para permitir la adaptación del predictor. La adaptación se interrumpía si $Y_t < 4 \text{ mm}$ (umbral).

La figura (5.9) representa el predictor a 4 pasos, suponiendo

do un modelo de orden 3 para el numerador y denominador de la función de transferencia. La fig (5.10) representa un detalle del comienzo de la predicción, allí se puede observar que luego de un período de adaptación, el predictor se acerca mucho a los valores reales del caudal de salida.



MM. FIG 5.9 Predictor lineal, 4 pasos adelante, modelo de tercer orden.

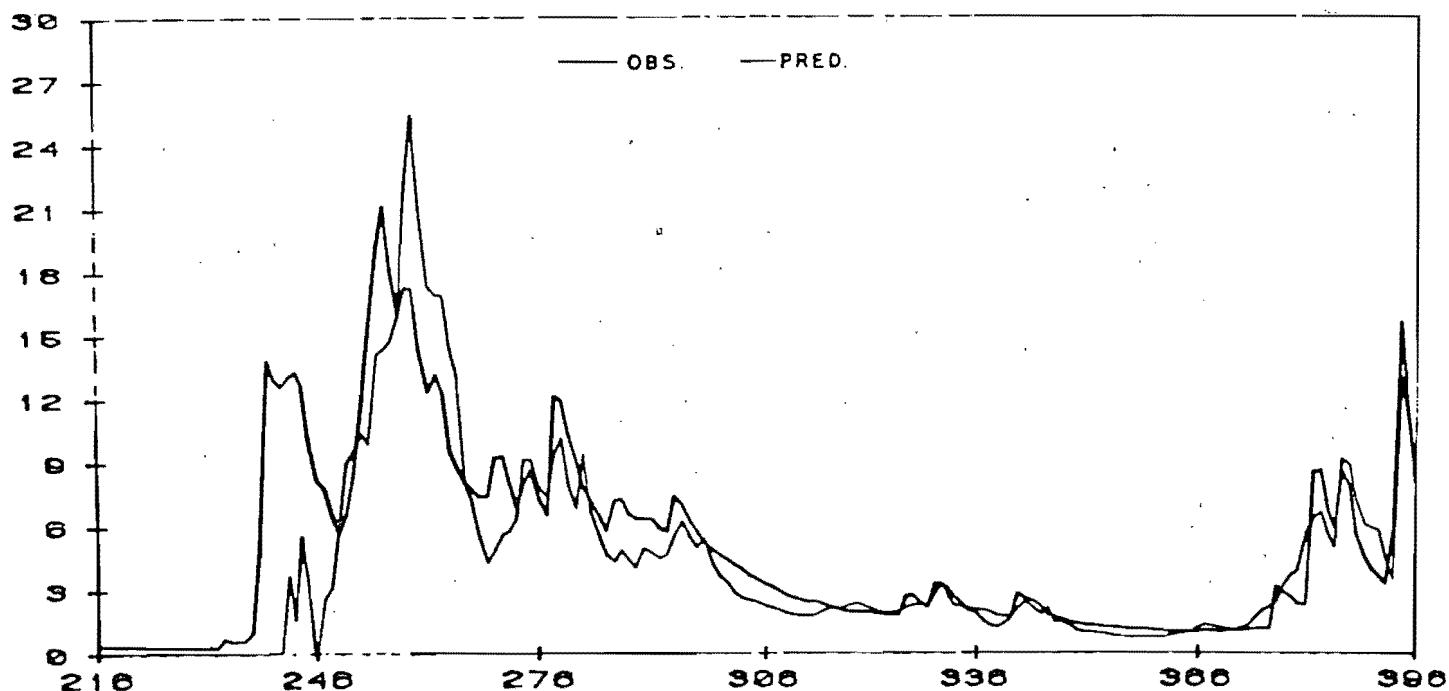


FIG 5.10 Detalle de las primeras predicciones (45 días).

La prueba de adaptación es bastante rigurosa: el predictor está inicializado con los parámetros de cuenca muy mojada, y los datos corresponden a las primeras lluvias luego de un prolongado período de sequía.

De las pruebas realizadas con datos reales obtenidos de la cuenca de Cantón concluimos que es inútil proponer un modelo de separación, no lineal, por cuanto un simple modelo lineal autoadaptable arroja resultados similares, incluso mejores en algunos puntos.

En segundo lugar, si se le compara con los modelos fijos, no lineales que se han intentado ajustar en forma global (R. Dominguez, M. Chin [26]), el predictor lineal autoadaptivo parece arrojar resultados muy similares para caudales medios y altos, y francamente mejores para caudales bajos. Esto se observa en ensayos con los mismos datos usados en este trabajo.

Debe tenerse en cuenta sin embargo que el modelo global citado [26], es un predictor de horizonte no acotado, mientras que el que hemos usado es a un horizonte fijo: r pasos (en nuestras pruebas $r = 4 = 24$ horas).

5.4 Pruebas con datos simulados.

Como vimos anteriormente, los datos observados sobre la cuenca de Cantón no eran muy confiables, especialmente (ver ref [26]), en lo que se refiere a la medición de las precipitaciones.

Probado el predictor lineal sobre dichas observaciones, con los resultados ya descritos en el párrafo anterior, surge otra pregunta: ¿cómo se comportaría este predictor si las observaciones siguieran una ley regida por uno de los modelos propuestos para describir la cuenca?

Se consideró el modelo no lineal propuesto por R. Domínguez, M. Chin (op.cit.) con los parámetros ajustados con las observaciones ya utilizadas de la cuenca de Cantón. Estos parámetros habían sido ajustados para minimizar el error cuadrático medio en el gasto total observado.

Dicho modelo, es esencialmente el discutido en la sección 5.1 (diferiendo solamente en el modelo no lineal de separación) y se resume:

$$\begin{array}{ll}
 Y_t = E_t + B_t & \text{gasto total} \\
 E_t = A_1 E_{t-1} + A_2 E_{t-2} + B_0 S_{t-1} + B_1 S_{t-2} + B_2 S_{t-3} & \text{escurr. superficial} \\
 B_t = a B_{t-1} + b X(t-2) & \text{gasto base} \\
 X_{t+1} = (1-\gamma) X_t + I_t & \text{estado de humedad} \\
 \left. \begin{array}{l}
 P_t = S_t + I_t \\
 I_t = \min(\phi_t, P_t) \\
 \phi_t = k_0 e^{-\alpha X_t}
 \end{array} \right\} & \text{modelo de separación}
 \end{array}$$

Los parámetros ajustados (ver ref [26] sobre el detalle del ajuste) fueron:

$$k = 57.578$$

$$a = 0.007535$$

$$\gamma = 0.010266$$

$$a = 0.34695$$

$$b = 0.006647$$

$$A_1 = 0.8356$$

$$A_2 = 0.0029029$$

$$B_0 = 0.16259$$

$$B_1 = -0.0038345$$

$$B_2 = 0.0053874$$

Con este modelo se simuló el comportamiento de la cuenca.

Como entrada se construyó una precipitación, que tuviera características como la observada en la cuenca de Cantón y que además sometiera el predictor a la predicción en temporadas de sequía y en temporada de lluvias altas.

Para ello se consideró la precipitación observada en la cuenca de Cantón durante 150 mediciones consecutivas (llamémosla serie A) y se construyó una serie simulada de precipitaciones concatenando la serie A, a una serie de 150 ceros (sequía absoluta) y dos veces la serie A según el esquema de la figura 5.11.

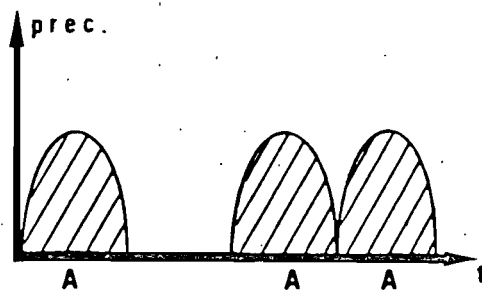


FIG 5.11

Suponiendo esta precipitación, se calculó con el modelo de la cuenca, el gasto total simulado, que se presenta en la figura 5.12 .

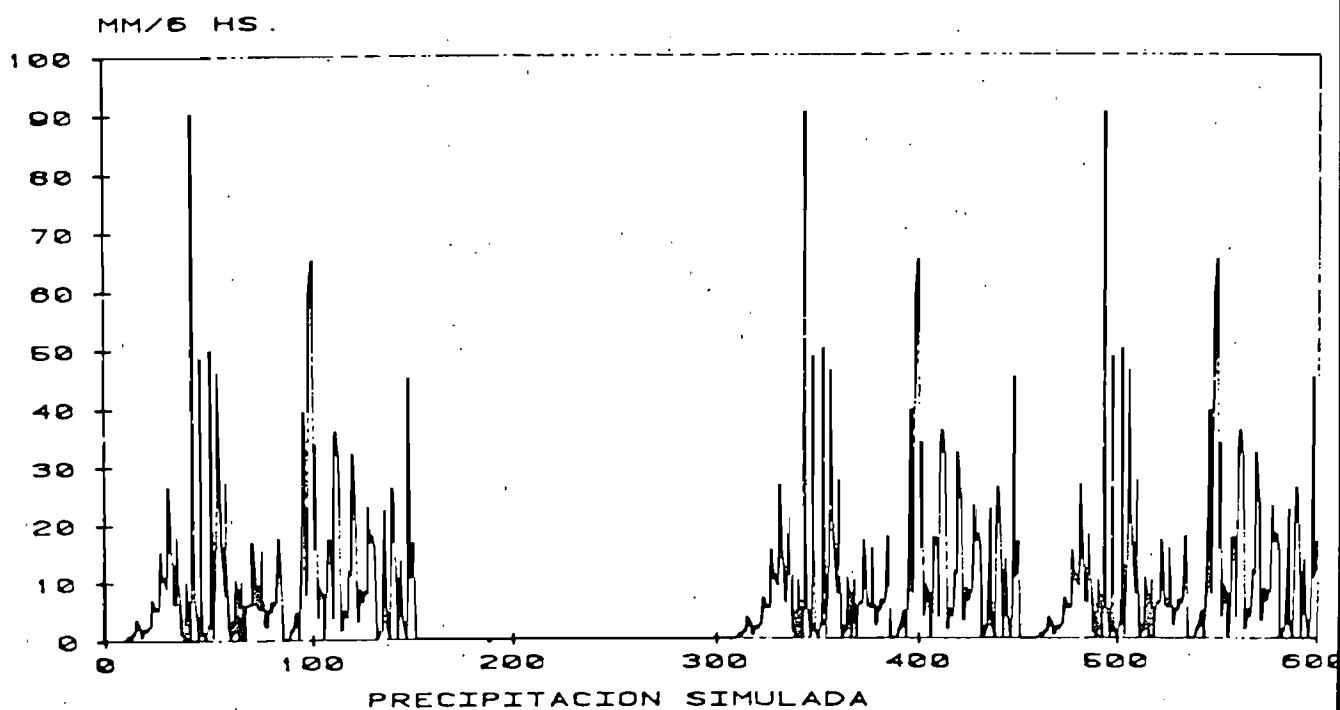
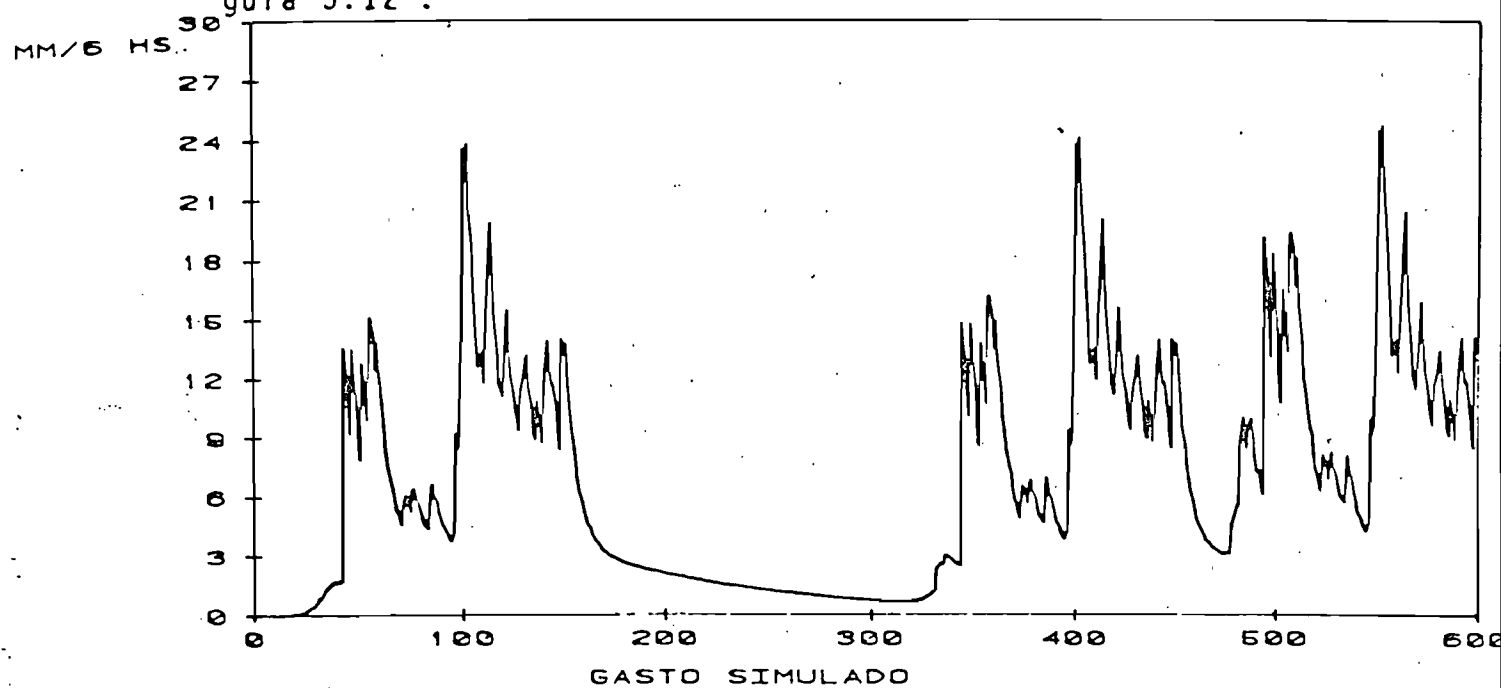


FIG 5.12

Sobre estos datos simulados (precipitación y gasto totales), se ajustó un predictor lineal autoadaptivo con dos esquemas distintos de adaptación (en la identificación del predictor óptimo): factor de olvido constante y factor de olvido variable calculado según el esquema de Fortescue et. al.

El umbral de adaptación del predictor lineal fue $Y_t = 2.5 \text{ mm/6hs}$, es decir que si el gasto total estaba por debajo de este valor, no se actualizaba el predictor lineal, y siempre la predicción fue hecha a 4 pasos.

En todos los casos, el predictor fue inicializado con parámetros adecuados para cuenca muy húmeda, mientras que el modelo simulado parte de una sequía total.

Tanto para factor de olvido constante como variable, se ensayó con modelos de orden 2 y 3.

La figura 5.13 muestra el comportamiento del predictor, con factor de olvido constante 0.98 (ventana efectiva = 50 pasos), y con modelos de orden 2 y 3.

En línea gruesa destaca el gasto simulado y en delgada el predicho a cuatro pasos adelante, conocidas las precipitaciones futuras.

Puede observarse en las figuras, un comportamiento muy bueno del predictor, y bastante similares los resultados con ordenes 2 ó 3.

Es notorio el transitorio de adaptación (de unos 35 pasos) al principio, donde el predictor se adapta de la inicialización a las condiciones de cuenca muy seca, y luego de $t=300$ un transitorio menor donde se adapta a las condiciones de cuenca semihúmeda.

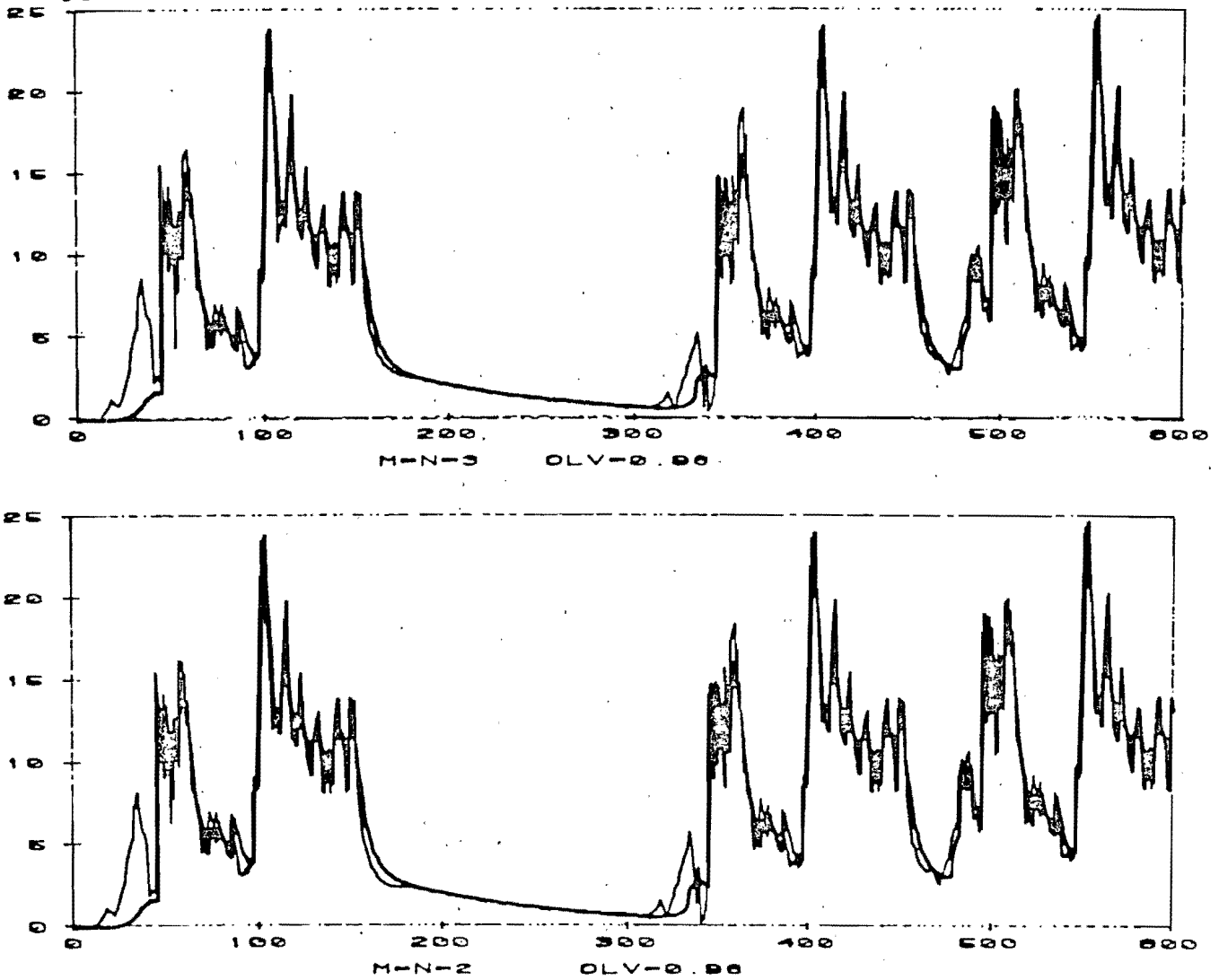


FIG 5.13 Predictor a 4 pasos, factor de olvido constante $\lambda=0.98$.

La figura 5.14 muestra el comportamiento del predictor con factor de olvido variable, de orden 3, y con 4 valores distintos de la cantidad de información: $I=\Sigma_0=0.5, 2.5, 10, 50$.

La figura 5.15 muestra los factores de olvido λ_t , instantáneos, asociados respectivamente a las 4 pruebas.

La figura 5.14 muestra que a pesar del rango amplio de la cantidad de información, el comportamiento del predictor no varía apreciablemente. La mayor adaptabilidad con $\sigma_0=2.5$, 10 que con $\sigma_0=50$ es ligeramente visible, pero ya con valores tan bajos de σ_0 como 0.5 comienzan a aparecer problemas en los transitorios: p. ej. véase que el transitorio desde $t = 305$ no se acaba en $t = 340$ como en los otros casos sino que se prolonga unos cuantos períodos de muestreo más.

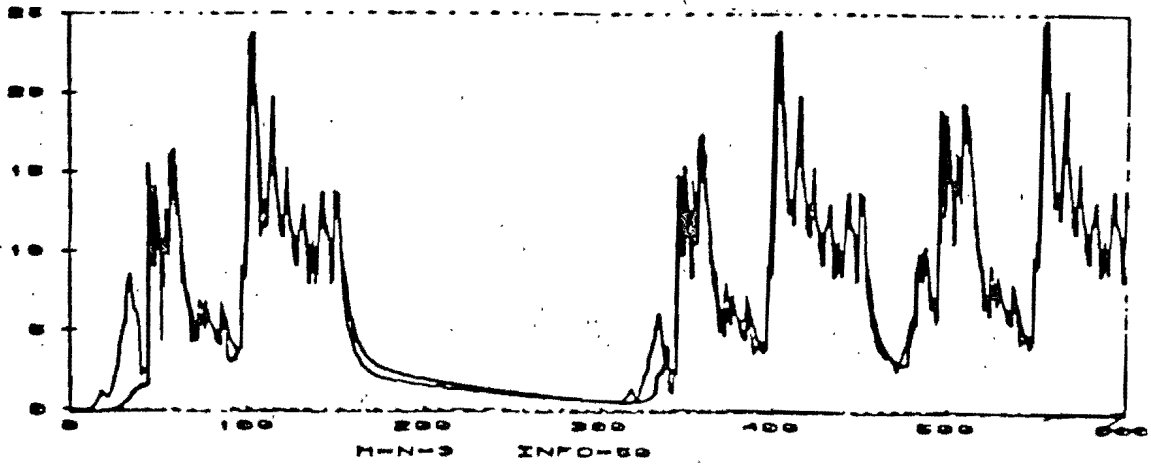
Finalmente en la figura 5.16 se muestran los resultados de aplicar el mismo predictor con $\sigma_0=30$, para dos ordenes distintos. Se omiten las gráficas de λ_t .

Los ordenes son 2 y 3 respectivamente.

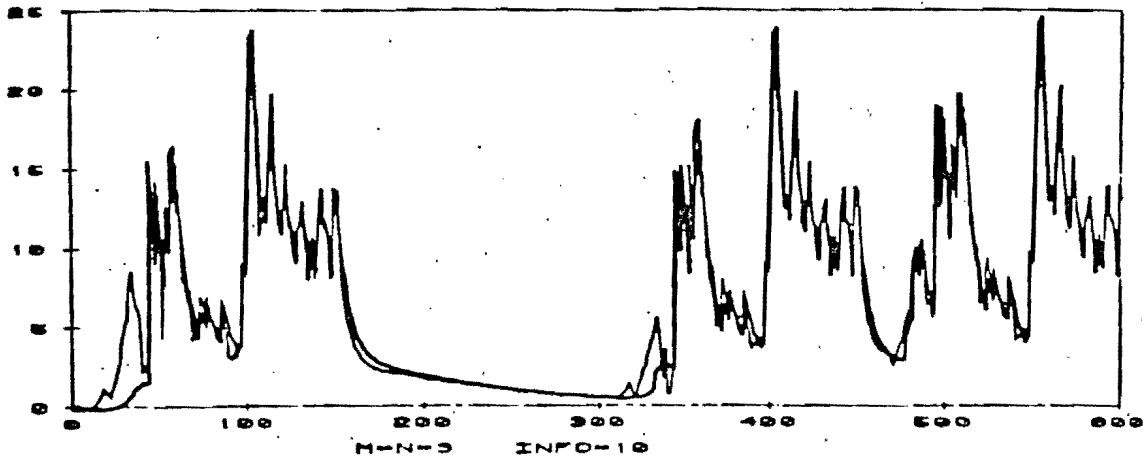
Los resultados son bastante similares entre sí.

En cuanto a la comparación entre los resultados que se obtienen con factor de olvido constante y variable, son análogos. En particular compárense las gráficas (para orden 3) que corresponden a λ constante = 0.98, y a variable con $\sigma_0=2.5$. Son prácticamente indistinguibles.

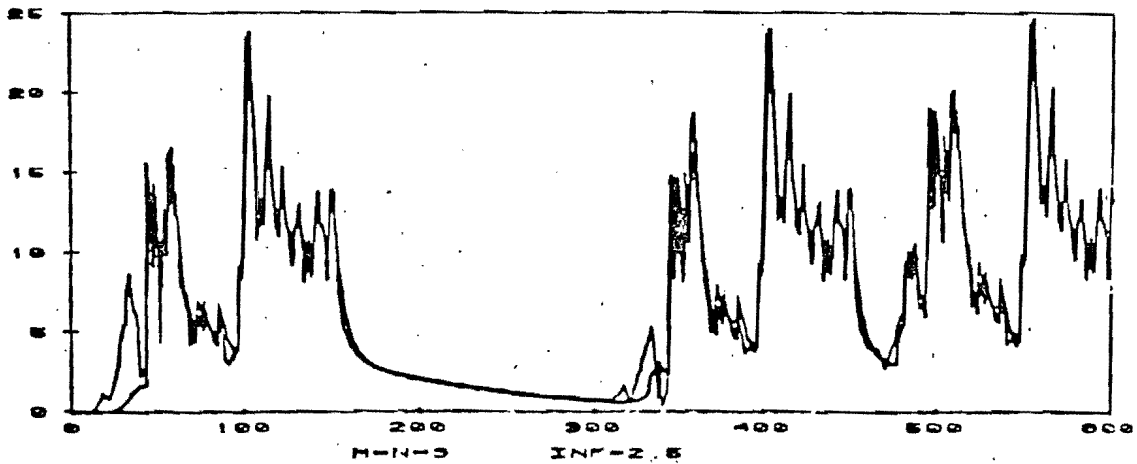
De la gráfica de λ_t con $\sigma_0=2.5$ no parece improbable que en promedio, λ_t estuviera alrededor de 0.98.



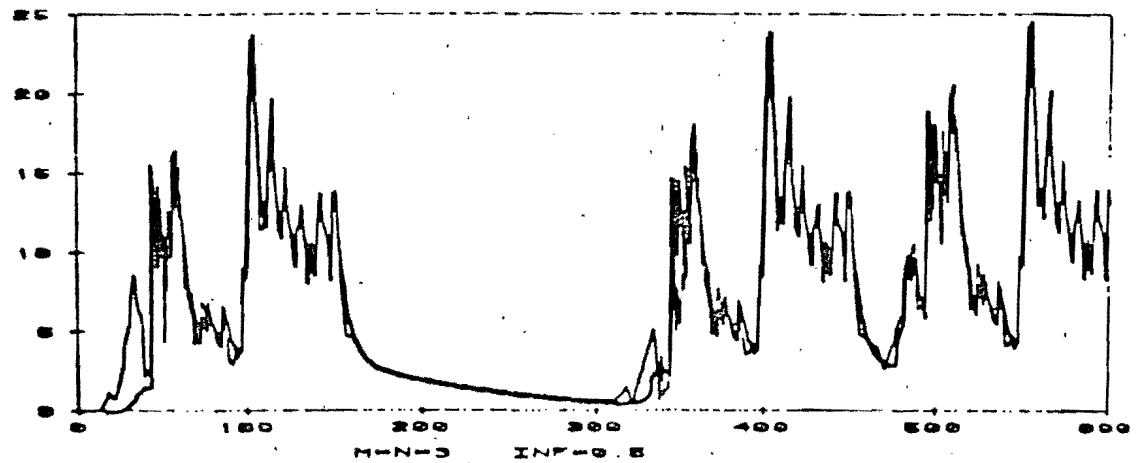
MM/O HS.



MM/O HS.



MM/O HS.



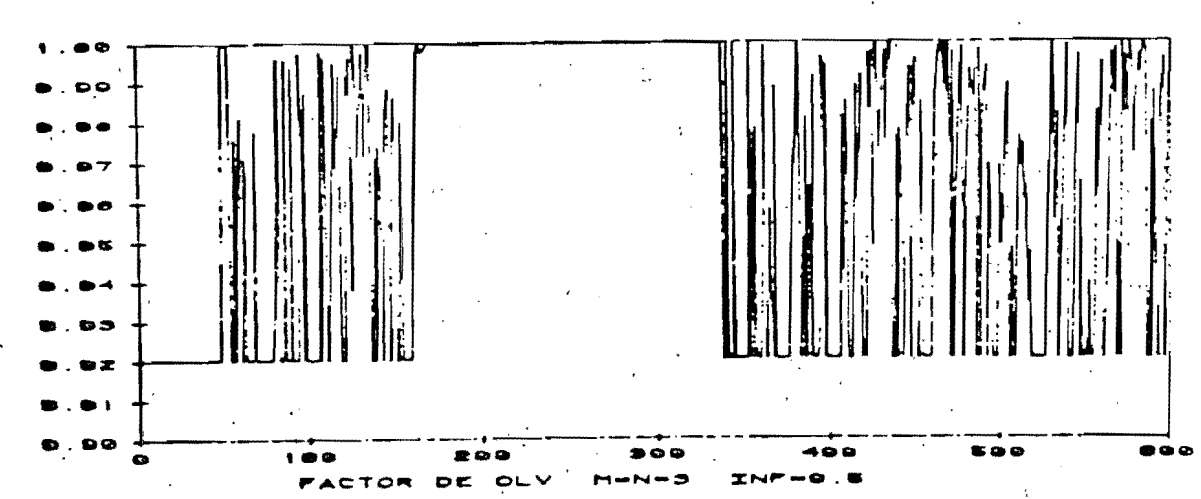
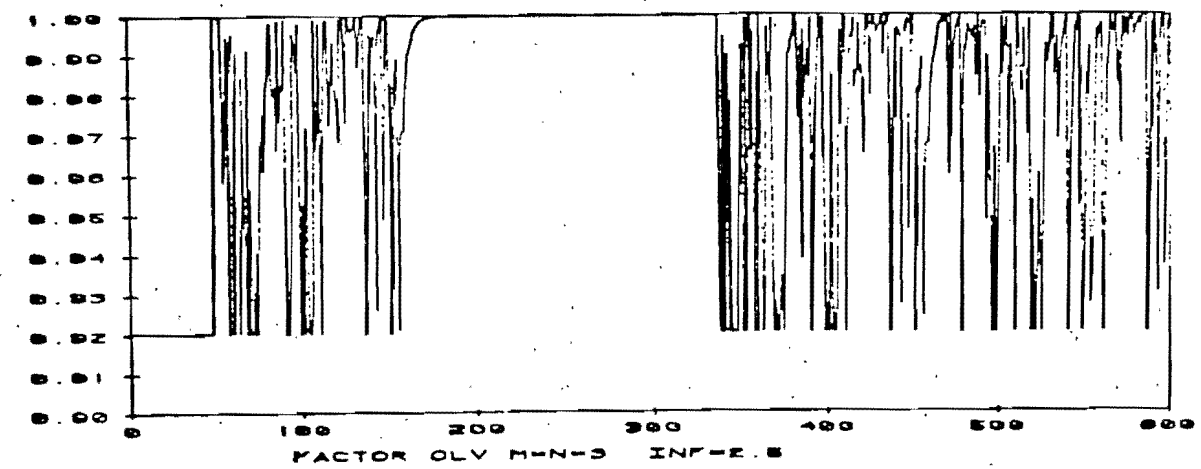
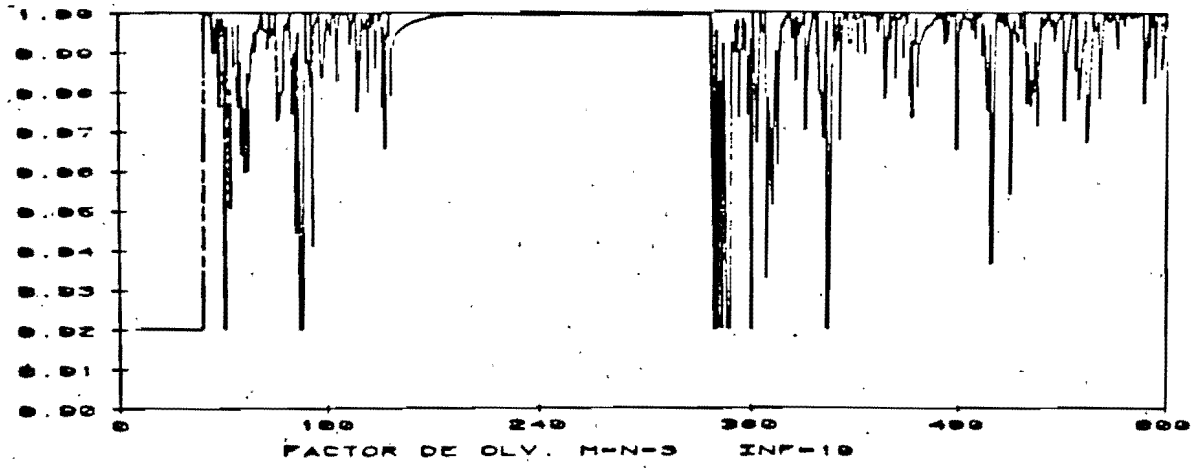
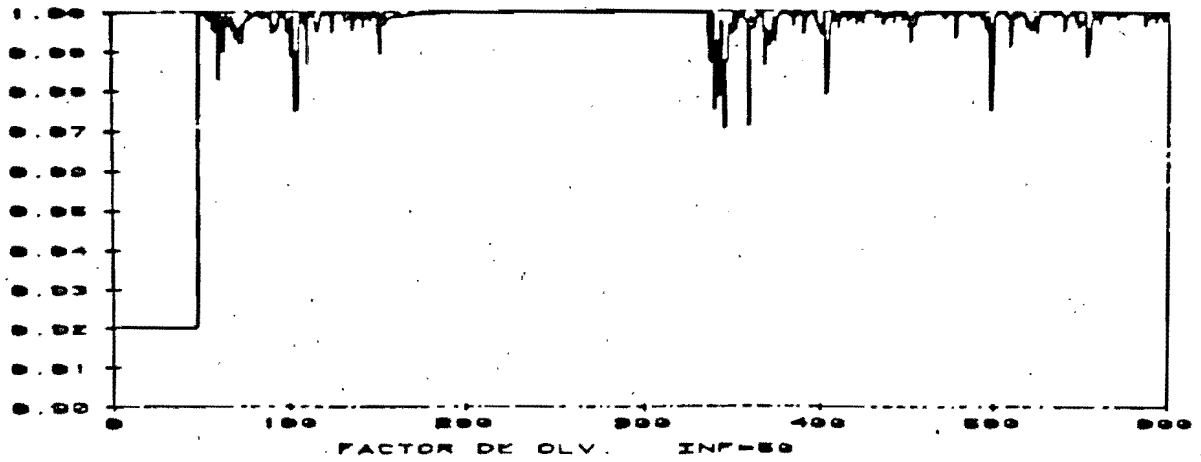


Figura 5.15

Podemos concluir que por las características del modelo no lineal, y la permanente variación de la linealización alrededor de su punto de operación (hecho que se deduce de las gráficas de λ_t), la preeicción se puede hacer con un predictor lineal autoadaptable con resultados muy satisfactorios, como lo muestran las figuras 5.13 a 5.16 .

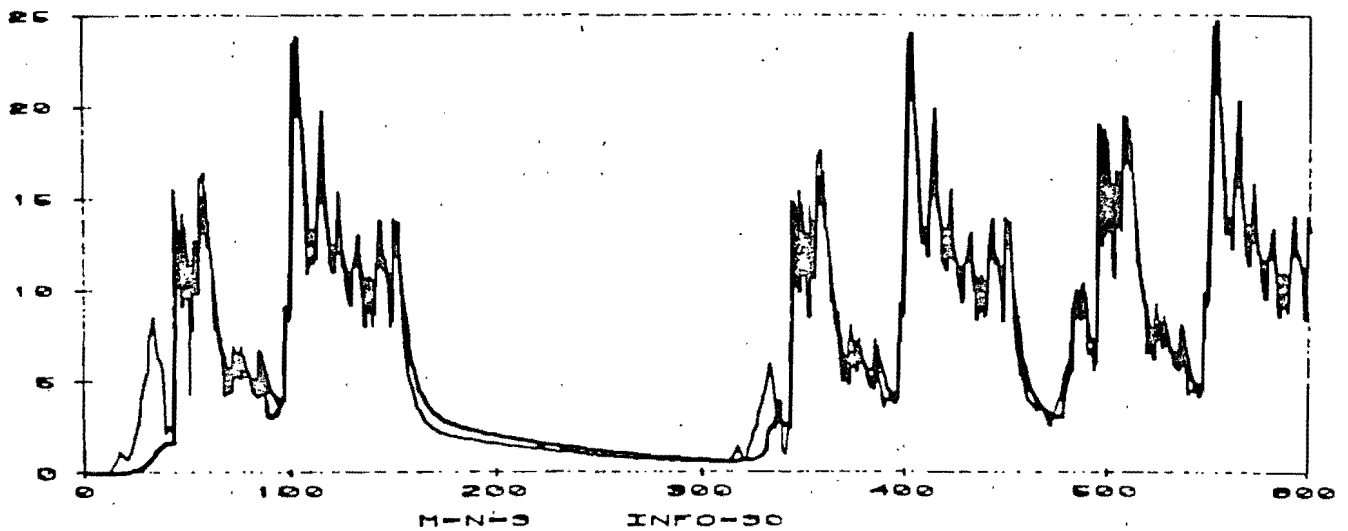
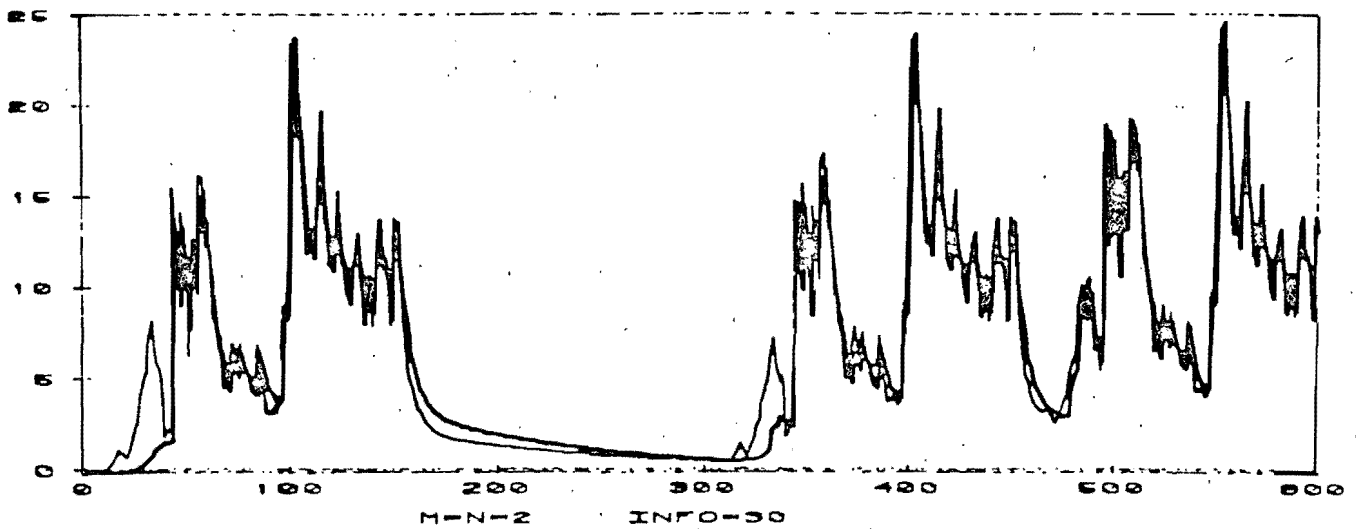


FIG 5.16

5.5. Conclusiones.

A través de las pruebas sobre observaciones reales en una cuenca (Cantón) se ha podido observar que la utilización del predictor lineal autoadaptivo arroja resultados similares a la utilización de modelos matemáticos bastante más complejos.

Esto se debe seguramente a que en la elaboración de dichos modelos se hacen un conjunto de simplificaciones y de suposiciones sobre la invariancia de ciertos parámetros entre períodos de calibración y de predicción, que resultan a la larga en una imprecisión del modelo.

Se aplicó el predictor lineal en dos formas:

En la primera se hizo una estimación del gasto base fuera de línea, y se utilizó el predictor para la predicción del escurrimiento superficial. Para ello se desarrolló una metodología de estimación del gasto base, fundamentada en el modelo supuesto para la cuenca.

En la segunda, se aproximó linealmente el comportamiento global de la cuenca, con un modelo lineal y variante en el tiempo.

Los resultados similares en las dos formas de utilización del predictor hacen suponer que no se consigue nada tratando de descomponer a "grosso modo" el gasto total en una parte li-

neal (escurrimiento superficial) y otra parte no lineal (gasto base).

Sin embargo, no fue posible distinguir si ésto es debido a las características de la cuenca o a las mediciones no muy confiables que se tenían de las precipitaciones.

Finalmente se comprobó que si las precipitaciones y el gasto total en la cuenca estuvieran vinculadas a través de un modelo no lineal ya estudiado en el Instituto, el predictor lineal autoajustable predice con gran precisión el comportamiento del gasto total supuestas conocidas las precipitaciones futuras.

APENDICE A. MATRICES ELEMENTALES DE HOUSEHOLDER

Definición

Diremos que una matriz $A \in R^{p \times p}$ es elemental según Householder (ver Householder [36]) si se cumple que:

$$A = I - \rho uv^T \quad (\text{A.1})$$

donde:

$$\begin{aligned} I &= I_p \\ \rho &\in R \\ u, v &\in R^p \end{aligned} \quad (\text{A.2})$$

Propiedades. Las siguientes propiedades de las "matrices elementales de Householder" son directamente verificables:

1) Determinante

$$\det \{I - \rho uv^T\} = 1 - \rho v^T u \quad (\text{A.3})$$

2) Traza

$$t_r \{I - \rho uv^T\} = p - \rho v^T u \quad (\text{A.4})$$

3) Inversa

$$\{I - \rho uv^T\}^{-1} = \{I - \sigma uv^T\} \quad (\text{A.5})$$

donde

$$\sigma^{-1} + \rho^{-1} = v^T u \quad (\text{A.6})$$

o sea:

$$\{I - \rho uv^T\}^{-1} = \left\{ I - \frac{1}{v^T u - 1/\rho} uv^T \right\} \quad (\text{A.7})$$

4) Autovalores

Denominando $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ a los autovalores (ordenados) de la matriz A, y $x_1 \dots x_p$ sus autovectores asociados, entonces:

i) $\lambda_1 = \lambda_2 = \dots = \lambda_{p-1} = 1$, asociadas a una variedad ortogonal a v .

ii) $\lambda_p = 1 - \rho v^T u$, con $x_p = u$

Demostración

$$2) \quad t_r\{I - \rho uv^T\} = t_r\{I\} - t_r\{\rho uv^T\} = p - \rho v^T u$$

4) Es directamente comprobable que son autovectores: u , con autovalor $\lambda_p = 1 - \rho v^T u$, y cualquier vector $x \perp v$ con autovalores asociados 1.

$$1) \quad \det\{I - \rho uv^T\} = \lambda_1 \dots \lambda_p = 1 \cdot 1 \dots 1 \cdot (1 - \rho v^T u) = (1 - \rho v^T u)$$

3) Siendo $\det A \neq 0$, alcanza con verificar que el producto de las expresiones A.1 y A.7 dá como resultado la identidad.

APENDICE B. ALGUNAS DESIGUALDADES PARA AUTOVALORES

1. Sea una matriz $A(p \times p)$ real y simétrica, con autovalores ordenados:

$$\lambda_{\max}\{A\} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = \lambda_{\min}\{A\}, \text{ y } x \in \mathbb{R}^p, \quad x \neq 0$$

entonces:

$$\lambda_{\min}\{A\} \leq \frac{x^T Ax}{x^T x} \quad (\text{B.1})$$

$$\lambda_{\max}\{A\} \geq \frac{x^T Ax}{x^T x} \quad (\text{B.2})$$

$$\lambda_{\min}\{A\} = \min_x \frac{x^T Ax}{x^T x} \quad (\text{B.3})$$

$$\lambda_{\max}\{A\} = \max_x \frac{x^T Ax}{x^T x} \quad (\text{B.4})$$

Demostraciones de lo anterior pueden encontrarse en [44]

y [45]

2. Sean A, B matrices ($p \times p$), y $C = A + B$
(A, B reales y simétricas)

Entonces

$$\lambda_{\min}\{C\} \geq \lambda_{\min}\{A\} + \lambda_{\min}\{B\} \quad (\text{B.5})$$

$$\lambda_{\max}\{C\} \leq \lambda_{\max}\{A\} + \lambda_{\max}\{B\} \quad (\text{B.6})$$

Demostración:

$$\frac{x^T C x}{x^T x} = \frac{x^T A x}{x^T x} + \frac{x^T B x}{x^T x} \quad (\text{B.7})$$

por (B.1):

$$\frac{x^T C x}{x^T x} \geq \lambda_{\min}\{A\} + \lambda_{\min}\{B\}$$

pero si elegimos x tal que sea el autovector asociado al autovalor mínimo de C y $\|x\|_2 = 1$, entonces:

$$\lambda_{\min}\{C\} \geq \lambda_{\min}\{A\} + \lambda_{\min}\{B\}$$

Con lo que se demuestra B.5. Análogamente se demuestra B.6 considerando los autovalores máximos.

3. Sea $A \in \mathbb{R}^{p \times p}$ simétrica y positiva definida. Entonces:

$$\lambda_{\min}\{A\} = 1/\lambda_{\max}\{A^{-1}\} \quad (\text{B.8})$$

La demostración puede encontrarse en [44]

APENDICE C

Demostración de la Proposición 1

Llamando $\ell_{\min}\{P_t^{-1}\}$ y $\ell_{\max}\{P_t^{-1}\}$ a los autovalores mínimo y máximo de P_t^{-1} respectivamente. Por (3.11):

$$P_t^{-1} = \lambda_1 \lambda_2 \dots \lambda_t P_0^{-1} + \sum_{i=1}^t \left(\prod_{j=1}^t \lambda_j \right) x_i x_i^T \quad (C.1)$$

luego, por (B.5):

$$\begin{aligned} \ell_{\min}\{P_t^{-1}\} &\geq \ell_{\min}\left\{ \sum_{i=t-p+1}^t \left(\prod_{j=i}^t \lambda_j \right) x_i x_i^T \right\} + \ell_{\min}\left\{ \sum_{i=t-2p+1}^{t-p} \left(\prod_{j=i}^t \lambda_j \right) x_i x_i^T \right\} + \dots \\ &\dots + \ell_{\min}\left\{ \sum_{i=t-rp+1}^{t-rp+p} \left(\prod_{j=i}^t \lambda_j \right) x_i x_i^T \right\} + \ell_{\min}\{\lambda_1 \dots \lambda_t P_0^{-1}\} \end{aligned} \quad (C.2)$$

como $\lambda_j \geq \lambda_m \quad \forall j$:

$$\ell_{\min}\{P_t^{-1}\} \geq \ell_{\min}\left\{\sum_{i=t-p+1}^t \lambda_m^p x_i x_i^T\right\} + \dots + \ell_{\min}\left\{\sum_{i=t-rp+1}^{t-rp+p} \lambda_m^{rp} x_i x_i^T\right\} + \ell_{\min}\{\lambda_m^t P_0^{-1}\} \quad (C.3)$$

$$\geq \lambda_m^p \ell_{\min}\left\{\sum_{i=t-p+1}^t x_i x_i^T\right\} + \dots + \lambda_m^{rp} \ell_{\min}\left\{\sum_{i=t-rp+1}^{t-rp+p} x_i x_i^T\right\} + \lambda_m^t \ell_{\min}\{P_0^{-1}\} \quad (C.4)$$

luego, por (3.50):

$$\ell_{\min}\{P_t^{-1}\} \geq \lambda_m^p m + \lambda_m^{2p} m + \dots + \lambda_m^{rp} m + \lambda_m^t \ell_{\min}\{P_0^{-1}\} \quad \text{c.s.}$$

$$\ell_{\min}\{P_t^{-1}\} \geq m \frac{1 - \lambda_m^{rp}}{1 - \lambda_m^p} \lambda_m^p + \lambda_m^t \ell_{\min}\{P_0^{-1}\} \quad \text{c.s.} \quad (C.5)$$

como por (B.8) resulta

$$\ell_{\min}\{P_t^{-1}\} = 1/\ell_{\max}\{P_t\}$$

entonces

$$\ell_{\max}\{P_t\} \leq \left[m \lambda_m^p \frac{1 - \lambda_m^{rp}}{1 - \lambda_m^p} + \lambda_m^t \ell_{\min}\{P_0^{-1}\} \right]^{-1} \leq \frac{1}{m \lambda_m^p} \frac{1 - \lambda_m^p}{1 - \lambda_m^{rp}} \quad \text{c.s.} \quad (C.6)$$

con lo que se demuestra la segunda desigualdad de (3.56).

Por otro lado, notando que como $\lambda_m < 1$ y $r \geq 1$ ($t \geq p$), se cumple que

$$1 - \lambda_m^p \leq 1 - \lambda_m^{rp} \quad (C.7)$$

por lo que

$$\frac{1}{m \lambda_m^p} \frac{1 - \lambda_m^p}{1 - \lambda_m^{rp}} \leq \frac{1}{m \lambda_m^p} \triangleq \ell_M \quad (C.8)$$

con lo que se demuestra la cota (3.55).

Para demostrar la otra parte de la desigualdad, partimos de (C.1), y por (B.6):

$$\begin{aligned} \ell_{\max} \{P_t^{-1}\} &\leq \ell_{\max} \{\lambda_1 \dots \lambda_t P_0^{-1}\} + \ell_{\max} \left\{ \sum_{i=t-p+1}^t \left(\prod_{j=i}^t \lambda_j \right) x_i x_i^T \right\} + \dots \\ &\dots + \ell_{\max} \left\{ \sum_{i=t-rp+1}^{t-rp} \left(\prod_{j=i}^t \lambda_j \right) x_i x_i^T \right\} + \ell_{\max} \left\{ \sum_{i=1}^{t-rp} \left(\prod_{j=i}^t \lambda_j \right) x_i x_i^T \right\} \quad (C.9) \end{aligned}$$

por (3.50) y (3.52):

$$\ell_{\max} \{P_t^{-1}\} \leq \lambda_M^t \ell_{\max} \{P_0^{-1}\} + M \lambda_M \frac{1 - \lambda_M^{(r+1)p}}{1 - \lambda_M^p} \text{ c.s.} \quad (C.10)$$

como $\ell_{\max} \{P_t^{-1}\} = 1/\ell_{\min} \{P_t\}$

$$\ell_{\min} \{P_t\} \geq \left[\lambda_M^t \ell_{\max} \{P_0^{-1}\} + M \lambda_M \frac{1 - \lambda_M^{(r+1)p}}{1 - \lambda_M^p} \right]^{-1} \text{ c.s.} \quad (c.11)$$

con lo que se demuestra la primera parte de (3.56).

Además, de (C.11) resulta que $\forall_t \geq p$:

$$\ell_{\min}\{P_t\} \geq \left[\lambda_M^P \ell_{\max}\{P_0^{-1}\} + \frac{M \cdot \lambda_M}{1 - \lambda_M^P} \right]^{-1} \Delta \triangleq \ell_m \quad \text{c.s.} \quad (\text{C.12})$$

con lo que se verifica (3.54).

Proposición 2

De acuerdo a (3.34):
$$\tilde{\theta}_{t+1}^1 = \left[I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T \right] \tilde{\theta}_t^1 \quad (\text{C.13})$$

por otro lado, de (3.30):

$$\begin{aligned} \left[I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T \right] &= I - P_t x_{t+1} x_{t+1}^T + \frac{P_t x_{t+1} x_{t+1}^T P_t x_{t+1} x_{t+1}^T}{1 + x_{t+1}^T P_t x_{t+1}} = \\ &= I - \frac{P_t x_{t+1} x_{t+1}^T}{1 + x_{t+1}^T P_t x_{t+1}} = P_t P_t^{-1} - \frac{P_t x_{t+1} x_{t+1}^T P_t P_t^{-1}}{1 + x_{t+1}^T P_t x_{t+1}} \end{aligned} \quad (\text{C.14})$$

entonces:

$$\left[I - \lambda_{t+1} P_{t+1} x_{t+1} x_{t+1}^T \right] = \left[P_t - \frac{P_t x_{t+1} x_{t+1}^T P_t}{1 + x_{t+1}^T P_t x_{t+1}} \right] P_t^{-1} = \lambda_{t+1} P_{t+1} P_t^{-1} \quad (\text{C.15})$$

Sustituyendo (C.15) en (C.13):

$$\tilde{\theta}_{t+1}^1 = \lambda_{t+1} P_{t+1} P_t^{-1} \tilde{\theta}_t^1 \quad (\text{C.16})$$

Aplicando (C.4) regresivamente en el tiempo se concluye

$$\begin{aligned} \tilde{\theta}_{t+1}^T P_{t+1}^{-1} \tilde{\theta}_{t+1}^1 &= \lambda_{t+1} \tilde{\theta}_t^T P_t^{-1} \left[I - \frac{P_t x_{t+1} x_{t+1}^T}{1 + x_{t+1}^T P_t x_{t+1}} \right] \tilde{\theta}_t^1 = \\ &= \lambda_{t+1} \left[\tilde{\theta}_t^T P_t^{-1} \tilde{\theta}_t^1 - \frac{(\tilde{\theta}_t^T x_{t+1})^2}{1 + x_{t+1}^T P_t x_{t+1}} \right] \end{aligned} \quad (C.20)$$

y por las ecuaciones (3.39):

$$\tilde{\theta}_{t+1}^T P_{t+1}^{-1} \tilde{\theta}_{t+1}^1 = \lambda_{t+1} \left[\tilde{\theta}_t^T P_t^{-1} \tilde{\theta}_t^1 - \frac{\epsilon_{t+1}^2}{1 + x_{t+1}^T P_t x_{t+1}} \right] \quad (C.21)$$

Como $\lambda_{t+1} < 1$, $P_t > 0$ y $\epsilon_{t+1}^2 \geq 0$, entonces de (C.21) se deduce que la sucesión es decreciente. Es positiva puesto que cómo x_{t+1} es persistentemente excitante, entonces P_{t+1} (y también P_{t+1}^{-1}) es positiva definida debido a la proposición 1.

De (C.21):

$$\tilde{\theta}_{t+1}^T P_{t+1}^{-1} \tilde{\theta}_{t+1}^1 \leq \lambda_M^{t+1} \tilde{\theta}_0^T P_0^{-1} \tilde{\theta}_0^1 \quad (C.22)$$

siendo $\lambda_M < 1$, $P_0 > 0$, entonces se deduce (3.64)

Para demostrar ii):

$$\text{De acuerdo a (3.61)} \quad \|\tilde{\theta}_t^1\| \leq \left(\prod_{i=1}^t \lambda_i \right) \frac{\epsilon_{\max} P_t}{\epsilon_{\min} P_0} \|\tilde{\theta}_0^1\|$$

(3.60), con lo que se demuestra i).

Para demostrar ii):

$$\text{de (3.60)} \quad \|\tilde{\theta}_t^1\| \leq \prod_{i=1}^t \lambda_i \ell_{\max}^{P_t} \ell_{\max}^{P_0^{-1}} \|\tilde{\theta}_0^1\| \quad (\text{C.17})$$

como $\ell_{\max}^{P_0^{-1}} = 1/\ell_{\min}^{P_0}$ (ya que $P_0 > 0$), entonces:

$$\|\tilde{\theta}_t^1\| \leq \left(\prod_{i=1}^t \lambda_i \right) \frac{\ell_{\max}^{P_t}}{\ell_{\min}^{P_0}} \|\tilde{\theta}_0^1\| \quad (\text{C.18})$$

Con lo que se demuestra (3.61).

Para demostrar iii) se procede como antes, pero tomando autovalores mínimos en lugar de máximos.

Proposición 3

Para demostrar i):

De acuerdo a (C.16)

$$\tilde{\theta}_{t+1}^{1T} P_{t+1}^{-1} \tilde{\theta}_{t+1}^1 = \lambda_{t+1}^2 \tilde{\theta}_t^{1T} P_t^{-1} P_{t+1} P_{t+1}^{-1} P_{t+1} P_t^{-1} \tilde{\theta}_t^1 = \lambda_{t+1}^2 \tilde{\theta}_t^{1T} P_t^{-1} P_{t+1} P_t^{-1} \tilde{\theta}_t^1 \quad (\text{C.19})$$

de acuerdo a (C.15), (C.14) y (C.19):

Como $\|\tilde{\theta}_0^1\| < \infty$, y $\ell_{\text{máx}} P_t < \ell_m$ c.s. (constante) por la proposición 1, entonces

$$0 < \|\tilde{\theta}_t^1\| \leq \left(\prod_{i=1}^t \lambda_i \right) \frac{\ell_m}{\ell_{\text{mín}} P_0} \|\tilde{\theta}_0^1\| \triangleq C_t \quad \text{c.s.}$$

como $\lim_{t \rightarrow \infty} C_t = 0$, c.s., se deduce (3.66)

Proposición 4

Aplicaremos (3.68) para cada salto en el parámetro y por su-
perposición:

$$\|\tilde{\theta}_{t_{0+\tau}}^2\| \leq \sum_{i=t_{0+1}}^{t_{0+\tau}} \left(\prod_{j=i}^{t_{0+\tau}} \lambda_j \right) \frac{M}{m} \frac{1 - \lambda_m^p}{\lambda_m^p} \frac{\lambda_M}{1 - \lambda_M^p} \|\Delta\theta_i\|$$

sacando la constante como factor común, y el máximo de $\|\Delta\theta_i\|$:

$$\|\tilde{\theta}_{t_{0+\tau}}^2\| \leq \frac{M}{m} \frac{1 - \lambda_m^p}{\lambda_m^p} \frac{\lambda_M}{1 - \lambda_M^p} \left(\max_{t_{0+1} \leq i \leq t_{0+\tau}} \|\Delta\theta_i\| \right) \sum_{i=t_{0+1}}^{t_{0+\tau}} \left(\prod_{j=i}^{t_{0+\tau}} \lambda_j \right) \quad (\text{C.23})$$

pero como

$$\sum_{i=t_{0+1}}^{t_{0+\tau}} \left(\prod_{j=i}^{t_{0+\tau}} \lambda_j \right) \leq \sum_{i=t_{0+1}}^{t_{0+\tau}} \lambda_M^{t_{0+\tau}-i+1} = \lambda_M \frac{1 - \lambda_M^\tau}{1 - \lambda_M} < \frac{\lambda_M}{1 - \lambda_M} \quad (\text{C.24})$$

Sustituyendo (C.24) en (C.23) se deducen (3.73) y (3.74).

Demostración de la proposición 5:

De (3.36) y (C.15):

$$\begin{aligned}
 \tilde{\theta}_{t+1}^3 P_{t+1}^{-1} \tilde{\theta}_{t+1}^3 &= \lambda_{t+1} \tilde{\theta}_t^{3T} P_t^{-1} \tilde{\theta}_t^3 - \lambda_{t+1}^2 \tilde{\theta}_t^{3T} x_{t+1} x_{t+1}^T P_{t+1} P_t^{-1} \tilde{\theta}_t^3 - \\
 &- \lambda_{t+1} x_{t+1} e_{t+1} + \lambda_{t+1}^2 \tilde{\theta}_t^{3T} x_{t+1} x_{t+1}^T P_t x_{t+1} e_{t+1} \\
 &- \lambda_{t+1}^2 x_{t+1} P_{t+1} P_t^{-1} \tilde{\theta}_t^3 e_{t+1} + \lambda_{t+1}^2 e_{t+1}^2 x_{t+1} P_{t+1} x_{t+1}
 \end{aligned}
 \tag{C.25}$$

Nótese que el 3°, 4° y 5° sumandos tienen media nula y pueden ponerse como $z_{t+1} e_{t+1}$ con z_{t+1} independiente de e_{t+1} .

Como

$$\lambda_{t+1} x_{t+1} x_{t+1}^T P_{t+1} P_t^{-1} = \frac{1}{1 + x_{t+1}^T P_t x_{t+1}} x_{t+1} x_{t+1}^T$$

y

$$\lambda_{t+1} x_{t+1}^T P_{t+1} x_{t+1} = \frac{x_{t+1}^T P_t x_{t+1}}{1 + x_{t+1}^T P_t x_{t+1}}$$

entonces (C.25) queda:

$$\begin{aligned}
 \tilde{\theta}_{t+1}^3 P_{t+1}^{-1} \tilde{\theta}_{t+1}^3 &= \lambda_{t+1} \tilde{\theta}_t^{3T} P_t^{-1} \tilde{\theta}_t^3 - \frac{\lambda_{t+1}}{1 + x_{t+1}^T P_t x_{t+1}} (\tilde{\theta}_t^{3T} x_{t+1})^2 \\
 &+ z_{t+1} e_{t+1} + \lambda_{t+1} \frac{x_{t+1}^T P_t x_{t+1}}{1 + x_{t+1}^T P_t x_{t+1}} e_{t+1}^2
 \end{aligned}
 \tag{C.26}$$

como el segundo sumando es negativo o nulo:

$$\tilde{\theta}_{t+1}^{3T} P_{t+1}^{-1} \tilde{\theta}_{t+1}^3 \leq \lambda_{t+1} \tilde{\theta}_t^{3T} P_t^{-1} \tilde{\theta}_t^3 + z_{t+1} e_{t+1} + \lambda_{t+1} \frac{x_{t+1}^T P_t x_{t+1}}{1+x_{t+1}^T P_t x_{t+1}} e_{t+1}^2 \quad (C.27)$$

como: $x_{t+1}^T P_t x_{t+1} \leq M \ell_{\max} P_t$

entonces: $\lambda_{t+1} \frac{x_{t+1}^T P_t x_{t+1}}{1+x_{t+1}^T P_t x_{t+1}} \leq \lambda_{t+1} \frac{M \ell_{\max} P_t}{1+M \ell_{\max} P_t}$

$$E\{\tilde{\theta}_{t+1}^{3T} P_{t+1}^{-1} \tilde{\theta}_{t+1}^3\} \leq E\{\lambda_{t+1} \tilde{\theta}_t^{3T} P_t^{-1} \tilde{\theta}_t^3\} + E\left\{\lambda_{t+1} \frac{M \ell_{\max} P_t e_{t+1}^2}{1+M \ell_{\max} P_t}\right\} \quad (C.28)$$

aplicando C.28 recursivamente, y usando 3.56

$$E\{\tilde{\theta}_{t+1}^{3T} P_{t+1}^{-1} \tilde{\theta}_{t+1}^3\} \leq E\left\{\lambda_1 \dots \lambda_{t+1} \underbrace{\tilde{\theta}_0^{3T} P_0^{-1} \tilde{\theta}_0^3}_{\tilde{\theta}} + \lambda_1 \dots \lambda_{t+1} e_1^2 \frac{M \ell_{\max} P_0}{1+M \ell_{\max} P_0} + \sum_{j=i}^{t+1} \frac{(\prod_{j=i}^{t+1} \lambda_j) M (1 - \lambda_m^P)}{m \lambda_m^P (1 - \lambda_m^{r_i P}) + M (1 - \lambda_m^P)} e_i^2\right\} \quad \text{c.s. (C.29)}$$

donde $r_i = \xi \{i-1\}$

demostración de i)

Como $E\{\tilde{\theta}_{t+1}^{3T} P_{t+1}^{-1} \tilde{\theta}_{t+1}^3\} \geq \ell_{\min} P_{t+1}^{-1} E\{\|\tilde{\theta}_{t+1}^3\|^2\} \quad (C.30)$

reemplazando (C.29) en (C.30) y

$$C_1(t) \triangleq \lambda_M^t \sigma^2 \frac{M \ell_{\text{máx}} P_o}{1 + M \ell_{\text{máx}} P_o}$$

$$E\{\|\tilde{\theta}_{t+1}^3\|^2\} \leq \frac{1}{\ell_{\text{mín}} P_{t+1}^{-1}} \left[C_1(t) + E\left\{ \sum_{i=2}^{t+1} \frac{(\pi \lambda_j)^{M(1-\lambda_m^p)} e_j^2}{m \lambda_m^p (1-\lambda_m^{rp}) + M(1-\lambda_m^p)} \right\} \right] \text{ c.s. } \quad (\text{C.31})$$

de (C.31) y (3.56):

$$E\{\|\tilde{\theta}_{t+1}^3\|^2\} \leq \frac{1}{\ell_{\text{mín}} P_{t+1}^{-1}} \left[C_1(t) + \sigma^2 \sum_{i=2}^{t+1} \frac{(\pi \lambda_j)^{M(1-\lambda_m^p)}}{\lambda_M^{t+2-i}} \right] \text{ c.s.}$$

$$\leq \frac{(1-\lambda_m^p)}{m(1-\lambda_m^{rp})\lambda_m^p} \left[C_1(t) + \sigma^2 \sum_{i=2}^{t+1} \lambda_M^{t+2-i} \right] \text{ c.s.}$$

$$\leq \frac{(1-\lambda_m^p)}{m(1-\lambda_m^{rp})\lambda_m^p} \left[C_1(t) + \sigma^2 \frac{\lambda_M - \lambda_M^{t+1}}{1-\lambda_M} \right] \text{ c.s.}$$

$$\therefore \limsup_{t \rightarrow \infty} E\{\|\tilde{\theta}_{t+1}^3\|^2\} \leq \frac{\sigma^2(1-\lambda_m^p)}{m \lambda_m^p} \cdot \frac{\lambda_M}{1-\lambda_M} \quad (\text{C.32})$$

con lo que se demuestra i.

demostración de ii)

de (C.31)

$$E\{\|\tilde{\theta}_{t+1}^3\|^2\} \leq \frac{1}{\ell_{\min}^{p-1} t+1} \left[C_1(t) + \sigma^2 \sum_{i=p+3}^{t+1} \frac{\lambda^{t+2-i} M(1-\lambda^p)}{m \lambda^p (1-\lambda^{riP})} + \sum_{i=2}^{p+2} \sigma^2 \lambda^{t+2-i} \right] \text{ c.s.}$$

$$\therefore \limsup_{t \rightarrow \infty} E\{\|\tilde{\theta}_{t+1}^3\|^2\} \leq \lim_{t \rightarrow \infty} \left[\frac{1}{\ell_{\min}^{p-1} t+1} \sum_{i=p+3}^{t+1} \frac{\lambda^{t+2-i} M(1-\lambda^p)}{m \lambda^p (1-\lambda^{riP})} \right] \text{ c.s.}$$

$$\leq \left[\frac{1-\lambda^p}{m \lambda^p} \right] \cdot \lim_{t \rightarrow \infty} \sum_{i=p+3}^{t+1} \left(\frac{\lambda^{t+2-i} \sigma^2 M(1-\lambda^p)}{m \lambda^p (1-\lambda^{riP})} \right) \text{ c.s.}$$

y en virtud del siguiente lema queda demostrada la segunda parte

LEMA

$$\lim_{t \rightarrow \infty} \sum_{i=p+3}^{t+1} \frac{\lambda^{t+2-i}}{1-\lambda^{riP}} \leq \frac{1}{\ln(1/\lambda)}$$

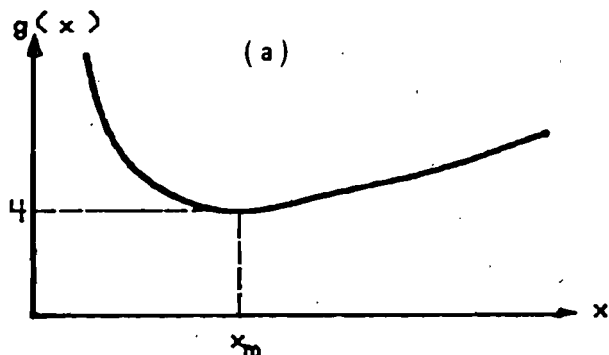
En efecto:

$$\sum_{i=p+3}^{t+1} \frac{\lambda^{t+2-i}}{1-\lambda^{riP}} \leq \sum_{i=p+2}^{t+1} \frac{\lambda^{t+2-i}}{1-\lambda^{i-p-1}} = \lambda^{t+1-p} \sum_{i=p+2}^{t+1} \frac{\lambda^{p-i+1}}{1-\lambda^{i-p-1}} \quad (\text{C.33})$$

llamando $g(x) = \frac{1}{\lambda^x(1-\lambda^x)}$, $\lambda < 1$, $x > 0$

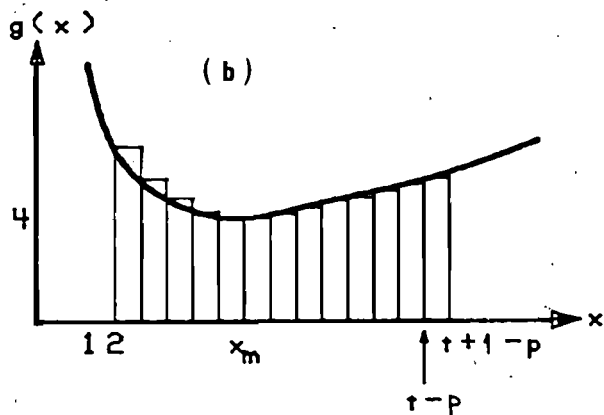
$g(x)$ tiene una forma como la de la figura (a)

$$\text{con } \begin{cases} x_m = -\ln 2 / \ln \lambda \\ g(x_m) = 4 \end{cases}$$



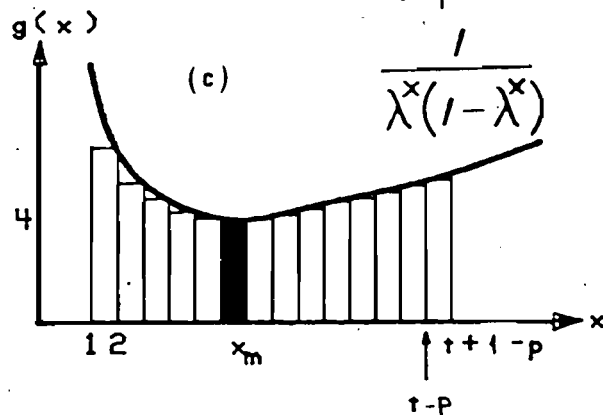
$$\sum_{i=p+3}^{t+1} \frac{\lambda^{p-i+1}}{(1-\lambda^{i-p-1})} = \sum_{i=p+3}^{t+1} g(i-p-1) \quad (\text{C.34})$$

luego, la suma C.34 está representada por el área de los rectángulos de la figura (b). Es fácil demostrar, que habiendo un solo mínimo relativo ($x > 0$) de $g(x)$, rotando todos los rectángulos a la izquierda de x_m , dicha área queda acotada por la integral:



$$\sum_{i=p+3}^{t+1} g(i-p-1) < \int_1^{t+1-p} g(x) dx$$

$$\therefore \sum_{i=p+3}^{t+1} g(i-p-1) < \int_1^{t+1-p} \frac{dx}{\lambda^x(1-\lambda^x)} =$$



$$= \frac{1}{\ln \lambda} \left[-\frac{1}{\lambda^{t+1-p}} - \ln(1-\lambda^{t+1-p}) + (t+1-p) \ln \lambda + \frac{1}{\lambda} + \ln \frac{1-\lambda}{\lambda} \right] \quad (\text{C.35})$$

de (C.35) y (c.33)

$$\lim_{t \rightarrow \infty} \left\{ \lambda^{t+1-p} \sum_{i=p+3}^{t+1} \frac{\lambda^{p-i+1}}{1-\lambda^{i-p-1}} \right\} < \frac{1}{\ln(1/\lambda)} \quad (\text{C.36})$$

con lo que se demuestra el lema.

REFERENCIAS

- [1] DUMONT, G.A. "Self tuning control of a chip refiner motor load" Automática Vol. 18-3, pp 307-314, 1982.
- [2] FORTESCUE T.R., KERSHENBAUM L.S., YDSTIE B.E. "Implementation of self tuning regulators with variable forgetting factors". Automática Vol 17-6, pp 831-835, 1981.
- [3] FROMME, G. "Self optimizing controller employing microprocessor for plants with Slowly or discontinuously Varying parameters". Process automation 1982-2, pp 93-99, 1982.
- [4] ISERMAN, R. "Parameter adaptive control algorithms-A tutorial" Automática Vol 18-5, pp 513-528, 1982.
- [5] MAYNE, D.Q., OSORIO-CORDERO A. "Deterministic convergence of a Selftuning regulator with variable forgetting factor". IEE Proc. D Vol 128-1, enero 1981.
- [6] NORTON, J.P. "Recursive identification and prediction of time-varying systems using extra information". IFAC Symp.

- on Identification and System Parameter Estimation. Tbilisi URSS-1978.
- [7] RAJBMAN, N.S.,
CHADEEV V.M. "On the A-S-I concept" Proc., 8th. Congress IFAC, Vol 2, 1982.
- [8] SCHUMANN, R. "Digital parameter-adaptive control of an air-conditioning plant". Automatica Vol 18-5, pp 569-575, 1982.
- [9] STANKOVIĆ, SS "A new adaptive real-time identification algorithm" Ifac Symp. on Identification and System Parameter Estimation, Tbilisi URSS, 1976.
- [10] WISS, C. "Predictive identification and control applied to zone melting process" Elec. Letters Vol 17-18, September 1981.
- [11] WITTENMARK B. "A two-level estimator for time-varying systems". Automatica Vol 15, pp 85-89, 1979.
- [12] CLARKE D., GAWTHROP P.J. "Self tuning controller". Proc. IEE Vol 122, pp 929-934, 1975.
- [13] ALBERT SITTNER "A method of computing least squares estimators that keep up with the da-

- ta". Siam Journal of Control, Ser. A
Vol 3-3, 1966.
- [14] LOZANO R. "Adaptive control with forgetting factor" Proc. 8th. IFAC Congress, Vol 2, 1982.
- [15] LOZANO R., LANDAU I. "Redesign of explicit and implicit discrete time model reference adaptive control schemes" Int. J. Control Vol. 33-2, pp. 247-268, 1981.
- [16] LOZANO R. "Identification of time varying parameters" 1983.
- [17] CANETTI R. "Algunos algoritmos de identificación de sistemas discretos". Reporte proyecto 5103, Instituto de Ingeniería, UNAM, 1986.
- [18] CANALES R., CANETTI R. "Síntesis de algoritmos de control autosintonizable en una microcomputadora 8748". Reporte proyecto 2123, Instituto de Ingeniería, UNAM, 1984.
- [19] GOODWIN G.C., PAYNE R.L. "Dynamic system identification: Experiment design and data analysis". Academic Press 1977.
- [20] GOODWIN G.C., SIN K.S. "Adaptive filtering, prediction and control" 1984.

- [21] SAGE A.P.,
MELSA J.L. "Estimation theory with applications to communications and control",
Mc. Graw Hill, New York.
- [22] ANDERSON B.,
MOORE J. "Optimal filtering" Prentice Hall,
New Jersey 1979.
- [23] YOUNG P. "Recursive approaches to time series analysis" Annals of the Institute of Mathematics and its Applications, Mayo-Junio 1974.
- [24] ESPAÑA M., DOMINGUEZ R., FLORES C. "Modelos matemáticos de cuencas hidrológicas. Aplicación al Río Temporal". Reporte Proyecto 2309, Instituto de Ingeniería, UNAM 1983.
- [25] ESPAÑA M. "Algoritmos de predicción lineal óptima para sistemas discretos monovariables". Serie Azul del Instituto de Ingeniería, # 484, UNAM 1984.
- [26] DOMINGUEZ R.,
CHIN M. "Estudios hidrológicos para la predicción de avenidas en la cuenca del Río Pánuco. 3a. Etapa". Reporte Proyecto 3127, Instituto de Ingeniería, UNAM 1985.
- [27] LANDAU I., LOZANO R. "Unification and evaluation of discrete time explicit model adap-

- tive control designs" Automatica 17, 1981.
- [28] JENSEN K., WIRTH N. "Pascal User Manual" Springer Verlag 1974.
- [29] ASTROM K.J. "Introduction to stochastic control theory". Academic Press 1970.
- [30] PAPOULIS A. "Probability, random variables and stochastic processes". Mc. Graw Hill Co. 1965.
- [31] WITTENMARK B. "A self-tuning predictor". IEEE Trans. on A.C. Vol 19, Diciembre 1974.
- [32] ASTROM K.J., WITTENMARK B. "On self-tuning regulators". Automatica Vol. 9 pp. 185-189, 1973.
- [33] TODINI E., WALLIS J.R. "Using C.L.S. for daily or longer period rain fall-runoff modelling". Proc. Workshop on Mathematical Models for Surface Water Hydrology. Pisa, Italia. John Wiley & Sons 1977.
- [34] DE KEYSER R., VAN CAUWENBERGHE A. "A self-tuning predictor". Automatica 17, 1. Enero 1981.
- [35] CANETTI R. "Programas de Identificación y predicción de sistemas dinámicos". Reporte Proyecto 5103, Instituto de Ingeniería, UNAM 1986.

- [36] HOUSEHOLDER A.S., "The theory of matrices in numerical analysis". Dover Pub. Inc., New York 1975.
- [37] GAUSS K.F., "Theory of the motion of the heavenly bodies" ("Theoria motus corporum coelestium", 1809). Dover Pub. Inc., New York 1969.
- [38] PLACKETT, R.L., "Some theorems in least squares", Biometrika Vol. 37, p 149, 1950.
- [39] LOZANO, R. "Identification of time-varying linear models" Proceedings 22nd C.D.C. IEEE, San Antonio Texas, 1983.
- [40] ANDERSON B.D.O. "Adaptive systems, lack of persistency of excitation and Bursting phenomena" Automatica, Vol 21-3, pp, 247-258, 1985.
- [41] YDSTIE, B.E. "Adaptive control and estimation with forgetting actors" 7th IFAC Symp. on Identification and system parameter

- Estimation, New York, Julio 1985.
- [42] ASTROM K.J. "Lectures on the identification problem, the least-squares method". Report 6806. Sept. 1968. Lund Institute of Technology. Division of Automatic Control. Lund. Suecia.
- [43] STREJC, V. "Least squares in identification theory". Kybernetika Vol 13 pp 83-105. Praga 1977
- [44] BELLMAN R.E. "Matrix Analysis" Mc Graw-Hill, Nueva York. 1968.
- [45] GANTMACHER F.R. "The theory of matrices" (Traducción de: "TEORIYA MATRITS") Chelsea Pub. Co., Nueva York 1977.
- [46] ASTROM K.J., WITTENMARK B. "Self-tuning controllers based on pole-zero placement". IEE Proceedings Vol 127, Pt. D No 3, Mayo 1980.
- [47] HAGGLUND T. "New Estimation Techniques for Adaptive Control". Tesis Doctoral, Universi

dad de Lund, Depto. de Control Automá
tico. Suecia 1983.

|48| JOHNSTONE R.M., ANDERSON B.D.O. "Exponential convergence
of recursive least squares with expo-
nential forgetting factor adaptive
control". Systems & Control Letters,
Vol 2 No 2, Agosto 1982.